

NASA/CP-2010-216205



# Selected Papers Presented at MODSIM World 2009 Conference and Expo

*Edited by  
Thomas E. Pinelli  
Langley Research Center, Hampton, Virginia*

---

March 2010

## NASA STI Program . . . in Profile

Since its founding, NASA has been dedicated to the advancement of aeronautics and space science. The NASA scientific and technical information (STI) program plays a key part in helping NASA maintain this important role.

The NASA STI program operates under the auspices of the Agency Chief Information Officer. It collects, organizes, provides for archiving, and disseminates NASA's STI. The NASA STI program provides access to the NASA Aeronautics and Space Database and its public interface, the NASA Technical Report Server, thus providing one of the largest collections of aeronautical and space science STI in the world. Results are published in both non-NASA channels and by NASA in the NASA STI Report Series, which includes the following report types:

- **TECHNICAL PUBLICATION.** Reports of completed research or a major significant phase of research that present the results of NASA programs and include extensive data or theoretical analysis. Includes compilations of significant scientific and technical data and information deemed to be of continuing reference value. NASA counterpart of peer-reviewed formal professional papers, but having less stringent limitations on manuscript length and extent of graphic presentations.
- **TECHNICAL MEMORANDUM.** Scientific and technical findings that are preliminary or of specialized interest, e.g., quick release reports, working papers, and bibliographies that contain minimal annotation. Does not contain extensive analysis.
- **CONTRACTOR REPORT.** Scientific and technical findings by NASA-sponsored contractors and grantees.

- **CONFERENCE PUBLICATION.** Collected papers from scientific and technical conferences, symposia, seminars, or other meetings sponsored or co-sponsored by NASA.
- **SPECIAL PUBLICATION.** Scientific, technical, or historical information from NASA programs, projects, and missions, often concerned with subjects having substantial public interest.
- **TECHNICAL TRANSLATION.** English-language translations of foreign scientific and technical material pertinent to NASA's mission.

Specialized services also include creating custom thesauri, building customized databases, and organizing and publishing research results.

For more information about the NASA STI program, see the following:

- Access the NASA STI program home page at <http://www.sti.nasa.gov>
- E-mail your question via the Internet to [help@sti.nasa.gov](mailto:help@sti.nasa.gov)
- Fax your question to the NASA STI Help Desk at 443-757-5803
- Phone the NASA STI Help Desk at 443-757-5802
- Write to:  
NASA STI Help Desk  
NASA Center for AeroSpace Information  
7115 Standard Drive  
Hanover, MD 21076-1320

NASA/CP-2010-216205



# Selected Papers Presented at MODSIM World 2009 Conference and Expo

*Edited by*  
*Thomas E. Pinelli*  
*Langley Research Center, Hampton, Virginia*

Proceedings of a conference sponsored by the  
National Aeronautics and Space Administration  
and the Virginia Modeling, Analysis & Simulation Center  
and held at the Virginia Beach Convention Center  
Virginia Beach, Virginia  
October 14-16, 2009

National Aeronautics and  
Space Administration

Langley Research Center  
Hampton, Virginia 23681-2199

March 2010

Trade names and trademarks are used in this report for identification only. Their usage does not constitute an official endorsement, either expressed or implied, by the National Aeronautics and Space Administration.

Available from:

NASA Center for Aerospace Information (CASI)  
7115 Standard Drive  
Hanover, MD 21076-1320  
443-757-5802

## PREFACE

Selected papers from MODSIM World 2009 Conference & Expo are contained in this NASA Conference Publication (CP). Managed by the Center for Public/Private Partnership (CP3), MODSIM World is a unique multi-disciplinary international conference and exposition for the exchange of modeling & simulation knowledge, research, and technology across industry, government, and academia. MODSIM World 2009 was held in Virginia Beach, VA, October 14-16, 2009, at the Virginia Beach Convention Center. The theme of the 2009 conference & expo was "21<sup>st</sup> Century Decision-Making: The Art of Modeling & Simulation". MODSIM World 2009 included pre-conference workshops, presentations, papers, working group meetings, exhibits, and a K-20 student career outreach program.

The 79 papers submitted for presentation were peer-reviewed and 64 were accepted for presentation. As a condition of acceptance, the first authors were responsible for securing/obtaining all permissions associated with the general release and public availability of each paper. To be included in the proceedings, first authors also agreed to grant NASA the right to publish their papers. There are 53 papers in this NASA CP.

Progress in simulation-based engineering and science holds great promise for the pervasive advancement of knowledge and understanding through discovery. As an interdisciplinary activity, simulation-based engineering and science enables more accurate design, development, and integration of complex engineering systems, sub-systems, and individual components; improves the conduct and understanding of engineering and science; provides more precise representation of operational environments; and improves the predictive capabilities of system performance based on quantifiable margins and uncertainties. Modeling and simulation-based engineering allow engineers and scientists to model processes too costly or impossible to investigate in other ways. Tools such as modeling and simulation represent new ways of doing science, along with experimentation to test hypotheses. Modeling and simulation-based engineering science has become an indispensable tool for investigating and resolving complex scientific and technological problems in a multitude of disciplines. Simulation-based engineering and science is also essential to the success of NASA's research, missions, and projects.

MODSIM World 2009 Conference & Expo was divided into six "standalone" tracks and one "cross-cutting" track. The 53 papers are distributed (in this publication) according to the following tracks: Defense & Homeland Security (7), Education & Training (4), Engineering & Science (6), Health & Medicine (5), Human Dimensions (Factors) (14), Serious Games (12), and Transportation & Logistics (6). There are two appendices in this publication. Appendix A contains the names of the conference organizing committee members and Appendix B includes the names of the individuals who chaired and organized the technical tracks. MODSIM World 2010 Conference & Expo will be held October 13-15, 2010, at the Hampton Roads Convention Center located in Hampton, VA.



# CONTENTS

Preface

Selected Papers

## I. Defense & Homeland Security Track

Bell, Benjamin; Phillip Short; and Stewart Webb – <i>Walking the Walk/Talking the Talk: Mission Planning with Speech-Interactive Agents</i> .....	1
Deller, Sean; Rabadi, Ghaith A; Michael I. Bell; Shannon R. Bowling; and Andreas Tolk – <i>Modeling the Information Age Combat Model: An Agent-Based Simulation of Network Centric Operations</i> .....	9
Faisal, Mahmud; S.G.M. Hossain; and Jobair Bin Alam – <i>Low-Cost Rescue Robot for Disaster Management in a Developing Country: Development of a Prototype Using Locally Available Technology</i> .....	17
Hanold, Gregg T. and David T. Hanold – <i>Route Generation for a Synthetic Character (BOT) Using a Partial or Incomplete Knowledge Route Generation Algorithm in UT2004 Virtual Environment</i> .....	25
Hase, Chris – <i>Adaptive Planning: Understanding Organizational Workload to Capability/Capacity through Modeling and Simulation</i> .....	31
Rahmani, Sonya and Otto von def Hoff – <i>The Application of Modeling and Simulation in Capacity Management within the ITIL Framework</i> .....	37
Tokgoz, Berna Eren; Volkan Cakir; and Adrain V. Gheorghe – <i>Handling Emergency Management in [an] Object Oriented Modeling Environment</i> .....	43

## II. Education & Training Track

Bowman, Arthur W. – <i>Professor Created On-line Biology Laboratory Course</i> .....	49
Hamilton, Paul Lawrence and Robert Brittain – <i>The Umbra Simulation and Integration Framework Applied to Emergency Response Training</i> .....	55
Rheinsmith, Richard; Jeffery Wallace; Warren Bizub; Andy Ceranowicz; Dannie Cutts; Edward T. Powell; Paul Gustavson; Robert Lutz; and Terrell McCloud – <i>Joint Composable Object Model and LVC Methodology</i> .....	61
Tocosa, Corliss A. and Taryn Cuper – <i>Using Game Theory to Qualitatively Analyze Global Relations</i> .....	69

### III. Engineering & Science Track

Arin, Arif; Ghaith Rabadi; and Resit Unal – <i>Tuning Parameters in Heuristics by Using Design of Experiments Methods</i> .....	73
Khasawneh, Mahmoud T.; Jun Zhang; Nevan E.N. Shearer; Raed M. Jaradat; Elkin Rodriguez-Velasquez; and Shannon R. Bowling – <i>An Exploratory Study of the Butterfly Effect Using Agent-Based Modeling</i> .....	79
Kuhn, D. Richard; Raghu Kacker; and Yu Lei – <i>Random vs. Combinatorial Methods for Discrete Event Simulation of a Grid Computer Network</i> .....	83
Russell, Michael S. – <i>Assessing Requirements Volatility and Risk Using Bayesian Networks</i> .....	89
Schwartz, Richard J. And Andrew C. McCrea – <i>Virtual Diagnostic Interface: Aerospace Experimentation in the Synthetic Environment</i> .....	103
Shamseldin, Ramez Ahmed and Shannon R. Bowling – <i>Using Agent-Base[d] Models to Optimize Large Scale Network[s] for Large System Inventories</i> .....	109

### IV. Health & Medicine Track

Anton, John J. – <i>The Application of Modeling and Simulation to the Behavioral Deficit of Autism</i> .....	115
Fowlkes, Jennifer; Sandra Dickinson; and Todd Lazarus – <i>Blended Training for Combat Medics</i> .....	123
Hester, Robert L.; Richard L. Summers; Radu Ilescu; Joyee Esters; and Thomas G. Coleman – <i>DigitalHuman (DH): An Integrative Mathematical Model of Human Physiology</i> .....	129
Liu, Miao; Lee A. Belfore; Yuzhong Shen; and Mark W. Scerbo – <i>Uterine Contraction Modeling and Simulation</i> .....	135
Quartuccio, Katherine; Patricia Lisk; and Karen Mittura – <i>Proposition: High-Fidelity Simulations for Remediation</i> .....	141

### V. Human Dimensions (Factors) Track

Cover, Mathew and Thomas Schnell – <i>Modeling Pilot Behavior for Assessing Integrated Alert and Notification Systems on Flight Decks</i> .....	145
Cuper, Taryn – <i>Modeling the Effects of Stress: An Approach to Training</i> .....	151
Daiker, Ron and Thomas Schnell – <i>Development of a Human Motor Model for the Evaluation of an Integrated Alerting and Notification Flight Deck System</i> .....	157
Ellis, Kyle and Thomas Schnell – <i>Eye Tracking Metrics for Workload Estimation in Flight Deck Operation</i> .....	163
Feary, Michael – <i>A Toolset for Supporting Iterative Human Automation: Interaction in Design</i> .....	169

Grande, Darby; J. Todd Black; Jared Freeman; Tim Sorber; and Daniel Serfaty – <i>Use of Dynamic Models and Operational Architecture to Solve Complex Navy Challenges</i> .....	175
Handley, Holly A.H. and Nancy P. Houston – <i>NATO Human View Architecture and Human Networks</i> .....	181
Johnson, W. Lewis <i>A Simulation-Based Approach to Training Operational Cultural Competence</i> .....	187
Mahmud, Faisal and Hasan Samiul – <i>Propagation of New Innovations: An Approach to Classify Human Behavior and Movement from Available Social Network Data</i> .....	195
Phillips, Rachel and Poornima Madhavan – <i>Effects of Secondary Task Modality and Processing Code on Automation Trust and Utilization During Simulated Airline Luggage Screening</i> .....	199
Rohloff, Kurt – <i>Automated Discovery and Modeling of Sequential Patterns Preceding Events of Interest</i> .....	205
Santarelli, Thomas; Aaron Pepe; Larry Rosenzweig; John Paulus; and Ahn Na Yi – <i>Representation of Cultural Role-Play for Training</i> .....	211
Vaccaro, James and Clark Guest – <i>Modeling Social Influence via Combined Centralized and Distributed Planning Control</i> .....	217
Zhang, Guangfan; Roger Xu; Wei Wang; Jiang Li; Tom Schnell; and Mike Keller – <i>Individualized Cognitive Modeling for Close-Loop Task Mitigation</i> .....	223

## **VI. Serious Games Track**

Bohil, Corey J. And Frank A. Biocca – <i>Cognitive Modeling of Video Game Player User Experience</i> .....	229
Durlach, Paula J.; Heather Priest; Glenn A. Martin; and Jay Saffold – <i>Developing Collective Training for Small Aerial Systems Employment</i> .....	235
Jeppsen, Isaac Ben – <i>Methods Used in Game Development to Foster FLOW</i> .....	241
Jones, Phillip and Taryn Cuper – <i>Making Games Not Work: Paradoxes Embedded in Game-Based Training and Concepts for Overcoming Them</i> .....	247
Jones, Richard L. – <i>QuickStrike ASOC Battlefield Simulation: Preparing the War Fighter to Win</i> .....	253
Manepalli, Suchitra; Yuzhong Shen; Hector M. Garcia; and Kaleen Lawsure – <i>Virtual VMASC: A 3D Game Environment</i> .....	259
Maulitz, Russell; Thomas Santarelli; Joann Barnieu; Larry Rosenzweig; Ahn Na Yi; Wayne Zachary; and Bonnie O'Connor – <i>TEACH (Train to Enable/Achieve Culturally Sensitive Healthcare)</i> .....	265
McGowan, Clement and Benjamin Pecheux – <i>Serious Games that Improve Performance</i> ..	271
Nguyen, Duc T.; Ahmed Ali Mohammed; and Subhash Kadiam – <i>A Chess-Like Game for Teaching Engineering Students to Solve Large System of Simultaneous Linear Equations</i> .....	277

Phillips, Mark; Jackie Scolaro; and Daniel Scolaro – <i>The Emergence of Agent-Based Technology as an Architectural Component of Serious Games</i> .....	283
Weil, Shawn A.; Alexandra Geyer; Paul Picciano; Alice Leung; and William Ferguson – <i>Pedagogy and Assessment in an Alternate Reality Game (ARG) Training Event</i> .....	289
<b>VII. Transportation &amp; Logistics Track</b>	
Arnaout, George M. And Shannon R. Bowling – <i>Improving Customer Waiting Time at a DMV Center Using Discrete-Event Simulation</i> .....	295
Foytik, Peter and Mike Robinson – <i>Implementing and Simulating Dynamic Traffic Assignment with Intelligent Transportation Systems in Cube Avenue</i> .....	301
Garcia, Christopher and Ghaith Rabadi – <i>An Optimization Model for Scheduling Problems with Two-Dimensional Spatial Resource Constraints</i> .....	307
Hanssen, Joel – <i>Fundamental Travel Demand Model Example</i> .....	313
Wang, Jie; Yuzhong Shen; and Asad Khattak – <i>Visualization of Traffic Accidents</i> .....	317
Wang, Jie; Yiannis Papelis; Yuzhong Shen; Ozhan Unal; and Mecit Cetin – <i>High-Fidelity Roadway Modeling and Simulation</i> .....	323
Appendix A – MODSIM World 2009 Conference Organizing Committee .....	329
Appendix B – MODSIM World 2009 Track Chairs and Track Organizers .....	331

# Walking the walk/Talking the talk: Mission Planning with Speech-Interactive Agents

Benjamin Bell<sup>1</sup>; Philip Short<sup>2</sup>; Stewart Webb<sup>2</sup>

<sup>1</sup>CHI Systems, Inc; <sup>2</sup>Ael Ltd

*bbell@chisystems.com; phil.short@baesystems.com; stewart.webb@baesystems.com*

**Abstract.** The application of simulation technology to mission planning and rehearsal has enabled realistic overhead 2-D and immersive 3-D “fly-through” capabilities that can help better prepare tactical teams for conducting missions in unfamiliar locales. For aircrews, detailed terrain data can offer a preview of the relevant landmarks and hazards, and threat models can provide a comprehensive glimpse of potential hot zones and safety corridors. A further extension of the utility of such planning and rehearsal techniques would allow users to perform the radio communications planned for a mission; that is, the air-ground coordination that is critical to the success of missions such as close air support (CAS). Such practice opportunities, while valuable, are limited by the inescapable scarcity of complete mission teams to gather in space and time during planning and rehearsal cycles. Moreover, using simulated comms with synthetic entities, despite the substantial training and cost benefits, remains an elusive objective. In this paper we report on a solution to this gap that incorporates “synthetic teammates” – intelligent software agents that can role-play entities in a mission scenario and that can communicate in spoken language with users. We employ a fielded mission planning and rehearsal tool so that our focus remains on the experimental objectives of the research rather than on developing a testbed from scratch. Use of this planning tool also helps to validate the approach in an operational system. The result is a demonstration of a mission rehearsal tool that allows aircrew users to not only fly the mission but also practice the verbal communications with air control agencies and tactical controllers on the ground. This work will be presented in a CAS mission planning example but has broad applicability across weapons systems, missions and tactical force compositions.

## 1. MISSION PLANNING, REHEARSAL GAPS

Mission planning and mission rehearsal are routinely performed using sophisticated automation and simulation technology. Planners, commanders and their personnel are now able to “fly-through” a mission, employing threat models and advanced visualization tools that can render accurate geospatial and terrain data. Such realistic simulations help prepare tactical teams for conducting missions in unfamiliar locales. For instance, detailed terrain data can prepare aircrew to recognize relevant landmarks and hazards, and threat models can provide a comprehensive glimpse of potential hot zones and safety corridors.

There is one aspect of mission performance that is critical to success which has remained beyond the reach of even the most advanced mission planning tools: verbal communication. Missions such as close air support (CAS) depend heavily on timely, succinct, correct and relevant spoken dialogue between air and ground elements. Joint Terminal Attack Controllers (JTACs) and CAS-rated aircrew typically train on live ranges to reach some critical performance level. But once deployed, practice opportunities are severely limited by the inescapable scarcity of complete mission teams to gather in space and time during planning and rehearsal cycles.

## 2. POTENTIAL SOLUTIONS

Mission planning and rehearsal should allow users to practice the radio communication along with the other aspects of mission performance. In CAS, for instance, the air-ground coordination is critical to the success and safety of the mission and should be represented in walk-through/fly-through activities. Unfortunately this is seldom the practice, due largely to the separation in time and space of the respective staffs in the air and ground elements planning and rehearsing the mission.

In general there are two constructs for meeting this gap: (1) use of live confederates as role-players; and (2) software simulations of entities in the scenario.

### 2.1 The “Wetware” Option

Option 1 is the use of live personnel and requires no sophisticated technology. But there are cost and access penalties incurred by the use of live role-players:

1. When participants are drawn from the trainee ranks their time is spent on providing cues to keep the scenario moving rather than on effective mission rehearsal;
2. When drawn from the instructor ranks, role-playing interferes with performance

assessment, since instructors are called upon to divide their attention between evaluating mission success and role-play;

3. It creates variability that makes standardizing rehearsal difficult due to the human element influencing events in each scenario.
4. Costs arise from compensating, transporting and lodging role-players at dedicated facilities.
5. Availability is compromised because expert role-players can be exceedingly difficult to arrange, particularly for missions in new areas of operation or that employ novel tactics or recent equipment changes.

The consequence is that access to mission planning and rehearsal is measured and scheduled and conducted at specific facilities.

## 2.2 The Software Option

Option 2 is to employ software simulations of entities in the scenario in lieu of live role-players. We are exploring this option by introducing intelligent, interactive agents into a mobile mission planning package. We commenced this investigation by defining the core capabilities needed for synthetic teammates. To provide interaction effectively for mission planning and rehearsal, our analysis revealed that synthetic teammates must possess the following capabilities:

1. simultaneous execution of: taskwork (e.g., flying the aircraft, working the console); teamwork (interacting with other members of the team); and measurement (for subsequent analysis and feedback);
2. interaction via spoken language (required for rehearsing mission communications);
3. modulating behaviors to replicate various error modes, to allow for varying the proficiency of the synthetic team members (important for playing out contingencies and stress-testing the plan).

We expect that the above generic requirements extend well beyond conventional computer-generated forces (CGFs), semi-automated forces (SAFs), and game-based artificial intelligence, or "AI"s – largely scripted entities with limited abilities to respond to events beyond a predefined range of simple behaviors. CGF/SAF technologies do have an important role to play, but for our purposes they fall short of addressing specific needs that remain unmet. To meet these needs, we are employing

cognitive modeling using CHI Systems' computational development tool, iGEN®, for encapsulating human expertise and behavior in synthetic agents (Zachary, LeMentec & Ryder, 1996). Sophisticated agents, such as those which may be built using iGEN, can provide dialogue-capable synthetic teammates to reduce reliance on human role-players and make mission planning and rehearsal more accessible, less costly, and more standardized.

## 2.3 Previous Work: On-Demand Team Training

Mission planning and rehearsal each share a simulation dimension with training, where this technique has received the most attention. We first integrated the cognitive modeling approach with full speech interaction for a US Navy program called Synthetic Cognition for Operational Team Training (SCOTT) (Zachary, *et al.*, 2001). SCOTT is a simulation-based practice and training environment in which a single human crewmember of an E-2C tactical crew can train in cross-platform coordination skills by interacting verbally with synthetic teammates, both on and off the E-2C. More recently, we developed Synthetic Teammates for Realtime Anywhere Training and Assessment (STRATA), a Close Air Support (CAS) trainer built on the progress made under SCOTT but using more sophisticated cognitive modeling and more advanced speech technologies (Bell, Johnston, Freeman & Rody, 2004). The emphasis in STRATA was to validate "on-demand team training" by making the instructor and the other CAS team members, such as the Forward Air Controller, entirely optional. Most recently, we developed the Virtual Interactive Pattern Environment and Radiocomms Simulator (VIPERS). VIPERS offers users opportunities for guided practice and feedback in radio communications skills and decision making in a simulated pattern environment (Bell, Ryder & Pratt, 2008). The format of this practice is simulation-based training with intelligent software agents performing in both tutoring roles and synthetic teammate roles, in a laptop-based portable application for anytime/anywhere training. Specifically, VIPERS provides three types of speech-interactive entities: (1) a synthetic instructor that provides coaching and feedback during scenarios and makes assessments to be used in a debrief; (2) a synthetic controller that maintains knowledge of all aircraft in the pattern and verbally responds to clearance requests and issues directives to all aircraft in the pattern; and (3) synthetic pilots/aircraft in the pattern behaving appropriately and making radio calls.

### 3. CAS MISSION PLANNING & REHEARSAL: AN EXPLORATORY STUDY

Work reported in this paper was aimed at applying some of the capabilities we had developed in the training domain to explore more realistic and more accessible mission planning and rehearsal tools. Our focus was on users in high OPTEMPO contexts, engaged in missions requiring a great deal of teamwork. We looked particularly at cases where teams are distributed and where verbal communication enjoys a key role in mission coordination, selecting CAS for this study. To accelerate our research, we employed a fielded mission planning and rehearsal tool, so that we could devote our attention to investigating the utility of speech-interactive synthetic teammates rather than on creating a suitable testbed. The tool we employed is called the Combined Arms Gateway Environment (CAGE).

#### 3.1 Summary of CAGE

CAGE, developed by Ael, is a mission support tool that enables operators to plan, rehearse and then conduct platform specific or independent missions under a wide variety of operational conditions. The system can be configured to support the operational needs of any given operator or platform configuration. CAGE is based on an open architecture JAVA framework.

CAGE allows planners to employ the rehearsal capability to create routes, inspect and deconflict airspace, view corridors and define threat cones. Planners and mission personnel can view the mission in 2-D (top-down) and 3-D. The 3-D view provides dynamic lighting (sun, shade, moonlight) to assess the tactical implications of time of day and visibility effects (fog, haze, cloudbase) to project the visibility under the forecast weather conditions.

#### 3.2 A Human-Centric Approach

Our analysis started with a human factors integration approach by considering what features would be required of a speech interactive agent for training, planning and rehearsal; and what the associated benefits were. This was to ensure that the use of such technology was driven by the needs of the warfighter, rather being implemented as a technology push simply for its own sake. This analysis yielded four required characteristics of a speech-interactive agents:

1. Real-time – includes element of time pressure on decision making and actions;
2. Unpredictable – able to include unanticipated / unexpected events;

3. Dynamic – able to respond to user actions;
4. Replicates the modality of real dialogue – user must process information in same way (e.g., cannot simply read prompts from a screen).

We also identified the following anticipated benefits:

1. Reduced instructor input – elements of automation mean that multiple users can train or rehearse concurrently on multiple systems, without the need for multiple instructors or mission commanders.
2. Increased combat readiness – availability of multiple, less costly systems reduces reliance on expensive, scarce simulators;
3. Reduced flying / simulator hours – system enables training that was previously only possible in the air or on a full mission simulator;
4. Reduced cost - as a result all of the above.

#### 3.3 Needs Analysis

A high-level needs analysis was performed for a CAS scenario. This was a limited analysis, in alignment with the exploratory nature of this research, and so was focused specifically on voice interaction. This entailed performing a Hierarchical Task Analysis (HTA) for the scenario, and reviewing each relevant step<sup>1</sup> to identify:

- The objective for that step.
- How to gauge that the objective has been achieved, *i.e.* the measure of effectiveness (MoE);
- The required inputs for that step (what the instructor has to include over and above the synthetic agent component in order to accomplish the step);
- The specific benefits that the synthetic agent provides, which would not have been achieved by other means (*e.g.* by displaying the dialogue as text on a screen);
- What the technology must be able to do in order to provide the required benefit.

<sup>1</sup> By 'relevant step' we mean those steps that involve the user doing something, as the HTA also covers the actions of the Joint Terminal Attack Controller (*i.e.* the actor being 'played' by the synthetic agent).

The results of the HTA were captured against the following criteria (example outcomes shown in parentheses):

- Task: (*Look for described area and features*).
- Objective: (*Rapidly and accurately identify areas based on description of the visual scene*).
- MoE: (*Identify target within elapsed time parameters*).
- Required inputs: (*A representation of the visual scene that relates to the descriptions being provided*).
- Benefit: (*Synthetic agent allows natural interaction between user and JTAC, with correct sensory input (auditory) and output (speech)*).
- Requirement for agent: (*able to provide descriptions that relate to the visual scene provided*).

### 3.4 Technical Approach

To bound the scope of our initial experiment, we created a set of CAS scenarios, focusing on dialogue between the pilot and JTAC, allowing for alternative dialogue branches and error correction. The complexity of the scenarios determines the necessary sophistication of the grammar, synthesized voice, and agent model. For this exploratory effort, therefore, the scenarios were limited to specific phases of a representative CAS mission.

In order to efficiently introduce voice capabilities to CAGE, a TCP socket-based network protocol using XML-based messaging was employed to enable communication between CAGE and our existing speech-enhanced synthetic agent framework. The XML schema was directly derived from the High Level Architecture (HLA) interactions used in some of our previous work (e.g., Chapman, Ryder, Bell, Wischusen & Benton, 2004). A network-based API was chosen based on direct routine calls as this approach involved minimal modification of the existing systems, each of which was able to retain its modes of operation, largely independent of the others. CAGE is responsible for loading the dialogue information (in the form of an XML file) and sharing that information with the agent framework as the scenario progresses. Data is shared at a semantic level. Position data, for instance, is shared to allow the agent component to generate the appropriate synthetic speech. CAGE determines the pace of the exercise by sending the information at the desired time intervals based on user responses and progress through the mission.

The speech-enabled agent framework consists of modules to synthesize and recognize speech, an agent implementation to respond appropriately, and a messaging framework called the Socket Executive to mediate communication among the modules and with CAGE over TCP (see Figure 1).

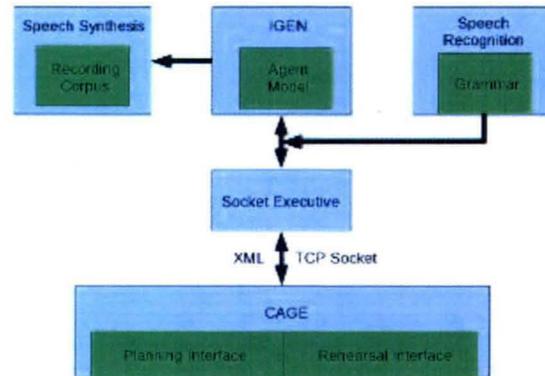


Figure 1: Architecture integrating CAGE, iGEN

We built an iGEN model to play the role of the ground-based observer - the Joint Terminal Attack Controller (JTAC) - and exported it to XML. The speech dialogues from the scenarios were divided into phrases spoken by the user and those spoken by the synthetic teammate. Speech synthesis was accomplished using the Festival speech synthesis engine (Black & Taylor, 1997) and Limited Domain Synthesis (LDOM) (Black & Lenzo, 2000), which uses phonemes derived from recordings to synthesize speech. This approach provides speech that is tactically realistic and based on concatenated recordings of domain experts; but also preserves the capability to dynamically generate speech in real-time, voicing variables such as coordinates, call signs and mission times.

One requirement of the LDOM approach is that recorded samples be collected for any lexical token in the vocabulary. This is a minimal requirement since word pair, tuples and longer phrases are permitted as well. We enhanced realism by recording phrase variants similar to those expected during mission planning and rehearsal. By carefully examining the dialogues, and constructing phrases covering the expected vocabulary including all possible numerals, call-signs, and directions, a corpus of phrases was created and then recorded by a domain expert. The recordings were volume normalized, broken into phonemes, and indexed for use by the Festival engine at runtime. Additional recordings made to accommodate revisions to the vocabulary were incorporated into the previous

corpus. Pauses were inserted into some of the communications (e.g., reading coordinates) to more realistically capture the manner in which such phrases are spoken operationally.

The user-spoken dialogue was represented semantically as a string identifier followed by any variables (e.g., coordinates, directions) optional words or phrases, and modifiers (such as "not"). These forms were coded into a Backus-Naur Form (BNF) grammar for the speech recognition engine. These semantic representations were also encoded into the grammar so that when a user-spoken phrase is recognized, the engine would return not only a plain-text representation but also the semantic frame with optional parameters included. This enabled the agent to more easily understand the recognized speech.

The socket executive uses a publish-subscribe mechanism to distribute information among the synthetic teammates, and constructs XML messages for communicating with CAGE over TCP. The synthetic teammate framework and components were built with native C++ code.

#### 4. RESULTS

##### 4.1 Scenario Definition

The implemented scenario demonstrates a mission rehearsal with a user in the role of lead CAS pilot, interacting with a synthetic JTAC. When a scenario is started, the components load their required data (CAGE loads its scenario data, the speech components load the grammar and voice data, and iGEN loads the JTAC model) and each initializes the appropriate communication channels. The user selects a call sign from a set of nominal identifiers and two-digit suffixes. The user then begins the mission and initiates communication by checking in with the chosen call-sign. Figure 2 shows a representative display at this point in the mission, with a 3-D view on the left and the 2-D view on the right.



Figure 2: Representative display in CAGE

The JTAC agent transmits a 9-line brief, based on information given to it by CAGE (the user can request a re-transmit at any point during the mission). The user then repeats the 9-line and the read-back is checked by the synthetic JTAC for accuracy. If an error is found in the readback, the user is asked to repeat any incorrect portions of the communication until it is correct (and only the incorrect portions). The JTAC agent then directs the user to the target, who must read back the targeting information, which is again checked for accuracy. Following an accurate read-back, the JTAC clears the user for attack. After attack the JTAC responds with a battle damage assessment, and the user signs off. During each exchange the JTAC waits for the appropriate response from the user, and asks the user to repeat any communication that is incorrect or unrecognizable. A representative transcript is shown in Figure 3.

User: Widow 76 this is Vader 28 checking in as fragged  
 JTAC: *Vader 28, Widow 76 Loud and dear, this is a Type 1 control, call ready to copy.*  
 User: Vader 28 Type 1 control, ready copy  
 JTAC: *IP U278, Heading 055 magnetic, Distance 9260 meters, Elevation 70 feet. Target is a Helicopter parked on western edge of dispersal. Location North 51 00.89 West 002 38.01. Mark Laser 1111 LTL 355 Magnetic. Friendlies 1000 South, Egress North to Bad Wolf. Advise when ready for remarks*  
 User: Ready to copy remarks  
 JTAC: *Final attack heading 055 through 030*  
 User: Elevation 70 feet, Location North 51 00.89 West 002 38.01. Friendlies 1km South. Laser 1111 LTL 355 magnetic. Attack heading 055 through 030 magnetic  
 JTAC: *Readback correct, report leaving IP*  
 User: Leaving IP, abort alfa romeo sierra  
 JTAC: *Widow 76, abort alfa, romeo, sierra you r target is one of 2 helicopters on the western edge of a dispersal.*  
 User: Helicopter, western edge, dispersal. Vader 28 leaving IP.  
 JTAC: *Short of target, airfield*  
 User: Short of target, airfield  
 JTAC: *North of runways, group of 8 hangars. From there, 12 o'clock 500, further set of 3 hangars, North East corner airfield. Laser on. Friendlies to South of all runways.*  
 User: Contact 10 seconds. Further 3 hangars Laser on. Visual friendlies  
 JTAC: *Right of hangars is large dispersal, in sunlight, target is helicopter on right hand side*  
 User: Contact Target, left of target further helicopter against building.  
 JTAC: *Affirm, deared hot*  
 User: In hot. Rifle away. Terminate  
 JTAC: *Terminate, Vader 28, widow 76, Delta Hotel, helicopter destroyed, End of mission.*  
 User: Target destroyed, Delta Hotel, End of Mission.

Figure 3: Representative dialogue between aircraft (user) and JTAC agent

## 4.2 Synthetic Teammate Interactions

An important design consideration is the degree of variability in whether user utterances are treated as "legal". Too restrictive an approach erroneously emphasizes syntax over semantics, frustrates users, and undermines mission planning and rehearsal objectives. Too accommodating an approach not only adds complexity to the recognition process but could introduce non-doctrinal phraseology.

There is no quick-fix solution; striking a proper balance depends on thoughtful, comprehensive consultations with subject matter experts, guided by a cognitive task analysis methodology (e.g., Zachary, Ryder & Hicinbothom, 2000). For our exploratory study we employed a CAS-rated RAF pilot and implemented logic in the JTAC agent that permits lexical and syntactic variations based on the tactical context. Each communication spoken by the user can thus be phrased in different ways; optional wording can be omitted and some alternate wordings are accepted.

This flexible grammar, combined with the selective requests for read-back (i.e., only incorrect portions of the 9-line need be repeated) afford a transparent dialogue capability. The work reported here was speaker-independent – no training to a specific voice was required. Our testers consisted of both U.K. and U.S. speakers with no noticeable differences in recognition rates among them.

Initial results showed that there was an immediate benefit to being able to practice techniques as they would be performed for real while remaining in a benign environment. For early-stage training, this removes the stress of the real situation in order to put the trainee at ease; for planning and rehearsal the realism is sufficient to provide the necessary situational awareness to adequately exercise the plan and measure an individual's performance in executing it.

Early feedback from end-users also indicates the scalability of this technology. There is significant potential to increase the richness of the training experience, including using the synthetic agent to increase the user's exposure to operational stress; to augment the simulated environment with more diverse players and to provide voice interaction in situations where it is not currently available.

## 4.3 Broader Implications

The investigation reported here provides preliminary support for the utility of speech-interactive synthetic teammates in the mission

planning and rehearsal domain. We recognize that our results are based on a limited scenario, and we are currently planning to develop more comprehensive, complex scenarios, which will require behavioral, speech and grammar components with additional sophistication.

To achieve the performance reported here in richer scenarios, we require more robust speech recognition and discourse management. We will address this by employing a dynamic grammar, where an intelligent agent activates and deactivates sub-grammars as the tactical situation changes, an approach we have reported in previous work (Bell, Johnston, Freeman & Rody, 2004). Our work has indicated that there is significant training benefit to be gained from using speech interactive agents through increased richness or improved efficiency of the training environment (Bell, Ryder & Pratt, 2008).

We are also expanding the reach of this approach through integration of the capabilities reported here with a more sophisticated testbed called the Distributed Synthetic Air Land Training (DSALT) facility operated by the UK MOD. Results from that experiment will provide a firmer foundation for assessing the utility of speech-capable synthetic teammates for training, mission planning and rehearsal.

## 5. CONCLUSION

New simulation capabilities that extend the benefits of synthetic training can yield parallel advances in mission rehearsal and mission planning. For missions that rely on effective communication and coordination, though, the verbal exchange among tactical teammates is trained, planned and rehearsed only if and when suitable role-players are available, co-located in time and place.

In this paper we have introduced speech-interactive synthetic teammates as a capability that overcomes these limitations and provides on-demand team simulation. Using CAS as a mission representative of the need for effective tactical communication and coordination, we present a mission planning and rehearsal system that is augmented with a synthetic JTAC agent. This added capability allows commanders and aircrew to plan and fly through a CAS mission while communicating verbally with the synthetic JTAC.

By employing the knowledge encapsulated in an intelligent agent, we can overcome many of the challenges faced in human-computer dialogue, and continue to enrich synthetic training while migrating the benefits of this approach into the realms of mission planning and rehearsal.

## REFERENCES

- Bell, B., Johnston, J., Freeman, J., & Rody F. (2004). STRATA: DARWARS for Deployable, On-Demand Aircrew Training. In *Proceedings of the Interservice/Industry Training, Simulation, and Education Conference (IITSEC)*, December, 2004.
- Bell, B., Ryder, J., and Pratt, S. (2008). *Communications and coordination training with speech-interactive synthetic teammates: A design and evaluation case study*. In D. Vincenzi, J. Wise, P. Hancock and M. Mouloua (Eds.), *Human Factors in Simulation and Training*. CRC Press. Boca Raton.
- Black, A.W. and Taylor, P.A. (1997). The Festival Speech Synthesis System: System documentation. Technical Report HCRC/TR-83, Human Communication Research Centre, University of Edinburgh, Scotland, UK, 1997.
- Black, A.W. and Lenzo, K. (2000). Limited Domain Synthesis. In *Proceedings of the Sixth International Conference on Spoken Language Processing (ICSLP 2000)*. Beijing, China October 16-20, 2000.
- Chapman, R. J., Ryder, J., Bell, B., Wischusen, D. & Benton, D. (2004) STRATA (Synthetic Teammates for Real-time Anywhere Training and Assessment): An integration of cognitive models and virtual environments for scenario based training. In *Proc. of Human Factors & Ergonomics Society 49th Annual Mtg.* October, 2004, New Orleans, LA.
- Zachary, W., LeMentec, J.C., & Ryder, J. (1996) Interface agents in complex systems. In C. Ntuen & E. Park (Eds.), *Human Interaction with Complex Systems: Conceptual Principles and Design Practice*. Norwell, MA:Kluwer Academic Publishers, pp 35-52.
- Zachary, W., Ryder, J. M., and Hicinbothom, J. H. (2000). Building Cognitive Task Analyses and Models of a Decision-Making Team in a Complex Real-Time Environment. In Chipman, Shalin & Schraagen, Eds. *Cognitive Task Analysis*. New Jersey: Lawrence Erlbaum, 2000. 365-384.
- Zachary, W. Santarelli, T., Lyons, D., Bergondy, M. and Johnston, J. (2001). Using a Community of Intelligent Synthetic Entities to Support Operational Team Training. In *Proceedings of the Tenth Conference on Computer Generated Forces and Behavioral Representation*. Orlando: Institute for Simulation and Training. pp: 215-224.



# Modeling the Information Age Combat Model: An Agent-Based Simulation of Network Centric Operations

Sean Deller, Old Dominion University  
Ghaith A. Rabadi, Old Dominion University  
Michael I. Bell, Alidade Incorporated  
Shannon R. Bowling, Old Dominion University  
Andreas Tolk, Old Dominion University

**Abstract.** The Information Age Combat Model (IACM) was introduced by Cares in 2005 to contribute to the development of an understanding of the influence of connectivity on force effectiveness that can eventually lead to quantitative prediction and guidelines for design and employment. The structure of the IACM makes it clear that the Perron-Frobenius Eigenvalue is a quantifiable metric with which to measure the organization of a networked force. The results of recent experiments presented in Deller, et al., (2009) indicate that the value of the Perron-Frobenius Eigenvalue is a significant measurement of the performance of an Information Age combat force. This was accomplished through the innovative use of an agent-based simulation to model the IACM and represents an initial contribution towards a new generation of combat models that are net-centric instead of using the current platform-centric approach. This paper describes the intent, challenges, design, and initial results of this agent-based simulation model.

## Introduction

The Information Age Combat Model (IACM), recently introduced by Cares (2005), attempts to describe combat (or competition) between distributed, networked forces or organizations. The basic objects of this model are not platforms or other entities capable of independent action, but rather nodes that can perform elementary tasks (sense, decide, or influence) and links that connect these nodes. Information flow between the nodes is generally necessary for any useful activity to occur.

Once the IACM has been defined in terms of a network of nodes and links, the language and tools of graph theory (see, for example, Chartrand 1984) can be used for both description and analysis. A concise description of any graph is provided by the adjacency matrix  $A$ , in which the row and column indices represent the nodes, and the matrix elements are either one or zero according to the rule:  $A_{ij} = 1$ , if there exists a link from node  $i$  to node  $j$  and  $A_{ij} = 0$ , otherwise. One method used in studying the evolution of complex adaptive systems (chemical, biological, social, and economic) is calculation of the principal (maximum) eigenvalue of the adjacency matrix (Jain and Krishna, 1998). The existence of a real, positive principal eigenvalue of  $A_{ij}$  is guaranteed by the Perron-Frobenius theorem. This eigenvalue,  $\lambda_{PFE}$ , represents the ability of a network to produce feedback effects in general and combat power specifically in the case of the IACM.

The structure of the IACM makes it clear that the Perron-Frobenius Eigenvalue ( $\lambda_{PFE}$ ) is a quantifiable metric with which to measure the

organization of a networked force. The results of recent experiments presented in Deller, et al., (2009) indicate that the value of the  $\lambda_{PFE}$  is a significant measurement of the performance of an Information Age combat force. This paper describes the intent, challenges, design, and initial results of this agent-based simulation model.

## The Information Age Combat Model (IACM)

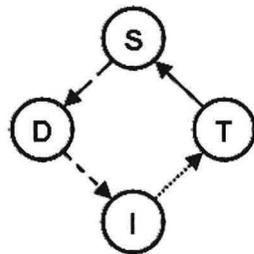
The IACM employs four types of nodes defined by the following properties:

- *Sensors* receive signals about observable phenomena from other nodes and send them to Deciders;
- *Deciders* receive information from Sensors and make decisions about the present and future arrangements of other nodes;
- *Influencers* receive directions from Deciders and interact with other nodes to affect the state of those nodes;
- *Targets* are nodes that have military value but are not Sensors, Deciders, or Influencers.

Each node belongs to a "side" in the competition, of which there are at least two. We will restrict the present discussion to two sides, conventionally termed BLUE (depicted in black in the figures) and RED (depicted in gray). In principle, any pair of nodes can interact, regardless of side, but some restrictions will be found to occur for both theoretical and practical reasons. It is worth noting that Influencers can act on any type of node, and Sensors can detect any type. The Target type was introduced primarily to reflect the fact that not all

military assets fall into one of the other three types. In most situations, however, an Influencer will target an adversary Sensor, Decider, or Influencer. The figures in this paper utilize the basic elements of graph theory.

The basic combat network shown in Figure 1 represents the simplest situation in which one side can influence another. The BLUE Sensor (S) detects the RED Target (T) and informs the BLUE Decider (D) of the contact. The Decider then instructs the BLUE influencer (I) to engage the Target. The Influencer initiates effects, such as exerting physical force, psychological or social influence, or other forms of influence on the target. The process may be repeated until the Decider determines that the desired effect has been achieved. It should be noted that the effect assessment requires sensing, which means that this will be conducted in a new circle. This most basic combat network is also referred to as a *combat cycle*.



**Figure 1.** The basic combat network represents the simplest situation in which one side can influence another.

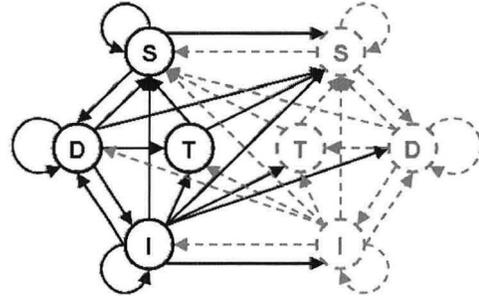
Each of the four links in Figure 1 is shown with a different type of line in order to emphasize the fact that the flows across these links may be very different. In particular, some links may represent purely physical interactions, while others may entail both physical processes and information flows. Two opposing combat cycles comprise the simplest two-sided combat network.

Cares (2005) described the simplest complete (two-sided) combat network as having 36 possible links. While the number of possible links for eight nodes (four each for BLUE and RED) is 64, we were able to exclude 28 and reduce that number to 36 based on the following important assumptions. The results are shown in Figure 2.

- Targets are passive; their only role is to be sensed and influenced. Therefore, 12 links from Targets to any nodes other than a Sensor were excluded.
- Sensors take no action; they provide information to Deciders and Sensors. Therefore, 10 links from Sensors to any

nodes other than a Sensor or own Decider were excluded.

- Deciders act only through Influencers but can be sensed. Therefore, 6 links from Deciders to any adversary nodes except a Sensor were excluded.



**Figure 2.** The simplest complete combat network represents all the ways in which Sensors, Deciders, Influencers and Targets interact meaningfully with each other.

When the BLUE/RED symmetry is taken into account, the number of link types is reduced to 18. These are listed in Table 1, where the nodes are identified as in Figure 2. Links between a node and itself in Figure 2 have been interpreted as connecting two different nodes of the same type and side.

**Table 1.** Types of links available in the IACM.

Link Type	From	To	Interpretation
1	$S_{BLUE}$ $S_{RED}$	$S_{BLUE}$ $S_{RED}$	S detecting own S, or S coordinating with own S
2	$S_{BLUE}$ $S_{RED}$	$D_{BLUE}$ $D_{RED}$	S reporting to own D
3	$S_{BLUE}$ $S_{RED}$	$S_{RED}$ $S_{BLUE}$	S detecting adversary S
4	$D_{BLUE}$ $D_{RED}$	$S_{BLUE}$ $S_{RED}$	S detecting own D, or D commanding own S
5	$D_{BLUE}$ $D_{RED}$	$D_{BLUE}$ $D_{RED}$	D commanding own D
6	$D_{BLUE}$ $D_{RED}$	$I_{BLUE}$ $I_{RED}$	D commanding own I
7	$D_{BLUE}$ $D_{RED}$	$T_{BLUE}$ $T_{RED}$	D commanding own T
8	$D_{BLUE}$	$S_{RED}$	S detecting adversary

	D <sub>RED</sub>	S <sub>BLUE</sub>	D
9	I <sub>BLUE</sub> I <sub>RED</sub>	S <sub>BLUE</sub> S <sub>RED</sub>	I attacking own S, or S detecting own I
10	I <sub>BLUE</sub> I <sub>RED</sub>	D <sub>BLUE</sub> D <sub>RED</sub>	I attacking own D, or I reporting to own D
11	I <sub>BLUE</sub> I <sub>RED</sub>	I <sub>BLUE</sub> I <sub>RED</sub>	I attacking own I, or I coordinating with own I
12	I <sub>BLUE</sub> I <sub>RED</sub>	T <sub>BLUE</sub> T <sub>RED</sub>	I attacking own T
13	I <sub>BLUE</sub> I <sub>RED</sub>	S <sub>RED</sub> S <sub>BLUE</sub>	I attacking adversary S, or S detecting adversary I
14	I <sub>BLUE</sub> I <sub>RED</sub>	D <sub>RED</sub> D <sub>BLUE</sub>	I attacking adversary D
15	I <sub>BLUE</sub> I <sub>RED</sub>	I <sub>RED</sub> I <sub>BLUE</sub>	I attacking adversary I
16	I <sub>BLUE</sub> I <sub>RED</sub>	T <sub>RED</sub> T <sub>BLUE</sub>	I attacking adversary T
17	T <sub>BLUE</sub> T <sub>RED</sub>	S <sub>BLUE</sub> S <sub>RED</sub>	S detecting own T
18	T <sub>BLUE</sub> T <sub>RED</sub>	S <sub>RED</sub> S <sub>BLUE</sub>	S detecting adversary T

The interpretation of some of the links (types 1, 4, 9, 10, 11, and 13 in Table 1) is ambiguous, and was recognized in the initial development of the IACM (Cares 2005). Overcoming this ambiguity was a necessary step in developing a simulation that would enable an analysis of the value of the  $\lambda_{PFE}$  as a significant measurement of the performance of an Information Age combat force. The simulations presented here are a step in this direction, since they employ only basic combat networks similar to Figure 1, but with the Target replaced by an adversary Sensor or Influencer. These combat cycles (Cares 2005) contain only links of types 2, 3, 6, 13, and 15. Of these, only type 13 is ambiguous.

#### A Basic Agent-Based Model Using the IACM

The structure of the IACM makes it clear that the  $\lambda_{PFE}$  is a quantifiable metric with which to measure the organization of a networked force, but is it an indicator of combat effectiveness? To determine this we constructed an agent-based simulation representation of the IACM and conducted a series of force-on-force engagements using opposing forces of equal assets and capabilities, but differing in their connectivity arrangements or configurations. These differences in connectivity

often, but not necessarily, lead to unequal  $\lambda_{PFE}$  values.

The agent-based paradigm was utilized for this purpose because the resulting models provide the ability to account for small unit organization, maneuver, and the networked effects that are the focus of our investigation. An additional advantage of utilizing an agent-based simulation was the ability to work around the ambiguities of link interpretation in the IACM. For example, instead of a mutually exclusive choice between defining a directional link from a BLUE Influencer to a RED Sensor (type 13 in Table 1) as either the Influencer "targeting" the Sensor or as the Sensor "sensing" the Influencer, both abilities can be represented in the agent-based simulation.

The first challenge in modeling the IACM concerned the adjacency matrix representation of the network. The IACM as originally described by Cares (2005) uses a single adjacency matrix to reflect the collective organization of both BLUE and RED forces. In this approach, the  $\lambda_{PFE}$  value is dependent on the configurations of both the BLUE and RED forces and might well represent the extent to which feedback effects occur in the engagement. Obviously, BLUE and RED each seek separately to maximize their own networked effects while minimizing those of the opposing force. This cannot be represented by a single  $\lambda_{PFE}$  value, so we calculate separate values ( $\lambda_{BLUE}$  and  $\lambda_{RED}$ ) to reflect the potential networked effects of the configurations of each of the opposing forces. These calculations required the adjacency matrices include a single Target node representative of all the enemy forces capable of being targeted. In other words, the values of  $\lambda_{BLUE}$  and  $\lambda_{RED}$  are determined solely by the arrangement of their respective assets, independent of the asset arrangement of the opposing force.

In order to best associate any difference in force effectiveness to the difference in connectivity, the opposing forces consisted of the same number of Sensors, Deciders, and Influencers, differing only in the manner in which they were arranged (i.e., linked). Since the potential value of a Sensor may not equal the potential value of an Influencer, the composition of each configuration considered in this work contained an equal number Sensors and Influencer to preclude any bias towards those configurations that have more of one or the other. Additionally, both types of nodes had identical performance capabilities (i.e., the sensing range was chosen equal to the influencing range, and the speeds of movement of the two types of node were equal).

In order to gain a "first order" understanding of the IACM, we made two key scoping decisions. First, each Sensor and Influencer would only be

connected to one Decider (but any given Decider could be connected to multiple Sensors and Influencers). Second, the connectivity within any force was limited to only those “vertical” links necessary to create combat cycles (i.e., link types 2, 3, 6, 13, and 15 in Table 1), which are the essence of the  $\lambda_{PFE}$  (the most basic element of the IACM).

### Developing the NetLogo Model

The agent-based simulation environment utilized for this research was NetLogo (Wilenski 1999). The code of the agent-based model closely follows the logic of the IACM, with a few notable exceptions. Agents served as Sensors, Deciders, and Influencers, but Targets were not included as they served no purpose other than to absorb losses. Given that this work represents a “first cut” effort, including Target agents with no detect, direct, or influence capabilities would only serve to clutter the results.

Additionally, Deciders cannot be destroyed in the present model. This was done in recognition of their unique role in connecting multiple Sensors and Influencers. Destruction of a Decider typically renders a number of other nodes useless (effectively destroyed), making it a particularly high value target. Since targets are detected and engaged in random order in our model, we wished to give all targets equal value in order not to generate atypical engagements that might bias the results.

The agent rules sets, themselves, function in accordance with the IACM. Sensors detect enemy nodes within the sensing range parameter, and communicate that information to their assigned (connected) Deciders. Deciders communicate the sensing information to their assigned Influencers. Influencers destroy the nearest enemy node that is both “sensed” by a Sensor connected to that Influencer’s Decider, and within the influencing range parameter. Deciders direct Sensor movement towards areas of suspected enemy nodes. Deciders direct Influencers to move towards the nearest “sensed” enemy node. All nodes are assumed to perform their functions perfectly and instantaneously. Agent interactions are deterministic, i.e., the probabilities of detect, communicate and kill are all “1”. A stochastic dimension to the model can be built once a better understanding of the research questions is gained, and this new dimension can be used to model errors and delays representing technological and human performance factors. Most importantly, the rules sets and parameter values for both BLUE and RED agents were identical.

Each agent in the model is defined as a part of an agentset (i.e., “breed”) associated with a particular Decider. Since the nodes of the IACM are generic, the most important defining characteristic of any agent is its connectivity. For example, all BLUE Sensors and Influencers connected to the BLUE Decider<sub>1</sub> are established by the following breeds:

```
breed [ BInfluencer1s BInfluencer1 ]
breed [ BSensor1s BSensor1 ]
```

The actual numbers of agents within these breeds will vary according to the configuration being tested. Sliders were utilized for this purpose, thereby enabling the BehaviorSpace feature to vary the configurations automatically and allowing us to execute the large number of engagements necessary to complete this research. The BLUE Decider<sub>1</sub> itself is also defined as a breed, but consists only of just that single agent. Similar agents for all other BLUE and RED Sensors, Influencers and Deciders were established.

The connectivity between these breeds represents the combat cycle links of the IACM (specifically link types 2, 3, 6, 13 and 15 as explained in Table 1). Link types 2 (“detection”), 6 (“order”), 13 (“LOF”)<sup>1</sup> and 15 (“LOF”) are defined in the simulation by the `directed-link-breed` keyword.

As mentioned earlier in this paper, link type 13 has an ambiguous meaning in the IACM. The `directed-link-breed` keyword defines the Influencer-to-Sensor link as the Influencer attacking an enemy Sensor. Both link type 3 and the other IACM interpretation of link type 13 (i.e., a Sensor detecting an adversary Influencer) will be defined by the sense procedure later in the code. Finally, all agents within each breed have certain variables that are tracked during the simulation, such as `side` (i.e., BLUE or RED), `dead` (i.e., agents that are attacked by an opposing Influencer may no longer act), and `sensed` (i.e., at any given tick count within the simulation an agent may be within sensing range of one or more opposing Sensors).

Given the large number of engagements within this experiment, it was imperative to utilize the BehaviorSpace feature of NetLogo. To enable this, each of the different force configurations were defined by using the `set` command to establish the appropriate numbers of Sensors and Influencers for each of the BLUE and RED Deciders. For example, BLUE Configuration (i.e., “BID”) #0 assigned 5 Sensors and 5 Influencers to

---

<sup>1</sup> LOF is an acronym for “line of fire,” which is a direct horizontal line from a firing weapon to its target.

BLUE Decider<sub>1</sub>, and one of each to the other 4 Deciders:

```
if BID = 0 [set Bconfig [5 1 1 1 1 5 1 1 1 1]]
if BID = 2 [set Bconfig [5 1 1 1 1 1 4 2 1 1]]
  set number-BSensor1s item 0 Bconfig
  set number-BSensor2s item 1 Bconfig
  set number-BSensor3s item 2 Bconfig
  set number-BSensor4s item 3 Bconfig
  set number-BSensor5s item 4 Bconfig
  set number-BInfluencer1s item 5 Bconfig
  set number-BInfluencer2s item 6 Bconfig
  set number-BInfluencer3s item 7 Bconfig
  set number-BInfluencer4s item 8 Bconfig
  set number-BInfluencer5s item 9 Bconfig
```

BLUE Configuration #2 is nearly identical, differing only in one link. Decider<sub>1</sub> now only has 4 assigned Influencers while Decider<sub>2</sub> now has 2. The movement of a single link is not trivial as it may have a significant impact on both the  $\lambda_{PFE}$  value and the average probability of Win for that particular configuration. All different force configurations were established in this manner, thereby allowing the BehaviorSpace feature to automatically cycle through all possible engagements between the BLUE and RED configurations instead of running the simulation one engagement at a time.

Since the focus of this effort is to gain insight into the relationship between the  $\lambda_{PFE}$  value and the effectiveness of a networked force, the agent-based simulation rules of engagement were quite simple. The battlespace (i.e., "world") within the model is deliberately featureless in order to focus on the configurations themselves. The agents are randomly distributed across the battlespace at the beginning of each engagement. Engagements continued until either all of the Sensors and Influencers of one force were annihilated, or both forces were incapable of continued combat (i.e., neither side contained a functioning combat cycle). A single run of the agent-based model will result in a BLUE win, a RED win, or an undecided result.

During each time tick of the simulation, the following procedures are executed: establish-links, sense, track, shoot, kill, move-Influencer, move-Sensor, and reset. The establish-links procedure establishes the links defined by the directed-link-breed keyword earlier in the code. It does so by breed, thereby ensuring each Sensor and Influencer is connected to only one Decider.

```
to establish-links
  ask BDecider1s [
    ask BSensor1s [create-detection-to
  myself [set color blue] ]
    ask BInfluencer1s [create-order-from
  myself [set color blue] ] ]
```

At this time, two of the four necessary links (types 2 and 6) of the IACM combat cycle have been established in the simulation. Link type 3 and one

of the two interpretations of link type 13 are established in the sense procedure. In this procedure, every Decider asks its assigned Sensors (i.e., "in-link-neighbors") to identify all adversary Sensors and Influencers within its sensing range (i.e., "s-range"). Upon identification, the specific sensed variable of the targeted agent for that particular opposing Decider is set to a value of "1." The s-range parameter remains constant for all Sensors, either BLUE or RED, over time. The sense procedure depicted below includes all opposing breeds (only RInfluencer1s is shown in this example) and is repeated for every BLUE and RED Decider breed:

```
to sense
  ask BDecider1s [
    ask in-link-neighbors [
      ask RInfluencer1s in-radius s-range [set
  sensedBD1 1]
```

The remaining links necessary to complete the IACM combat cycles (link type 15 and the alternate interpretation of link type 13) are established by the track, shoot, and kill procedures. During the track procedure, every Decider asks its assigned Influencers (i.e., "out-link-neighbors") to identify all adversary Sensors and Influencers within its influencing range (i.e., "i-range"). Upon identification, the targeted agent is linked to that particular Influencer using the create-LOF-from-myself keyword. The i-range parameter remains constant for all Influencers, either BLUE or RED, over time. The track procedure depicted below includes all opposing breeds (only RInfluencer1s is shown in this example) and is repeated for every BLUE and RED Decider breed:

```
to track
  ask BDecider1s [
    ask out-link-neighbors [
      ask RInfluencer1s in-radius i-range
  [create-LOF-from myself]
```

Now that the complete IACM combat cycle has been established, the shoot and kill procedures represent its execution. During this procedure, each Decider directs its assigned Influencers to identify the single closest opposing Sensor or Influencer with which it shares a LOF link. This limits all Influencers to the same rate of fire of one targeted node per time tick. Identification is portrayed by setting the dead variable equal to "1."

```
to shoot
  ask BDecider1s [
    ask out-link-neighbors [
      ask out-link-neighbors [
        let $targets-sensed turtles with
  [(sensedBD1 = 1) and (side = 2)]
        if any? $targets-sensed [
```

```

    ask min-one-of $targets-sensed
[distance myself] [set dead 1] ] ] ] ]

```

Following this identification, the `kill` procedure deletes all agents that have been “sensed,” “tracked” and “shot.” The purpose of separating the `kill` procedure from the `shoot` procedure is to allow simultaneous shots, thereby precluding any advantage that would be gained by the order of execution of the `shoot` procedure code.

```

to kill
  ask turtles with [(dead = 1)] [die]
end

```

The collective effect of the `sense`, `track`, `shoot`, and `kill` procedures is to require that a Sensor and an Influencer must be assigned to the same Decider and within their respective `s-range` and `i-range` in order to successfully complete a combat cycle (i.e., eliminate the targeted node).

Upon completion of all combat cycle execution, all remaining Sensors and Influencers are moved. The `move-Influencer` procedure directs all Influencers to move towards the nearest opposing Sensor or Influencer that has been sensed by a friendly Sensor assigned to the same Decider. If there are no qualifying opposing Sensors or influencers, then the Influencer will not move. Each time `tick` includes five iterative moves of a distance of “1” that are sequential between Deciders and sides in order to preclude any advantage of moving first or last. An example iteration for one Decider follows below:

```

to move-Influencer
  ask BDecider1s [
    ask out-link-neighbors [
      let $targets-sensed turtles with
[(sensedBD1 = 1) and (side = 2)]
      if any? $targets-sensed [
        set heading towards min-one-of
$targets-sensed [distance myself] forward 1 ]
    ] ]

```

The `move-Sensor` procedure directs all Sensors to move towards the nearest Sensor or Influencer that is not currently sensed by a friendly Sensor assigned to the same Decider. This procedure is necessary to enable both sides to eventually target those opposing Sensors and Influencers that did not start the simulation within any friendly Sensor’s `s-range`.

```

to move-Sensor
  ask BDecider1s [
    ask in-link-neighbors [
      let $targets-sensed turtles with
[(sensedBD1 =
1) and (side = 2)]
      if not any? $targets-sensed [
        let $targets-unsensed turtles with
[(sensedBD1 = 0) and (side =
2)]
        if any? $targets-unsensed [
          let $nearest-unsensed min-one-
of $targets-unsensed [distance myself]

```

```

    set heading towards $nearest-
unsensed forward 1 ] ] ] ]

```

The final procedure during each time `tick` is `reset`. During this procedure, all sensed variables are reset to “0” and all links, to include the LOF “tracking” links, are deleted in preparation for the `establish-links`, `sense`, `track`, `shoot`, `kill`, `move-Influencer`, `move-Sensor`, and `reset` procedures for the next time `tick`.

## Initial Results

The initial experiment consisted of all possible engagements of the 42 different configurations of two networked forces (BLUE and RED), each containing 7 Sensors, 3 Deciders, 7 Influencers, and 1 Target. The sole Target node is representative of all the opposing nodes vulnerable to destruction. Additionally, the capabilities for each of these node types were identical between the forces. A comprehensive test of each of these 42 configurations against each other required 1,764 different engagements. Each engagement was represented by 30 replications, each with a random distribution of the BLUE and RED nodes across the battlespace.

The results clearly show that the probability of a BLUE win increases for those BLUE configurations with a greater  $\lambda_{PFE}$  value. A simple linear regression confirmed this with a coefficient of determination ( $R^2$ ) equal to 0.896 for the following equation:

$$y = 1.0162(x) - 1.5780$$

where,  $y$  = the average probability of a BLUE win for that configuration

$x$  = the  $\lambda_{PFE}$  value of a configuration

## Summary & Conclusion

The agent-based simulation described in this paper was employed to analyze the value of the  $\lambda_{PFE}$  as a quantifiable metric with which to measure the organization of a networked force. This simulation was specifically designed to overcome the challenges of the IACM link ambiguity and the large number of engagements necessary to complete the research. The results of recent experiments presented in Deller, et al., (2009) indicate that the value of the Perron-Frobenius Eigenvalue is a significant measurement of the performance of an Information Age combat force.

## References

Deller, Sean, Bell, Michael I., Rabadi, Ghaith A., Bowling, Shannon R., and Tolk, Andreas (2009). Applying the Information Age Combat Model:

Quantitative Analysis of Network Centric Operations. Accepted for publication in *The International C2 Journal*.

Cares, Jeff. 2005. *Distributed Networked Operations*. iUniverse, New York.

Chartrand, Gary. *Introductory Graph Theory*. Dover Publications, 1984.

Jain, Sanjay, and Krishna, Sandeep. 1998. Autocatalytic Sets and the Growth of Complexity in an Evolutionary Model. *Physical Review Letters* 81: 5684-5687.

Jain, Sanjay, and Krishna, Sandeep. 1999. Emergence and growth of complex networks in adaptive systems. *Computer Physics Communications* 121-122: 116-121.

Wilenski, Uri. 1999. *NetLogo*. Center for Connected Learning and Computer-based Modeling, Northwestern University, Evanston, IL. <http://ccl.northwestern.edu/NetLogo>.



# Low- Cost Rescue Robot for Disaster Management In a Developing Country – Development of a Prototype Using Locally Available Technology

**Mahmud, Faisal. ; Hossain, S.G.M.<sup>1</sup>; Alam, Jobair Bin.<sup>2</sup>**

*Civil and Environmental Engineering Department, Old Dominion University, Norfolk, Virginia, USA*

*Department of Mechanical Engineering<sup>1</sup>, University of Nebraska at Lincoln, Nebraska, USA*

*Professor<sup>2</sup>, Department of Civil Engineering, Bangladesh University of Engineering & Technology (BUET),  
Dhaka, Bangladesh*

*E-Mail: [fmahm001@odu.edu](mailto:fmahm001@odu.edu), [smgmamur@huskers.unl.edu](mailto:smgmamur@huskers.unl.edu)<sup>1</sup>*

**Abstract.** The use of robots in different fields is common and effective in developed countries. In case of incident management or emergency rescue after a disaster, robots are often used to lessen the human effort where it is either impossible or life-threatening for rescuers. Though developed countries can afford robotic-effort for pro-disaster management, the scenario is totally opposite for developing and under-developed countries to engage such a machine-help due to high cost of the machines and high maintenance cost as well. In this research paper, the authors proposed a low-cost “Rescue-Robot” for pro-disaster management which can overcome the budget-constraints as well as fully capable of rescue purposes for incident management. Here, all the research works were performed in Bangladesh – a developing country in South Asia. A disaster struck structure was chosen and a thorough survey was performed to understand the real-life environment for the prototype. The prototype was developed considering the results of this survey and it was manufactured using all locally available components and facilities.

Keywords: Rescue-robot, pro-disaster management, cost-effectiveness.

## INTRODUCTION

Disaster causes can be classified into two main groups-natural and man-made. In the past, pro-disaster rescue work was performed by human where machine assistance was not effectively utilized as it was not improved enough. As the time passed, the unprecedented number and scales of natural and human-induced disasters has urged the emergency search and rescue community around the world to seek for newer, more effective equipment to enhance their efficiency. Rely on search dogs, camera mounted probes and human help is still considered as the assisting tools and technology for rescue as well as search purposes. But, with the advancement of science and technology, intelligent robots (i-robots) equipped with advanced sensors and detectors are attracting more and more attentions from researchers and rescuers. The developed countries though achieved the technology and went on production for rescue

robots, in case of developing or underdeveloped countries, this advanced and intelligent assisting tool remains a dream to-date. The reason is mainly for the budget constraints- without the money, no research work is possible, and without the research work, no innovation is capable-ultimately the aim for rescue robot as a mass production becomes a latent hope. In this paper, the researchers tried to illustrate how to develop the technology of a rescue robot within a limited budget or low-cost in a developing country as well as effectiveness of reusable materials in making this rescue robot.

## Background of the Research

Natural and human-induced disasters always cost in terms of money or lives. Natural disasters in recent years, such as the Hurricane Katrina and Rita in 2005, China earthquake in the capital of Sichuan in 2008, Cyclone Nargis in Burma and Bangladesh in 2008 and so others throughout the

world claimed deadly and costly tolls to the affected communities [1]. Human-induced disasters in the form of civil-war, terrorist attacks etc. also have direct casualties not fewer than the natural disasters. Collapsed buildings are common field environment for humanitarian search and rescue operations. Earthquakes, typhoons, tornados, weaponry destructions, and catastrophic explosions can all generate damaged buildings in large scales. The use of heavy machinery in such incidents is prohibited because they would destabilize the structure, risking the lives of rescuers and victims buried in the rubble [2].



Figure 1: Actual site after the building collapsed in Dhaka, Bangladesh.

Rescue specialists use trained search dogs, cameras and listening devices to search for victims from above ground. Though search dogs are effective in finding human underground, they are unable to provide a general description of the physical environment the victim locates. Camera mounted probes can provide search specialists a visual image beyond voids that dogs can navigate through, however their effective range is no more than 4-6 meters along a straight line below ground surface [2]. Robot assisted search and rescue systems though took much attention from different perspective but the actual field work was started from 2001. The first real research on search and rescue robot began in the aftermath of the Oklahoma City bombing in 1995 [3]. Robots were not used at the bombing response, but suggestions

as to how robots might have been applied were taken. In 2001, the first documented use of urban search and rescue robots took place during the 9/11 World Trade Center (WTC) disaster where mobile robots of different sizes and capacities were deployed. These robots range from tethered to wireless operated, and from the size of a lunch box to the size of a lawnmower [4].

The goal of this research project was to build a prototype of a robotic vehicle capable of working in the rescue process of some natural as well as human made disasters those generally occur in Bangladesh. According to the OFDA/CRED International Disaster Database, from 1942 to 1991, seven major wind storms caused the death of 570,413 people in Bangladesh [11]. Although earthquake did not claim life toll as wind storms but Ali et al. [12] explains that, the historical seismicity data of Bangladesh and adjoining areas indicate that Bangladesh is vulnerable to earthquake hazards. As Bangladesh is the world's most densely populated area, any future earthquake shall affect more people by unit area than any other seismically active regions of the world. So, taking into account principally these types of disasters where there may be destruction of structures, that may lead to confinement of human inside the debris, a disaster affected building was chosen to perform a survey on the environment that the prototype would have to face. Then the design procedure was performed basically concerning of the use of locally available and cost effective product and technology.

## LITERATURE REVIEW

Hi-tech robots and related works have been done for last couple of years, but robots with high efficiency within a low cost budget may not gain that type of attention. In Johns Hopkins University, Baltimore, Maryland, four undergraduate engineering students designed and built a remote controlled robotic vehicle to find deadly land mines in rugged terrain and mark their location with a spray of paint. The student spent about \$5000 to design and build their prototype. They estimated the vehicle could be mass-produced for \$1,000 or less, not including the cost of more sophisticated

detection sensors [6]. Educational robot like The Trikebot was claimed as a ground-up design effort chartered to develop an effective and low-cost educational robot for secondary level education and home use with its chassis to be produced economically (approximately \$500 per chassis) by its inventors [7]. In India a team of students from the Sree Chithra College of Engineering, Pappanamcode, developed a low-cost working model of robot called "RASOR" capable of functioning in domestic and industrial environments. Though the team would not like to reveal the exact production cost of RASOR, "for marketing reasons", they said that making a RASOR would take less than Rs. 30,000 (approximately \$615 USD) [8]. Albert Ko and Henry Y. K. Lau had worked for the low-cost autonomous robotic search and rescue system to design and cooperate in large quantity to search for survivors in rubbles. These robots were equipped with wireless communication module to facilitate data and video/audio transfer. These wireless robots, with no tethers, could navigate freely in obstructed environment but were difficult to track their locations once they wandered out of the operators' sights [2]. A team from the University of New South Wales, Sydney, Australia had built several "Redback" robots, each for a cost of approximately \$4,500 USD including the cost of the onboard PC, batteries, communications and sensors [9]. Utilizing off-the-shelf low cost parts, Mundhenk et al. had constructed a robot that was small, light, powerful and relatively inexpensive (< \$3900) [10]. From all these previous works, it is found that robots making with hi-technology and performance as well as low-cost budget sometimes cause hindrance to reach the goal. Our effort in this respect was to overcome all the previous failures as much as possible and to utilize reusable materials so as to negotiate with the budget limits.

## METHODOLOGY

### Field Work and Site Survey

On December 8, 2007, a 22 story high rise building (called RANGS Bhaban) in Dhaka, the capital of Bangladesh, collapsed while it was going on for demolition to make a link road. The casualties numbered more than 22. Dhaka, the capital of Bangladesh, is now one of the congested and

mostly populated cities in the world with the population density 14,608/km<sup>2</sup>[5]. With the population increase and urbanization, buildings are built for reasons, but in most cases without following proper guidelines and building codes. As a result, if a medium intensity of earthquake once strike Dhaka, the scenario will be disastrous. Keeping this fact in mind, a research work was carried out for the first time to make a machine help (hereinafter called rescue robot) for pro-disaster management and rescue purposes. The collapsed building site was investigated for getting the real life scenario after the disaster and also for the design purpose of the rescue robot to overcome all type of hindrance for rescue work.



Figure 2: Comparison of the gravel size on site with a pen.

The concrete structure building's 14<sup>th</sup> floor collapsed due to improper demolition methods which caused the collapse of all the floors up to 4<sup>th</sup> floor and took lives of several workers. The condition of the building was considered as quasi similar to a natural disaster struck building. A place was chosen on the 6<sup>th</sup> floor that had the possibility of having someone trapped in. The survey results showed that for a beeline course for the robot, on an area having the length of 3 meters and width of 1 meter, the average length of concrete blocks (the length being in the same direction of the robot's path) came to be as 198 mm. In this measurement, concrete blocks having length of 50 mm to 300 mm were considered as potential obstacles for the robot. Blocks with length less than 50 mm were

considered as too easy to overcome while those more than 300 mm were considered as too large that the robot could accommodate on it.

Similarly, the average height of the obstacles came as 106 mm and the average slope of the obstacle was measured as 60 degrees, that is, the robot would have to be capable of moving on the 60 degree slope to overcome that certain block.

These data were necessary to identify the size of the robot's wheels and also the overall dimensions of the total robot. This also helped to find the required torque for the drive motors.

### Search for Local Components

After the completion of the robot's mechanical component design, a thorough investigation was performed on the availability of those components in the local market. The researchers had to rely on the reusable components wherever possible to imply. For example, aluminum chips collected from the lathe machine refusals were used to prepare the wheels of the robot. But because of the presence of too much slug, they did not appear to be a good choice for casting.



Figure 3: Aluminum chips for robot wheel.

Then aluminum alloy automobile engine cylinders were melted and casted which showed very good performance for the casting. These cylinders were collected from the junkyards. Custom tires were prepared from heavy duty timing belts. This would help the robot to move about in a very rough terrain.



Figure 4: Robot wheel before finishing.

## DESIGN

### Design for Work Environment

The work environment for a rescue robot differs from the conventional robots principally because of the rough terrain caused by the debris. This is similar to the international rescue robots competitions which require robots to negotiate complex and collapsed structures, find simulated victims, and generate human readable maps of the environment [13]. The wheel size was optimized considering the torque supplied by the motors and the average height of the obstacles obtained from the field survey. Four shock absorbers were made using metal spiral springs and steel plates. These were provided to minimize the shock created on the robot's body as well as the effect of a dislocated center of gravity that might have caused the risk of falling down of the robot while moving through an inclined surface or passing a large obstacle. Triangular brackets were provided in the chassis of the robot to induce extra strength to the structure. But aluminum was used wherever possible to keep the total weight of the robot down which is important for economic power consumption.

High torque worm-wheel motors (from surplus store) were used as drive and these were coupled to rear wheels. Each of the rear wheels was coupled to the front wheel by a chain-sprocket system to enhance the torque and work as a four wheel drive vehicle. This would help in its vigorous

movement through the rough terrain and also to overcome obstacles.



Figure 5: Spring plate to resist bumping.

The tires of the wheels were made from heavy duty timing belts which had very good grip to move on an inclined surface. These had an additional property of heat resistance up to such a level which may help this robot perform rescue works in a hot, fire affected zone.

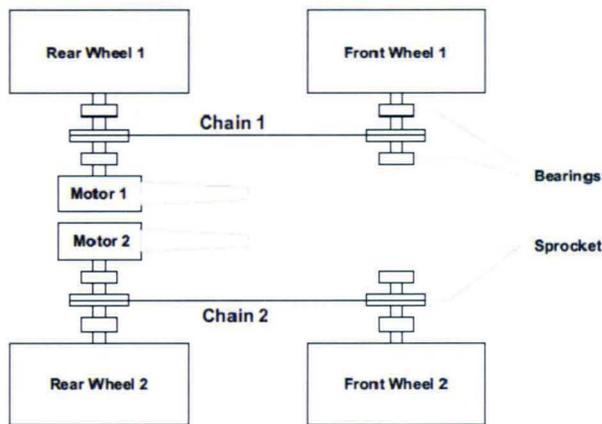


Figure 6: Schematic of the Robot's drive system.

### Design for Manufacturability and Assembly

Some of the Design for Manufacturability (DFM) and Design for Assembly (DFA) methods were followed even though this was a prototype. This would help in the manufacturing and assembly of several robots if the prototype would work successfully. These techniques would in turns help reduce the labor and capital cost and also to keep the cost of the robot down.

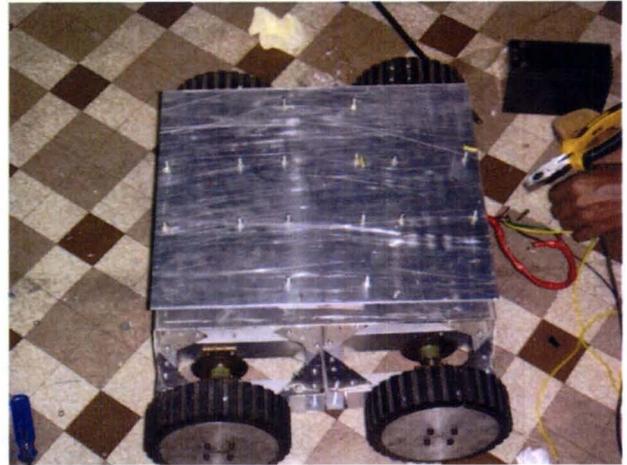


Figure 7: Rescuemate PDA (official name of the rescue robot).

Standardized parts were designed and used wherever possible. For example, standard sized screws were used in most of the parts of the assembly. The brackets and the aluminum channel stocks were interchangeable parts.

The whole chassis of the robot was composed of four symmetric and interchangeable sub-assemblies. Most of the screws were inserted as a top down manner. But in some cases the DFM and DFA were not possible to follow such as, parts count was tried best to minimize but to sacrifice for interchangeable and symmetric subassemblies, it increased to some extent. Secondary operations on the wheels were performed by machining the rough casted surface which also goes against DFM but this was important to reduce the material cost. If the wheels were manufactured from a large shaft then it would not only induce high cost for the material but also a large amount of aluminum would be wasted.

### COST ANALYSIS

The cost can be divided into three broad categories as the cost for mechanical components, cost of manufacturing and the cost of the electronics and programming components.

#### Mechanical Components

These components shared a cost of around \$550. The motors were collected from surplus stores which were still in very good working condition. These were power window motors of automobiles.

The chassis was made of mainly aluminum and these components were collected from the hardware stores which incurred a higher price. But the wheels were custom made from rejected aluminum alloy components from a junkyard. The chain sprockets were scrapped ship components.

### **Cost of Manufacturing**

Most of the manufacturing works were performed in the university facility which included lathe and shaper machine operations, foundry and welding operations etc. and these were not counted to the cost of manufacturing. But for some special operations such as delicate welding or machining, local machine shops and the expert help was sought which took about only \$100. This was possible as labor cost is very low in Bangladesh compared to any first world country.

### **Electronic Components**

These components had two extremes. Some components were prepared in the lab as the simple circuit boards and wirings. But for some components the estimated cost was sacrificed for the high-tech components such as camera for the robot vision, high intensity search lights, microcontroller circuit boards, and radio frequency generating and receiving circuits giving clean signals. Although the robot's test run was performed by using very simple motor control circuit, these components were essential for the full fledged working for the rescue robot and so these costs were included in the estimation. These components shared a price of about \$800.

Analyzing the costs, it can be concluded that the manufacturing cost and the cost for mechanical components were successfully kept low without affecting the quality much. The total cost including sophisticated sensor parts and some other costs were no more than \$2,000.

### **RESULTS**

From this research work, it is found that robotic assistance is quite important in Bangladesh for pro-disaster situation. In this three phase research

work, the authors demonstrated the first phase whereas the second and third phases were still going on. The first phase was that a robot will be able to overcome all the difficulties to run in a disaster strike site where rubbles and other obstacles were by produced. In the second phase, it will be able to identify any live person/ living animal by its audio-visual system. The last and ultimate part is to attach a robotic arm with the vehicle so that it can assist the trapped person to rescue. The robotic arm has already prepared in the BUET mechanical engineering lab, but the on-site test is still to run. With a very low budget, the researchers/ authors were not only able to meet the research objectives but also made it possible how to use the scrap goods/materials to make a new robot. From this perspective, we can call this robot an "Environmental Friendly" rescue robot. During the initial test run of the robot, it showed good performance to overcome moderate sized discrete obstacles (maximum height being 55mm and on a plane surface having a slope up to 40°). But these performances need to be improved to achieve the goals set by the initial survey.

### **Limitations**

There were some limitations like all other research works which the authors accept as of short funding and allocation of proper money in this project. The research work was carried out for only six months by the authors as they had to come to the USA for higher studies. But, before coming, they were able to complete the first phase which already have described. The micro-controller system was not completed for the first run of this robot. The robotic rescue arm was made by this time. The authors expect to contribute this kind of research work in the developing countries where the governments face budget constraints to do research work for making robotic rescue effort.

### **CONCLUSION AND FUTURE WORKS**

Within a limited low-cost budget and reusable materials using, the Rescuemate PDA showed its effectiveness to match the research goals. For field test purpose, wire-connected experiments were

done, but micro-processor control system was the research goal which is still under research options. Currently some works are going on for making it more affordable to the local agencies in Bangladesh in case of emergency rescue works. This was the first effort in Bangladesh to make a rescue robot from reusable materials and scrap goods. It is true that sometimes mass production can make anything available with low-cost, but the researchers at least tried to meet the need of rescue works by robotic effort within a low-budget in Bangladesh. In case of mass production, this type of robot can be built within \$1000. In future, this current rescue version of robot in Bangladesh will get more rescue capability and preciseness because of the "Central Robotic Research (CRR)", the authors would like to contribute to this goal.

#### ACKNOWLEDGEMENT

The authors would like to thank Dr. Jobair Bin Alam, Professor, Department of Civil Engineering, Bangladesh University of Engineering and Technology (BUET), Bangladesh for his generous support and inspiration as well as giving the opportunity for this research work. Both the authors were research assistants of Dr. Jobair Bin Alam at the time of this research. This research work was solely carried out for BUET and Bangladesh Government. All the photographs included in this paper were taken by the researchers from real site investigation.

#### REFERENCES

[1] <http://www.guardian.co.uk/world/natural-disasters>.

[2] Albert Ko and Henry Y. K. Lau. Robot Assisted Emergency Search and Rescue System, International Journal of Advanced Science and Technology Vol. 3, February, 2009.

[3] Murphy, R. R. (2004). " National Science Foundation Summer Field Institute for rescue robots for research and response (R4)". AI Magazine. Vol. 25, issue 2: pp.133-136.

[4] Snyder, R. (2001). "Robots assist in search and rescue efforts at WTC," IEEE Robot. Automation Magazine, vol. 8, pp. 26-28.

[5] <http://archive.wn.com/2009/07/10/1400/dhaka/index6.html>

[6] <http://www.scienceblog.com/cms/node/2947>. Retrieved on 06/07/2009.

[7] Hsiu, T and et al. (2002). "Designing a Low-cost, Expressive Educational Robot," The Robotics Institute, Carnegie Mellon University, Pittsburgh, PA.

[8] <http://www.hinduonnet.com/2004/07/04/stories/2004070408280300.htm>. The Hindu, online edition of India's National Newspaper, published on July 4th, 2004.

[9] Raymond Sheh. "The Redback: A Low-Cost Advanced Mobility Robot for Education and Research".

[10] Mundhenk et al. "Low cost, high performance robot design utilizing off-the-shelf parts and the Beowulf concept, The Beobot project."

[11] EM-DAT: The OFDA/CRED International Disaster Database, Université Catholique de Louvain, Brussels, Belgium (05/01/2003).

[12] Earthquake Database and Seismic Zoning of Bangladesh, Md. Hossain Ali, Bangladesh University of Engineering and Technology, INCEDE Report, February 1998, Bangkok.

[13] Test Arenas and Performance Metrics for Urban Search and Rescue Robots Adam Jacoff, Elena Messina, Brian A. Weiss, Satoshi Tadokoro, and Yuki Nakagawa; Proceedings of the 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems, Las Vegas, NV, October 27-31, 2003.



# ROUTE GENERATION FOR A SYNTHETIC CHARACTER (BOT) USING A PARTIAL OR INCOMPLETE KNOWLEDGE ROUTE GENERATION ALGORITHM IN UT2004

Gregg T. Hanold, Technical Manager

*Oracle National Security Group*

*gregg.hanold@oracle.com*

David T. Hanold, Animator

*DavidHanoldCG*

*dhanold@comcast.net*

**Abstract.** This paper presents a new Route Generation Algorithm that accurately and realistically represents human route planning and navigation for Military Operations in Urban Terrain (MOUT). The accuracy of this algorithm in representing human behavior is measured using the Unreal Tournament™ 2004 (UT2004) Game Engine to provide the simulation environment in which the differences between the routes taken by the human player and those of a Synthetic Agent (BOT) executing the A-star algorithm and the new Route Generation Algorithm can be compared. The new Route Generation Algorithm computes the BOT route based on partial or incomplete knowledge received from the UT2004 game engine during game play. To allow BOT navigation to occur continuously throughout the game play with incomplete knowledge of the terrain, a spatial network model of the UT2004 MOUT terrain is captured and stored in an Oracle 11g Spatial Data Object (SDO). The SDO allows a partial data query to be executed to generate continuous route updates based on the terrain knowledge, and stored dynamic BOT, Player and environmental parameters returned by the query. The partial data query permits the dynamic adjustment of the planned routes by the Route Generation Algorithm based on the current state of the environment during a simulation. The dynamic nature of this algorithm more accurately allows the BOT to mimic the routes taken by the human executing under the same conditions thereby improving the realism of the BOT in a MOUT simulation environment.

## 1. INTRODUCTION

Research on Human Behavior Representation (HBR) in synthetic agents (BOTS) has focused predominantly on Cognitive Modeling. A Cognitive Model attempts to represent human thinking or decision making and translate that to human action. The corollary to Cognitive Modeling is Behavior Generation, which we have defined as the representation of human behavior that mimics or emulates the human. We have found very little research which examines Human Behavior Modeling (HBM) in this context [2], [3].

The research indicates that a system today can be smart enough to give the illusion of life by concentrating on creating consistent believable high level behavior instead of natural looking human actions [1], [5]. The UnrealTournament™ 2004 (UT2004) game engine selected for this research, for example, provides an interface to the physics and AI components that generate the BOT behavior and actions that are used to implement the new route planning and navigation algorithm. The high level behavior elements received from the game engine form the inputs to the algorithm that plans and generates the routes that the BOT executes. BOT execution of the routes includes dynamic behavioral actions based on sensory information to better mimic the human thereby giving this illusion of life or realism as we

have defined it. To further improve the realism of BOT route planning and navigation we introduce the concept of using partial or incomplete knowledge of the environment. This concept results in the dynamic calculation of routes based on sensory information and behavioral actions that more closely mimic those available to humans executing similar actions. To be realistic the behavior must mimic that of the human.

The simulation of Human Behavior for the purpose of measuring realism requires a virtual environment that can closely resemble that of the real world. The game industry has successfully achieved this goal with the Massively Multiplayer Online (MMO) games and First Person Shooter (FPS) Games such as Quake III Arena™, Half-Life2™, and Unreal Tournament's Americas Army™ Mod. As previously indicated, in this paper we develop a new route generation algorithm that is executed using Unreal Tournament 2004 (UT2004). The simulation of this new algorithm is accomplished using the Gamebots 2004 (GB2004) UnrealScript package, the Pogamut BOT (agent) and Java Libraries and its Netbeans plug-in, and an Oracle 11g spatially enabled database.

## 2. Background

Generating realistic human behavior in a virtual environment continues to challenge the simulation community. In recent years the explosion in game technology and advances in multi-agent systems and behavior representation in BOTs, make possible the ability to mimic human actions such as route planning and navigation. A key issue in the virtual environment that remains is how to generate human-like behaviors for BOTs. In recent years, game developers are contributing more and more effort on game artificial intelligence (AI), further supporting the importance of the need for simulating realistic human behaviors.[6] Real world events and the rising training costs in response to them has further shifted emphasis in the simulation community toward realism. This increased emphasis on realism suggests that AI-driven BOTs should be able to act as opponents against human players or as team members to cooperate with human players in the virtual environment. One well studied BOT action in this context is route planning.

Route planning, in general, is a well studied problem with a wide range of application areas, including artificial intelligence in games, robotics, and military simulation. While many algorithms exist for discovering and producing routes or paths, when the terrain can be represented as a graph, A\* is arguably the most frequently used graph search technique. First described in 1968 [7], A\* has been intensely studied and developed and now has several specialized forms. A\* or one of its specialized forms is the basis of route planning in many computer games [8]. The route planning algorithm presented in this paper expands on the principles of A\*.

## 3. SIMULATION ENVIRONMENT

The simulation environment used to develop and implement a new route generation algorithm satisfied several unique factors. First, the application program interface (API) should not introduce bias or confounding variables into the experimental design resulting from the virtual environment. Second, the API must allow for the collection, measurement and storage of game and environmental parameters without impacting game engine performance. Third, the API must support integration of the client application with the game engine physics and artificial intelligence (AI) engines. Finally, the virtual environment (map or level) must have an interface to allow physical parameters collected from humans executing defined scenarios in the physical environment to be input for statistical comparison. The Unreal Tournament 2004™ game engine with the Gamebots 2004 (GB2004) UnrealScript package

and Pogamut BOT (agent) and Java Libraries and its Netbeans plug-in addressed these factors.

### 3.1 Unreal Tournament 2004™ (UT2004)

UnrealTournament (UT) was the first game to ship with synthetic agents or BOTs. UT provides a custom scripting language, UnrealScript, through which game developers can modify (MOD) the host game. UnrealScript provides a rich Object Oriented (OO) interface to the UT game engine producing MODS such as Ravenshield™ and Infiltration™. Other UT based games, such as America's Army™ and Vegas, extend the UT2004 game engine and lock or limit the ability to MOD through UnrealScript. With the rich OO interface and the availability of Pogamut Integrated Development Environment (IDE) with its Netbeans plug-in and BOT and Java Libraries, UT2004 was selected as the base game engine.

### 3.2 Gamebots 2004(GB2004)

Gamebots, an UnrealScript package, was jointly developed by USC and Carnegie Mellon University (CMU) as an interface between the server and client. The interface provides access to sensory information such as the location and direction of a player in the game world or a message received from a teammate through synchronous and asynchronous messages communicated between server and client. BOT action commands from client to server are also accessed through this interface. Andrew Marshall at USC-ISI created a higher-level interface based on the Gamebots protocol, called JavaBot API [Marshall, 2002] to handle the specific Gamebots protocol, network socket programming, message passing, and other related issues, which makes the development of BOT AI neater and simpler.

### 3.3 Pogamut

Expanding on the JavaBot API and extending the Gamebots' UnrealScript, Jakub Gemrot and Rudolf Kadlec developed the Pogamut plug-in to the Netbeans™ IDE (<http://artemis.ms.mff.cuni.cz>). The base Pogamut Architecture, shown in Figure 1, integrates the UT2004 Server through the GameBots 2004 (GB2004) API with the Client and Netbeans IDE.

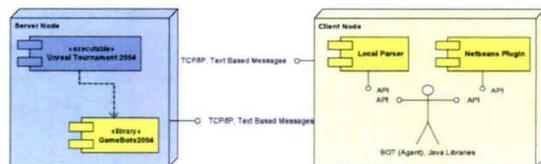


Figure 1 - Pogamut Architecture [4]

### 3.4 Oracle 11g Database with Spatial Data Objects

The basic Pogamut architecture integrated with UT2004 was extended with database functionality to permit advanced analytical processing of the environmental information available through the GB2004 and Pogamut interface. The Oracle 11g database provides two important functions. First, it provides for the parameter storage and subsequent analysis and retrieval based on BOT and Player sensory and action logic. Second, with the Spatial Data Objects (SDO), network and spatial analytics could be applied to the UT2004 map environment and collected during BOT initiation. Player/BOT monitoring functions added to the base GB2004 and Pogamut Core library provide near real time sensory and game parameters and spatially aware updates to the database. Figure 2 represents the simulation environment implemented.

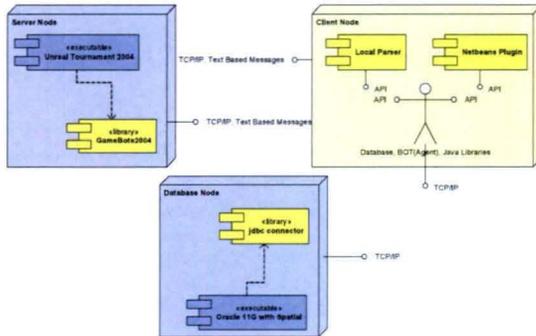


Figure 2 – Simulation Environment Architecture

## 4. ROUTE PLANNING ALGORITHM

Executing realistic route planning and navigation in the UT2004 game environment introduced several challenges. First, most game virtual environments (maps or levels) model the terrain as a graph by inserting pathnodes along navigable features such as urban streets, intersecting at distinct intersections, hall ways and stairs within buildings, and paths or trails in rural environments. A typical pattern is shown in Figure 3 for the McKenna MOUT virtual terrain used in this research. The UT2004 function to build AI Paths then computes the navigable edges between the nodes taking into account non-trafficable buildings, barriers and terrain. BOT logic programmed in the Pogamut Client instructs the BOT navigation AI to either execute a computed A\* route to destination or to proceed to a specified pathnode. Because these pathnodes are predetermined during map design, do not change (i.e. are always reachable) and do not represent all paths the human can travel within the map, realism is not achieved. Second, UT2004 stores the pathnodes and the calculated

edges in the map. BOT logic then retrieves this data during initialization and builds the navigation paths using A\*, nearest neighbor or pathnode lookup. Because all the information is available, the routes a BOT plans are always perfect (i.e. shortest distance, shortest time, or least cost). Leveraging the graph available through the Pogamut interface to UT2004 and GB2004 and the Oracle 11g database with SDO, we developed a route planning algorithm which has at its core A\*, but uses incomplete or imperfect knowledge in its execution. In addition to the SDO, the database makes available near real time dynamic information about the environment upon which the BOT logic can react. Thereby creating a more realistic BOT and addressing the challenges of the UT2004 environment.

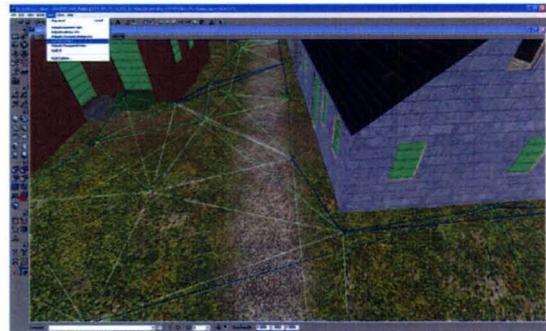


Figure 3 – Pathnodes with UT2004 edges

### 4.1 A\* (A-star) Algorithm

A\* is a best-first search algorithm that finds the least cost path from a given start pathnode to an end pathnode. Cost in A\* is an attribute of the graph edges and pathnodes included in the path. The cost values are defined to correspond to a desired property of the real-world terrain the graph represents. In UT2004, the value of the cost is determined by the map designer. Later, in our realistic design we will show how to change this to allow dynamic calculation of this cost based on the environment.

As previously noted A\*, is a well researched and documented algorithm. A good description and representation of the algorithm is found at: [http://en.wikipedia.org/wiki/A\\*\\_algorithm](http://en.wikipedia.org/wiki/A*_algorithm). The pseudo-code is shown in Figure 4. Execution of A\* begins with the identification of the start and the goal pathnode. Working partial paths are assembled from the start pathnode along the edges connecting surrounding pathnodes towards the goal pathnode. This process of following the edges from pathnode to pathnode is repeated, adding pathnodes to the partial paths until one reaches the goal pathnode. A heuristic function, denoted  $f(x)$ , where  $x$  is the end pathnode of a partial path currently being considered, is used to

calculate the next pathnode of the graph to extend the path to.

$$f(x) = g(x) + h(x) \quad (4.1)$$

where:  $g(x)$  is the total cost of the partial path

$h(x)$  is the distance from  $x$  to the goal

For A\* to be optimal, the distance function  $h(x)$  must underestimate the distance to the goal. The Pogamut implementation of A\* used in this research computes  $h(x)$  as the straight-line (Euclidean) distance to the goal pathnode, guaranteeing the underestimate (admissibility) criteria is met.

```
function A*(start,goal)
  closedset := the empty set % The set of nodes already evaluated.
  openset := set containing the initial node % The set of tentative nodes to be
  evaluated.
  g_score[start] := 0 % Distance from start along optimal path.
  h_score[start] := heuristic_estimate_of_distance(start, goal)
  f_score[start] := h_score[start] % Estimated total distance from start to goal
  through y.

  while openset is not empty
    x := the node in openset having the lowest f_score[] value
    if x = goal
      return reconstruct_path(came_from,goal)
    remove x from openset
    add x to closedset
    foreach y in neighbor_nodes(x)
      if y in closedset
        continue
      tentative_g_score := g_score[x] + dist_between(x,y)
      tentative_is_better := false
      if y not in openset
        add y to openset
        h_score[y] := heuristic_estimate_of_distance(y, goal)
        tentative_is_better := true
      elseif tentative_g_score < g_score[y]
        tentative_is_better := true
      if tentative_is_better = true
        came_from[y] := x
        g_score[y] := tentative_g_score
        f_score[y] := g_score[y] + h_score[y]
    return failure

  function reconstruct_path(came_from,current_node)
    if came_from[current_node] is set
      p = reconstruct_path(came_from,came_from[current_node])
      return (p + current_node)
    else
      return the empty path
```

Figure 4 – Wikipedia's or the A\* algorithm.

The Pogamut A\* multiplies an edge cost value determined by the level or map designer with the edge length to calculate cost variable ( $g(x)$ ). The value of the edge cost is always greater than or equal to 1, to ensure admissibility or the heuristic function,  $h(x)$ .

#### 4.2 Imperfect Knowledge Algorithm (IKA)

From the preceding discussion of A\* two limitations to the generation of realistic routes stand out. First, A\* guarantees an optimal path based on the heuristic, which in this case is distance. To mimic the route planning and subsequent navigation of humans, A\* must be modified to account for additional variables, such as cover, terrain type, slope, doors, and windows. Second, A\* calculations compute a-priori, a single optimal (based on the heuristic) path to the goal during BOT initialization. Updates to this path are computationally impractical during game play unless the number of pathnodes is minimized.

Minimizing the number of pathnodes would unrealistically limit the possible BOT paths.

To solve these limitations, we first modify the A\* heuristic function's cost variable  $g(x)$  to permit dynamic calculation of cost based on current game environment.

$$g'(x) = w_1 * g(x) + w_2 + w_3 + \dots + w_n \quad 4.2$$

where:  $w_1$  is the length cost factor

$w_2 - w_n$  are cost factors determined from BOT, Player and environment

The modified heuristic function:

$$f'(x) = g'(x) + h(x) \quad 4.3$$

where  $g'(x) > g(x)$

Second, we develop a mechanism for dynamic collection environmental parameters and limiting the available pathnodes for planning [9]. The dynamic collection of environmental parameters that include BOT, Player, and game parameters use the modified Pogamut architecture, which includes the database component and extension to the GB2004 UnrealScript Library and Pogamut Core Java Library for Player monitoring. The initial parameters ( $w_2 - w_n$ ) used include Player Visibility, Door, Window, terrain type and the standard edge length. To limit the available pathnodes, a network model of the UT2004 pathnodes and calculated edges is constructed at BOT initialization using the Oracle 11g database with Spatial Data Objects. The network model and SDO permit expanding the pathnode and edge density to more closely approximate the possible paths a human might execute. To ensure timely IKA route calculation, retrieval of pathnodes and edges is accomplished using the SDO geometry function:

$$\text{SDO\_FILTER}( p.geom, boundingbox) \quad 4.4$$

where  $p.geom$  is the SDO geometry column of the NavPoints Table, and

$boundingbox$  is the SDO geometry of the area visible to the BOT

The listing in figure 5 returns the navigable nodes and edges contained within the boundingbox ( $\text{SDO\_BB}(\text{player})$ ) modified pathnode pattern, shown in figure 6, for application in the IKA. The modified pathnode pattern ensures the possible paths available to the BOT are consistent with those available to the human player.

```
SELECT n.unrealid          //returns unrealid of navigable points
FROM navpoints n         //navpoints table constructed at BOT init
WHERE SDO_FILTER(n.geom,
  SDO_BB(player) ) = 'TRUE'; //SDO_BB(player returns Geometry of player
// location
```

Figure 5 – SQL listing for bounded pathnodes.

In addition to the SDO\_Geometry objects stored in the NavPoints table, we added the tables  $\text{tbl\_PPParams}$  and  $\text{tbl\_BParams}$  to provide the structure for storing the BOT, Player and game parameters that are used in computing the dynamic cost function of the Imperfect Knowledge Algorithm (IKA). This data structure also permits

dynamic near real time updates to the BOT, Player, and game parameters used in the heuristic, through the monitored player class added to the Pogamut Core library. The monitored player class performs sense functions on human players and BOTS running in the game and updates tbl\_PParams and tbl\_BParams with the sense results.

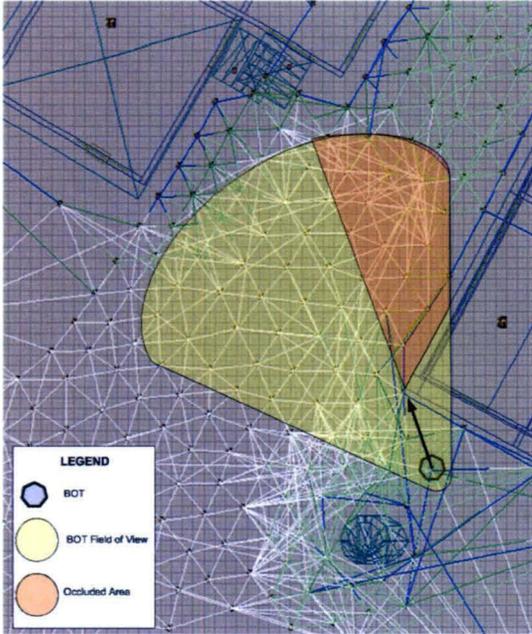


Figure 6 – UT2004 Network Model

The IKA requires additional modifications to the map or level design to identify specific terrain features such as doors, windows, and terrain type. UnrealScript is used to create additional pathnode actors that can be placed in the map or level to identify these features as well as the general planning pathnode that provides intermediate pathnodes from which the IKA will calculate the overview routes used to produce the IKA goal pathnodes. The concept of two pass route planning using UnrealTournament was introduced by Zhuoqian Shen to provide navigation through a multi-level (room) map [6].

```

// BotGPPathNode. General Purpose Node
class BotGPPathNode extends PathNode
    placeable;
// BotDestinationPathNode. Goal Node
class BotDestinationPathNode extends PathNode
    placeable;
// BotDoorPathNode. Door Node
class BotDoorPathNode extends PathNode
    placeable;
// BotWindowPathNode. Window Node
class BotWindowPathNode extends PathNode
    placeable;

```

Figure 7 – UnrealScript Listing for Pathnode extensions

As noted above, the IKA is implemented in two passes. The first pass is executed at BOT initialization to generate and store all goal

pathnodes that will be used by the IKA. During initialization, all pathnodes and their neighbors, as computed by the UT2004, are stored in a SDO within the NavPoints table. The IKA then uses A\* to compute the optimum path from the UT2004 pathnode, PlayerStartNode, to the pathnode, BOTDestinationPathNode, using only the BotGPPathNodes (Figure 8). Once the initial path has been generated as part of initialization, these points are used to determine the goal pathnodes for subsequent IKA path calculations.



Figure 8 – BOTGPPathNodes Pathnodes

Subsequent passes execute the IKA using the dynamically generated cost variable  $g'(x)$  (4.2) and the modified heuristic function  $f(x)$  (4.3). The IKA has built into the BOT's doLogic() function the sensory function that triggers route calculation, and when appropriate re-calculation of the IKA goal nodes. The pseudo-code for the IKA is shown in Figure 9.

## 5. MEASURING REALISM

The last section demonstrated the use of the UT2004 based simulation environment in the development of the Improved Knowledge Algorithm (IKA). The next step is to provide a quantitative approach to measuring its realism. Recall that in this paper we have defined realism to be a representation of human behavior that mimics or emulates the human. The UT2004 based simulation environment and the added monitor player class to the Pogamut Core Library and modifications to GB2004 provide the capability to record human player and BOT actions during a simulation run and store them in a data structure for subsequent numerical and statistical analysis. This feature will enable the development of a realism metric that measures the deviation of BOT actions from the human player. Combining the realism metric with the classic Turing test an objective and subjective assessment of the realism of the specific action being tested is possible.

```

function IKA (start, goal, pf) % pf = pass flag (0 or 1)
  bpclosedset := the empty BotGPPPathNode set
  bpopenset := set containing the initial BotGPPPathNode
  closedset := the empty set % The set of all pathnodes already evaluated.
  openset := set containing the initial node % The set of all tentative
  pathnodes nodes to be evaluated.
  g_score[start] := 0 % Distance from start along optimal path.
  h_score[start] := heuristic_estimate_of_distance(start, goal)
  f_score[start] := h_score[start] % Estimated total cost from start to goal
  through y.
  if ps = 1 % Compute intermediate route using BotGPPPathNodes
    closedset := bpclosedset
    openset := bpopenset
    w2[] := w3[] := w4[] := 0 % traditional A*
  else
    closedset := the empty set
    openset := set containing the initial node
  while openset is not empty
    x := the node in openset having the lowest f_score[] value
    if x = goal
      return reconstruct_path(came_from, goal)
    remove x from openset
    add x to closedset
    foreach y in neighbor_nodes(x)
      if y in closedset
        continue
      tentative_g_score := w1 * g_score[x] + w2[x] + w3[x] + w4[x] +
        dist_between(x,y)
      tentative_is_better := false
      if y not in openset
        add y to openset
        h_score[y] := heuristic_estimate_of_distance(y, goal)
        tentative_is_better := true
      elseif tentative_g_score < w1 * g_score[y] + w2[y] + w3[y] + w4[y]
        tentative_is_better := true
      if tentative_is_better = true
        came_from[y] := x
        g_score[y] := tentative_g_score
        f_score[y] := g_score[y] + h_score[y]
    return failure

function reconstruct_path(came_from, current_node)
  if came_from[current_node] is set
    p = reconstruct_path(came_from, came_from[current_node])
    return (p + current_node)
  else
    return the empty path

```

Figure 9 – IKA Psuedo-Code

## 6. CONCLUSIONS AND FUTURE WORK

The simulation architecture presented in this paper demonstrates an improved simulation architecture for development and testing of BOT actions. This simulation architecture was used to develop the Imperfect Knowledge Algorithm (IKA) which provides a more realistic representation of the BOT route planning action.

Continuation of this research will develop the realism metric, using the methodology described in this paper to provide a quantitative validation of the realism of the IKA.

The design and implementation of the virtual environment that supports the UT2004 based simulation architecture is very labor intensive. Functionality that can be implemented within the UnrealED to automate the terrain analysis for the placement of pathnodes and generation of the path edges should be studied and implemented.

In the current implementation of the IKA, the parameters that affect the route cost are manually set. A method for tuning the IKA parameters to optimize for realism as measured by the realism metric should be studied.

Finally, validation of the realism metric requires human data collected in a real world environment

that has been modeled in UT2004. The Fort Benning, McKenna MOUT site was selected for this research to allow for this validation. Future research should be undertaken to collect human data executing a defined scenario that can be compared using the realism metric with human player and BOT data executing the same scenario in the virtual world.

## REFERENCES

1. Beeker, E. (2004) "Potential Error in the Reuse of Nilsson's A Algorithm for Path-finding in Military Simulations", *JDMS*, vol. 1, no. 2, Apr, pp. 91–97
2. Brogan, D. & Johnson, N. (2003) "Realistic Human Walking Paths", *Proceedings of Computer Animation and Social Agents (CASA)*, pp. 94–101
3. Burgess, R. & Darken, C. (2004) "Realistic human path planning using fluid simulation", *Proceedings of Behavior Representation in Modeling and Simulation (BRIMS)*
4. Dorfler, M. (2007) "Pogamut 2 IDE for agents in UT2004", <http://artemis.ms.mff.cuni.cz/pogamut/tiki-index.php?page=Architecture>, last accessed 27 Jul 2009
5. Funge, J. et al, (1999) "Cognitive modeling: knowledge, reasoning and planning for intelligent characters", *Proceedings of the 26th annual conference on Computer graphics and interactive techniques, SIGGRAPH '99*, July
6. Shen, Z. & Zhou, S. (2006) "Behavior Representation and Simulation for Military Operations on Urban Terrain", *Simulation*, vol. 82, no. 9, Sep, pp 594-607
7. Nilsson, N. J. (1980). "Principles of Artificial Intelligence", Palo Alto, California: Tioga Publishing Company.
8. Beeker, E. (2004) "Potential Error in the Reuse of Nilsson's A Algorithm for Path-finding in Military Simulations". *JDMS* vol. 1, no. 2, April, pp 91-97
9. Hanold, G. & Hanold, D. (2009) "A Method for Quantitative Measurement of the Realism of Synthetic Character (BOT) Actions Within the UT2004 Virtual Environment", *MODSIM 2009 World (Submitted)*

# Adaptive Planning, Understanding Organizational Workload to Capability/Capacity Through Modeling and Simulation

Chris Hase, Ph.D.

*Whitney, Bradley & Brown Consulting*

*chase@wbbinc.com*

**Abstract.** In August 2003, the Secretary of Defense (SECDEF) established the Adaptive Planning (AP) initiative [1] with an objective of reducing the time necessary to develop and revise Combatant Commander (COCOM) contingency plans and increase SECDEF plan visibility. In addition to reducing the traditional plan development timeline from twenty-four months to less than twelve months (with a goal of six months)[2], AP increased plan visibility to Department of Defense (DoD) leadership through In-Progress Reviews (IPRs). The IPR process, as well as the increased number of campaign and contingency plans COCOMs had to develop, increased the workload while the number of planners remained fixed. Several efforts from collaborative planning tools to streamlined processes were initiated to compensate for the increased workload enabling COCOMS to better meet shorter planning timelines. This paper examines the Joint Strategic Capabilities Plan (JSCP) directed contingency planning and staffing requirements assigned to a combatant commander staff through the lens of modeling and simulation. The dynamics of developing a COCOM plan are captured with an ExtendSim® [3] simulation. The resulting analysis provides a quantifiable means by which to measure a combatant commander staff's workload associated with development and staffing JSCP [4] directed contingency plans with COCOM capability/capacity. Modeling and simulation bring significant opportunities in measuring the sensitivity of key variables in the assessment of workload to capability/capacity analysis. Gaining an understanding of the relationship between plan complexity, number of plans, planning processes, and number of planners with time required for plan development provides valuable information to DoD leadership. Through modeling and simulation AP leadership can gain greater insight in making key decisions on knowing where to best allocate scarce resources in an effort to meet DoD planning objectives.

## 1. INTRODUCTION

Secretary of Defense Donald Rumsfeld originated the Adaptive Planning initiative in August 2003 to improve the joint operation planning system. Secretary Rumsfeld believed that the joint operation planning process was too lengthy, existing plans were out dated in the new strategic environment, existing contingency plans could not quickly transition to execution, and the planning system was unable to coordinate military operations within the context of a whole of government response to a contingency or crisis situation [5].

Overcoming the shortcomings of the legacy joint operation planning system, the AP initiative included the following objectives [2]:

- Produce plans on demand in 1 year or less with plan revisions as needed,
- Shape plans throughout development through periodic dialogue among DoD leaders and planners,
- Perform planning in parallel and concurrent at strategic and operational levels,

- Link planning, readiness and force management processes and data in a virtual environment,
- Provide a full range and menu of military options to meet changing circumstances,
- Prioritize plans and planning effort to enable seamless transition to execution
- Determine force, logistic, transportation and operational feasibility throughout the planning process.

The AP initiative increased the workload in the planning community by reducing planning time, increasing planning visibility, increasing the number of plans and increasing the level of detail desired in the planning process.

Leadership in the AP community was sensitive to the increased workload with shorter planning timelines. However, they lacked tools necessary to quantify the increased workload and measure its impact on the existing planning capacity.

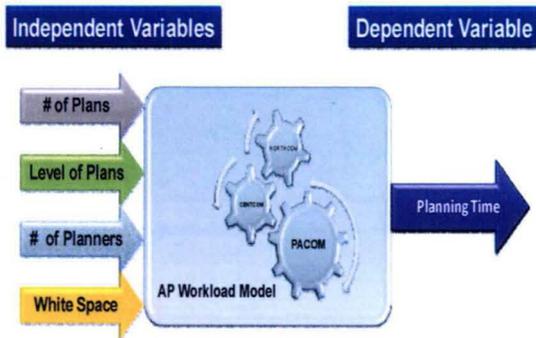
This paper addresses the adaptive planning workload to capacity study from a modeling and simulation perspective. Study details are initially presented. The value of modeling and simulation

as it supports complex process analysis is summarized at the conclusion of this paper.

## 2. ADAPTIVE PLANNING WORKLOAD TO CAPACITY MODEL

AP leadership initiated an adaptive planning workload study to assess the capacity of COCOM planners to develop contingency plans within the guidelines of the AP objectives.

Establishing a common framework from which to look at challenges brought on by AP initiatives was an important first step of this study. Defining metrics, variables and the relationship of those variables to each other was essential in providing a quantitative assessment of workload to capacity. AP documentation discussed timelines for plan development. Additionally, legacy plan development has also been interested in planning timelines. For these reasons, planning timelines was selected as the underlying metric to measure goodness within this study. Planning timelines was selected to become the dependent variable. There were many issues that affected planning timelines. These issues were divided into four themes or independent variables; number of plans, level of plans, number of planners, and white space. Figure 1 provides a graphical representation between independent variables, the work at COCOMs and the dependent variable in the AP Workload Model.



**Figure 1: AP Workload to Capacity Model**

### 2.1 Variables

Number of Plans was the first independent variable defined as the quantity of JSCP, Chairman of the Joint Chiefs of Staff (CJCS) and COCOM directed plans assigned to a combatant commander. This includes both the number of plans requiring development and those under review.

Level (Types) of Contingency Plans was the second independent variable defined by the JSCP, CJCS or COCOM directed planning level

(i.e. Level 1, 2, 3, 3T, 4). Contingency planning includes the preparation of four levels of planning detail. Planning levels are further defined as follows:

Level 1 planning (Strategic Concept) requires the least amount of planning detail and is normally completed in the shortest amount of time. Generally, COCOMs complete Level 1 planning in 10-12 weeks. From a COCOM perspective, completion of level 1 planning, documents an initial concept of operations.

Level 2 planning (Base Plan) require completion of strategic concept planning. A level 2 plan contains paragraphs one through five of the standard Operation Plan (OPLAN) format with annexes.

Level 3 planning (Concept Plan) require completion of strategic concept and completion of the base plan planning. A Level 3 plan is an abbreviated OPLAN that would require considerable expansion to be considered an OPLAN or Operation Order (OPORD). A Concept Plan (CONPLAN) contains the base plan, the commander's CONOPS with the appropriate annexes (A, B, C, D, J, K, S, V, Y and Z) and appendixes. If directed by the JSCP, planners do not calculate detailed support requirements or prepare detailed support requirements or prepare Time Phased Force Deployment Data (TPFDD) files. A level 3 plan that contains a TPFDD typically requires more detailed planning for the phased deployment of forces.

Level 4 planning (Operational Plan or OPLAN) contain the base plan, all required annexes with associated appendixes and a TPFDD. The OPLAN identifies planning assumptions and the specific forces, functional support, deployment sequence, and resources to execute the plan. The combatant commander may initiate contingency plans not directed by the JSCP. The format and content for a contingency plan is prescribed in Joint Operation Planning and Execution System [6].

The number of planners was the third independent variable defined as the quantity of planners and subjective quality of planners assigned to a COCOM staff developing and reviewing plans.

The last independent variable is defined as white space. White space was a measurement of the calendar time that it took for a planning function to be completed. More detail defining white space is presented in the next section of this paper.

The AP Workload to Capacity Model had one dependent variable called planning time. Planning time was defined as the calendar time



Most COCOMs had their lead planner sign off on the data submission.

The purpose of this model and the underlying analysis is not to be predictive on the amount of time it takes COCOMs to develop their plans but to show the relative impact of changes in planner workload, policies and resources. The number of plans and plan level determines workload. Planning inefficiencies and resources affect the planning timelines. And the number of planners, their availability and training determines resources.

### 3. DATA ANALYSIS

The primary objective of the data analysis was to provide a quantifiable basis to compare variables that affect planning timelines. Four independent variables (white space, plan complexity, number of plans, and number of planners) were assessed in isolation relative to their impact on planning, see Figure 4. Additional insights were gleaned by comparing the affect of the independent variables with each other. The independent variables were adjusted in 10% increments up to 30%. Adjustments were made that positively and negatively affected the dependent variable (planning time).

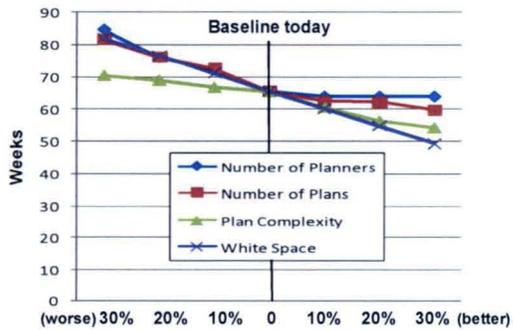


Figure 4: Comparison of Variables

The four independent variables appear fairly linear. Zero on the x axis represents status quo today. Values left of zero represent worsening conditions while values to the right of zero represent improving conditions. The y axis represents average days required for all of the plans to be processed through one cycle. This represents the total plan development and review workload. In reality, the concept of a living plan implies that plans are never complete. Once they have been approved at the appropriate level, review begins again based on changes that have occurred since the plan was initiated during the previous cycle. As discussed earlier, underlying assumptions are always being challenged. This simplistic view of averaging the time it takes for all of the plans to complete their journey from JSCP release to

plan approval (for a new plan) or plan review (for an existing plan) allows for a baseline from which to conduct analysis.

Figure 4 depicts four lines with varying slopes. Variables with greater slopes have a larger impact on planning timelines than variables with lesser slopes. For instance, changes in white space have a greater impact on planning timelines than plan complexity. In this example influencing the time it takes to complete all plan approvals/reviews has a greater impact than changing the level of some of the plans.

### 4. SENSITIVITY ANALYSIS

Sensitivity analysis is treated from two perspectives. The first perspective explores changes in the baseline date. The second perspective evaluates data ranges provided by COCOM surveys.

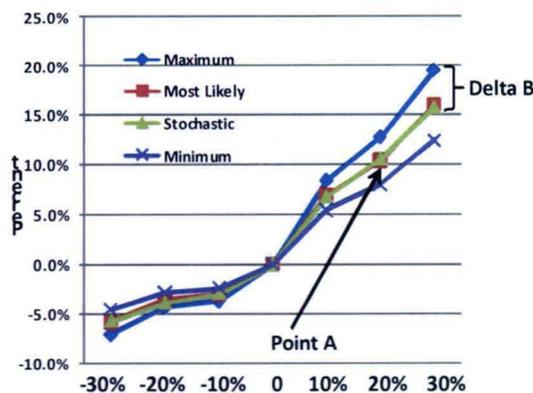
#### 4.1 COCOM Input Data

First, the point of the exploratory analysis is based on evaluating changes in the input data collected from the COCOMs. The primary analysis of this study is based on changes to the input data in an attempt to address how sensitive the input data is to change. For example, manning level sensitivity relative to the planning timeline is tested by changing planner manning levels by plus or minus ten percent increments.

#### 4.2 Baseline Data and Changes to Input Parameters

Second, as stated in the previous section, the COCOM data reflected a range in planning times from plan to plan and from COCOM to COCOM. Even within the same type of plan (level 3) at the same COCOM, planning times could be different. The independent (input) data was collected and placed into one of four bins; minimum time, maximum time, most likely time, and stochastic (based on Monte Carlo distribution). The minimum and maximum times bound the timeline range. The most likely data was obtained as the theoretical value that the data would represent if only one data point could be used. The stochastic data source represents the corresponding time it would take to develop and review plans averaged over 100 model runs if the time required to develop and review a plan was randomly picked between the minimum and maximum times the COCOMs provided. Since data analysis focused more on the delta between the baseline data and changes to input parameters, the sensitivity of input parameter was evaluated. Figure 5 presents a sensitivity comparison of plan quantities based on using maximum, most likely, stochastic, or minimum

planning timelines. The “X” axis represents changes in plan quantity from -30% to 30% in 10% increments. For example, if a COCOM developed or reviewed 16 plans (baseline value) with a given amount of resources, how would the planning timeline change if the number of plans varied from 11 to 21 utilizing the same resources? The “Y” axis represents the percent change in the planning timeline using the baseline COCOM data. Point A depicts that for a 20% increase in the number of plans there is a corresponding increase in the planning timeline by 11%.



**Figure 5:** Plan Quantity Sensitivity Analysis

The four lines represented in Figure 5 depict different perspectives of the COCOM survey data. Point A reflects no difference between the most likely COCOM data and the stochastic data. Delta B represents the difference in using the most likely or stochastic data and maximum data from the COCOM survey. Specifically, when the number of plans increased by 30% the time required to develop and review plans increased by about 16% when using the most likely or stochastic data and increased by about 19% when using the maximum time data from the COCOM surveys. If the analysis was not sensitive to the COCOM input specifying the length of time required to develop and review a plan, there would be no delta. Inspection of Figure 5 reveals less than a 0.2% delta between the most likely data source and the stochastic data source in model runs where the number of plans was varied by  $\pm 30\%$ . Figure 5 also reflects a delta of up to 4% between the most likely and maximum times as depicted by delta B. The delta is less for all other comparisons of data sources.

There is some sensitivity (up to 4%) between the most likely planning timeline and the maximum planning timeline when the number of plans is increased by 30%. There is less variation for other quantities of planning changes. The overall variation is relatively small, providing a degree of

confidence that the COCOM timeline data that was used for the data analysis accurately portrays COCOM planning timelines.

## 5. ANALYTIC INSIGHT

The data collected from the COCOMs was based on existing planning processes. One of the tools used during the TPFDD development phase for those plans that required TPFDDs was Collaborative Force Analysis, Sustainment, and Transportation (CFAST). CFAST [7] is a portal-based, collaborative campaign planning tool that provides a set of business tools and supporting infrastructure that shortens contingency and crisis action planning, command exercises, force modernization studies, and analysis. Real time data immediately updates interactive map displays and other visualizations for continuous monitoring and effective response to dynamic situations. CFAST was plagued with various problems that prevented the tool from working as effectively as was desired. After the COCOM data had been collected, CFAST as a planner's tool was terminated.

During the socialization of this study with some of the COCOMs, concern was raised that the data analysis provided over optimistic timelines. Those timelines were based on COCOM data that assumed CFAST was supporting TPFDD development. With the cancelation of CFAST, a couple of the COCOMs indicated their planning timelines during TPFDD development doubled and that three planning conferences were now required where only one was needed before. This information provided a unique opportunity to assess the loss of CFAST from the perspective of the model developed for this study. As expected, overall plan development timelines increased as a result of increases in TPFDD development timelines, though not as much as one would expect. Not all COCOMs have plans requiring TPFDDs and those that do, only a fraction of their plan portfolio contain plans with TPFDDs. Some TPFDDs contain only a few hundred units while others contain thousands. Overall, the model showed an increase on average of 7% across all plans. The data from one COCOM that has a very complex TPFDD demonstrated a far greater impact on plan development timelines based on the model. The model indicated as much as a 42% increase in the plan development time for one plan that has an extensive TPFDD. Insights from these additional model runs imply that the use of collaborative planning tools can have a direct impact on plan development timelines.

## 6. ANALYTIC CONCLUSIONS

The results of numerous model runs have provided useful information to AP leadership on where to focus limited resources. The value of trained planners was reinforced. The impact of collaborative planning tools and the loss of CFAST was quantified. The importance of streamlining planning processes and eliminating white space was validated.

More importantly, decisions by senior leaders in changing the requirement for the number of plans or planning detail can now be quantified. Together, this analysis supports the AP community as they continue to update their strategic roadmap with various initiatives in assisting the planning community in the years ahead.

## 7. SUMMARY

Military planning has been ongoing for centuries. Changes in the geo-political landscape, the mobility of modern forces, and collaboration technologies have brought about changes in the speed of government within the DoD as it relates to military planning. AP initiatives have codified objectives in changing the visibility, quantity, and frequency of plan development within the highest levels of government.

The AP Workload to Capacity study has provided a tool to assist decision makers on the best use of limited resources in meeting the Departments planning objectives. Reaching the most efficient balance of trained planners, collaboration technologies and policy decisions that affect planning processes forms the cornerstone in meeting planning timeline objectives.

The field of modeling and simulation was instrumental in providing AP decision makers the analysis and tools necessary to assist the Joint Planning and Execution Community. Future work by the Adaptive Planning Implementation Team will leverage the foundation of analysis brought about through computer simulation.

## REFERENCES

1. Rumsfeld, D. (2003) "Implementation of Adaptive Planning", Secretary of Defense Memorandum, Washington, DC.
2. Gates, R. (2008) "Implementation of AP Roadmap II", Secretary of Defense Memorandum, Washington, DC.
3. ExtendSim® (2008) Developed by Imagine That, <http://www.extendsim.com/>.

4. Mullen, M. (2008) "CJCSI 3110.01G, Joint Strategic Capabilities Plan FY208-2010 (S)", Washington DC.
5. Klein, R. (2007) "Adaptive Planning, Not Your Great Grandfather's Schlieffen Plan", *JFQ*, Issue 45, 2<sup>nd</sup> Quarter, 2007 pp. 84-88.
6. Sharp, W. (2007), "CJCSM 3122.03C, Joint Operation Planning and Execution System Volume II", Washington, DC.
7. Collaborative Force Analysis, Sustainment, and Transportation (n.d.) Developed by DPRA, <http://www.dpra.com/index.cfm/m/269>.

# The Application Of Modeling And Simulation In Capacity Management Within The ITIL Framework

Sonya Rahmani; Otto von der Hoff

*SRA International, Inc.*

[sonya\\_rahmani@sra.com](mailto:sonya_rahmani@sra.com), [otto\\_von\\_der\\_hoff@sra.com](mailto:otto_von_der_hoff@sra.com)

**Abstract.** Tightly integrating modeling and simulation techniques into Information Technology Infrastructure Library (ITIL) practices can be one of the driving factors behind a successful and cost-effective capacity management effort for any Information Technology (IT) system.

ITIL is a best practices framework for managing IT infrastructure, development and operations. Translating ITIL theory into operational reality can be a challenge. This paper aims to highlight how to best integrate modeling and simulation into an ITIL implementation.

For cases where the project team initially has difficulty gaining consensus on investing in modeling and simulation resources, a clear definition for M&S implementation into the ITIL framework, specifically its role in supporting Capacity Management, is critical to gaining the support required to garner these resources. This implementation should also help to clearly define M&S support to the overall system mission.

This paper will describe the development of an integrated modeling approach and how best to tie M&S to definitive goals for evaluating system capacity and performance requirements. Specifically the paper will discuss best practices for implementing modeling and simulation into ITIL. These practices hinge on implementing integrated M&S methods that 1) encompass at least two or more predictive modeling techniques, 2) complement each one's respective strengths and weaknesses to support the validation of predicted results, and 3) are tied to the system's performance and workload monitoring efforts. How to structure two forms of modeling: statistical and simulation in the development of "As Is" and "To Be" efforts will be used to exemplify the integrated M&S methods. The paper will show how these methods can better support the project's overall capacity management efforts.

## 1. Introduction

ITIL is a best practices framework and set of guidelines that define an integrated, process-based approach for managing information technology services. Translating the ITIL theory into operational reality can be a challenge. Methods of implementation and best practices using ITIL principles are out of scope for this paper. Rather, this discussion aims to highlight how best to integrate modeling and simulation into ITIL implementations.

A clear definition for M&S implementation into the ITIL framework especially its role in supporting Capacity Management is critical to gaining customer and stakeholder buy-in. In the case example, discussed later in this paper, the team had difficulty gaining consensus on investing in modeling and simulation resources. The benefits of modeling and simulation were unclear to the project's overall mission and as a result there was insufficient modeling resource allocation.

However, once M&S was tied directly to the system's Capacity Management activities as part of ITIL, the M&S efforts gained traction. Lessons learned from this case example have been leveraged as part of developing this paper's thesis.

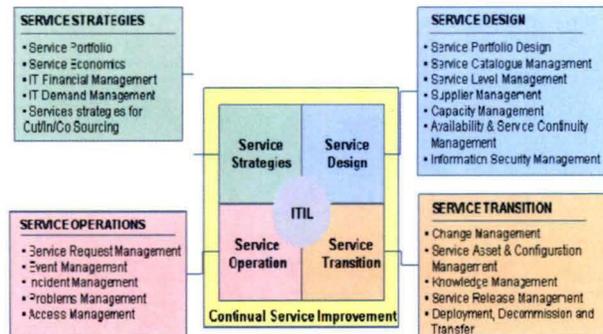
The successful implementation of M&S within ITIL will encompass the following characteristics: 1) use of at least two or more predictive modeling techniques, 2) methods complement each one's respective strengths and weaknesses to support the validation of predicted results, and 3) techniques are tied to the system's performance and workload monitoring efforts.

## 2. ITIL BACKGROUND

ITIL encompasses a set of concepts and policies for managing information technology infrastructure, development and operations. ITIL consists of the following five disciplines (illustrated in Figure 1):

- Service Strategy
- Service Design

- Service Transition
- Service Operation
- Continual Service Improvement



**Figure 1: M&S Integration into ITIL Framework**

## 2.1 M&S and the ITIL Framework

The scope of Service Design includes the design of new services, as well as changes and improvements to existing ones. Service Design consists of several areas; however, for purposes of this discussion, the focus will be on the Capacity Management area.

## 2.2 Implementing M&S Using ITIL Framework

Capacity Management is the discipline that ensures IT infrastructure is provided at the *right time* in the *right volume* at the *right price* and is used in the most efficient manner.<sup>1</sup> The real success lies in implementing an integrated M&S approach that 1) encompasses at least two or more predictive modeling techniques, 2) complement each techniques' respective strengths and weaknesses to support the validation of predicted results, and 3) is tied to the system's performance and workload monitoring efforts.

For system development and deployment projects that are still in early operational stages, additional model validation challenges may arise from the lack of a scalable Performance Test environment or a full system monitoring solution thereby limiting access to actual performance data. Using at least two types of modeling techniques can help to overcome this early validation challenge by raising confidence in model results where general agreement is obtained using disparate modeling techniques. In addition, the combination of M&S methods can successfully deliver capacity

forecasting flexibility for both large and small scale projects.

Projects with the following characteristics will most benefit from an M&S implementation tied to ITIL principles:

- Clear-cut performance analysis goals
- Strict Service Level Agreements (SLAs) or Operational Level Agreements (OLAs)
- Enterprise class applications
- Volumes experiencing significant growth
- Time-based mission critical or real-time systems
- Lack of a full-scale performance test environment (need for alternative system evaluation techniques)
- Cost sensitive capacity requirements
- Long lead-time resource acquisition

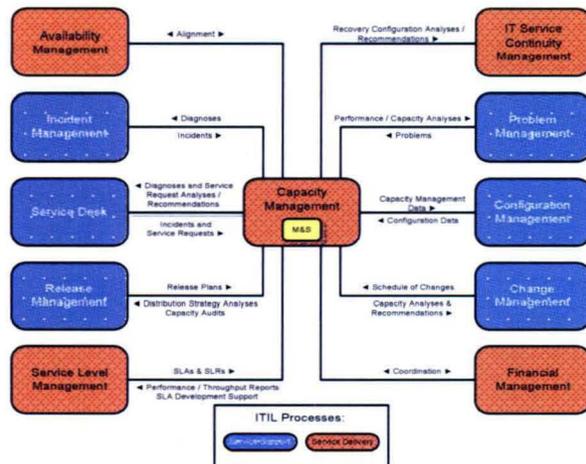
The M&S implementation should be driven by definitive goals for evaluating system capacity and behavior given clearly stated performance requirements. The M&S implementation team likewise needs to be equipped with performance analysis and engineering expertise together with target system subject matter knowledge. Furthermore, the project's ITIL framework should be tailored to tie M&S to the following ITIL activities: Monitoring, Demand Management, Performance Tuning and Application Sizing.

## 3. A CASE STUDY

A case study on a federal IT system is used below as an example to illustrate M&S implementation in ITIL's Capacity Management processes. The federal system contains over 100 million records and processes close to 50 million requests annually. In addition, the system specifically meets the program characteristics in Section 2.2 mentioned above.

All these factors underscored the need for a robust and flexible capacity management program. As a result, a formal Capacity Management Process was created using the ITIL framework. The ITIL framework was tailored to support the federal system's overall Service Delivery and Service Support functions. In creating the Capacity Management processes, the project implemented modeling and simulation activities as a set of integrated activities. Figure 2 illustrates the M&S relationship central to Capacity Management within the program's ITIL process framework:

<sup>1</sup> ITIL Open Guide. March 2, 2009. <  
<http://www.itlibrary.org>>



**Figure 2: M&S Central Relation to Capacity Management within Enterprise ITIL Framework**

As part of this implementation, M&S activities were joined to several ITIL activities (as described below):

- Monitoring – system performance data (e.g., resource utilization metrics, response times, throughput, etc.) and workload monitoring (e.g., arrival patterns, transaction volume, etc.) were collected and analyzed from both the Production and Test environments. M&S uses these data to build and update the models.
- Demand Management – M&S applies stochastic abstractions and transaction volume models to workload impact analyses.
- Performance Tuning – M&S supports project efforts to identify steps required to handle current and/or new workloads to optimize system performance or operational policy.
- Application Sizing – M&S supports identification of resources needed for a new system application or a change to existing application. For example, model results provide input into hardware acquisitions required for new system deployments.

One of the most significant factors that contributed to the success of the program's Capacity Management Process was the tightly integrated M&S implementation within the project's overall ITIL framework.

The ITIL framework references four modeling techniques and includes Trends Analysis, Analytical Modeling, Simulation Models and Baseline Models. This paper classifies both Trends Analysis and Analytical Modeling as forms of statistical techniques. In addition, Baseline Models are defined in the context of a simulation

model, and defined as a “benchmark” of the current (“As Is”) system performance.

This case example illustrates that it is the combination of both statistical and simulation modeling techniques that directly support making the program's Capacity Management Process a success.

### 3.1 M&S Techniques in Case Example

A combination of statistical and simulation model techniques were used to quantify performance, estimate capacity, provide subject matter input, and afford validation to the overall modeling activities. Statistical techniques included:

- Trending using ARIMA (Auto-Regressive Integrated Moving Average) models for time series data – these methods were used to support characterization of existing system workloads and forecasting of future growth patterns based on historical volumes.
- Analytical model development efforts – these were used for several different needs including deriving mathematical expressions of system workloads to characterize workload arrival patterns and critical resource capacity models. In addition, historical transaction data were also analyzed to identify key performance factors and develop reusable statistical descriptions of the system's behavior.

Figure 4 illustrates typical transaction workload regression trending models for two classes of system transactions. The blue line depicts historical data whereas the red line represents the regression predictions. The use of ARIMA modeling techniques suitably captures the temporal characteristics of workload seasonality as well as year over year background growth where present.



**Figure 3: Transaction Arrival Trending Models**

Statistical modeling strengths includes the ability to apply relatively simple methods that require shorter turnarounds to getting answers as well as requiring less detailed input data; weaknesses include a higher risk of being less accurate for predicting

response times and throughput, loss of predictive accuracy where future behavioral patterns vary substantially relative to historical patterns, and an inability to deal with queuing and resource contention analysis.

Simulation modeling is used to gain more accurate predictive results for response time, throughput and resource consumption. The simulation modeling techniques included:

- “As Is” simulation model development efforts that craft simulation models of the existing systems and validate against the performance in the production environment (baseline model which “benchmarks” the current system).
- “To Be” simulation models that leverage the “As Is” models to develop the anticipated views (i.e. future operating conditions).

Simulation modeling strengths include providing the capabilities for more accurate projections of system throughput and response times in support of hardware acquisition estimates and architecture validation efforts and the ability to predict and analyze dynamic queuing properties and resource contention conditions. Simulation modeling weaknesses can include requiring a longer turn around time and large volumes of detailed output performance data. Valid use of the simulation model results will depend on the accuracy of the performance data used to develop the models.

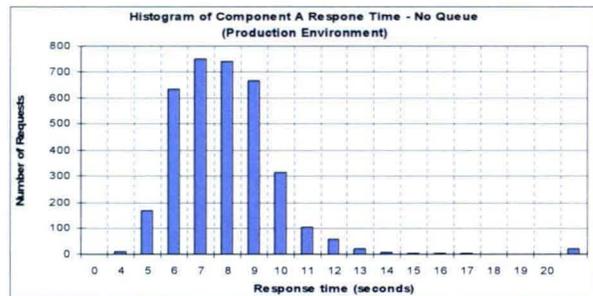
However, when used in collaboration, the two differing modeling techniques can be combined to generally support a broader set of performance analysis needs and introduce flexibility in satisfying the project’s capacity management objectives.

### 3.2 Developing the “As Is” Models

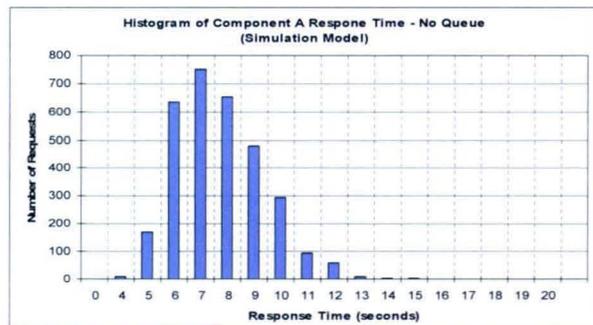
Early on, one of the biggest challenges was lack of production monitoring on the legacy system components. The project had an urgent need for precise simulation model results; however, most of the legacy system lacked any performance monitoring tools that would correlate workload to resource consumption (e.g., CPU, disk reads/writes, etc). As a result, the integrated M&S methods were tailored to tackle these challenges by modeling parts of the system as a “black box” and using a combination of statistical and simulation techniques.

The statistical analysis encompassed evaluating historical performance data (such as response time and throughput) to characterize statistical latency distributions under no queuing conditions.

These techniques were used to combat the lack of instrumented performance data on specific pieces of the system. For these components, historical response data were analyzed to identify a time where there was little or no queuing in the system. During these periods, the start and finish times of each transaction were collected and used to create a histogram illustrating resulting service times. The histogram data were used to build the best fit curve characterized as a probability distribution. Thereafter, the team used the distribution to represent the system service time in the simulation model. Figures 4 and 5 below illustrate the histogram of response times under no queuing conditions for Production and the Simulation Model.



**Figure 4: Histogram of Component A Response Time - No Queue (Production)**



**Figure 5: Histogram of Component A Response Time - No Queue (Simulation Model)**

Although, the model simulation was not able to entirely capture the long tail observed in Production data, the associated statistical data demonstrated that there was little difference in overall response time between the simulation and production data results (see Tables 1 and 2 below).

**Table 1: Production Statistics**

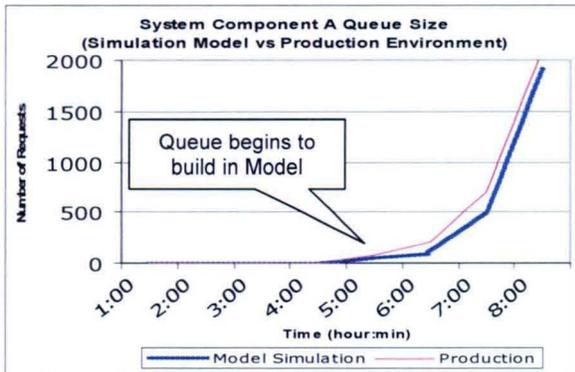
Mean	7.87
Median	8
Mode	7

Standard Deviation	1.79
--------------------	------

**Table 2: Simulation Model Statistics**

Mean	7.72
Median	8
Mode	7
Standard Deviation	1.65

After validating service times, the queuing behavior was analyzed using a time period starting with an empty queue that gradually built over time. The service request arrival times were also assessed for that period. The simulation model was run with the statistically derived service and arrival time models. Figure 6 depicts the validated simulation results:



**Figure 6: Simulation Model vs. Production Environment Queue Validation**

The team compared the model's simulation results to production data in order to successfully validate against the true system performance. In this manner, the team was able to leverage two different modeling techniques to successfully build the "As Is" simulation model. The statistical analysis facilitated service time characterization in a manner that could then be applied in the simulation models. This would not have been possible without these statistical models due to lack of production performance data. In addition, if we had used statistical techniques in isolation, we would not have been able to vary response time and correlate this to queuing behavior over the course of a day.

Simulation models were subsequently updated once production monitoring tools had later been deployed. Collected performance data were evaluated using analytical techniques to associate resource consumption with the workload executed

(viz. CPU, database reads/writes, etc). The simulation model was validated under full workload conditions by comparing results (response time, throughput, and CPU consumption) to the production environment. Production changes (e.g., new code deployed, architecture or platform changes, etc.) could then be quickly rendered in the simulation environment by leveraging monitored data against the validated baseline "As Is" model.

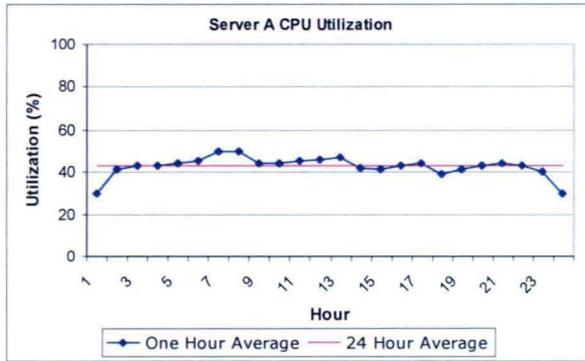
Implementing two different modeling techniques therefore proved critical to performing capacity management early in the system development lifecycle when performance data were not yet available. Model accuracy was in turn improved after production data became available.

### 3.3 Leveraging "As Is" to Forecast Impact of New Workloads

The program's Capacity Management forecasting responsibilities includes regular engagement with the system stakeholders to identify workload changes that may impact the IT system's performance and computational resource needs.

A recent workload addition of several million records exemplifies the important role M&S played in the Capacity Management process. The M&S team worked closely with the Demand Management office to characterize the new workload's yearly demand based on historical behavior of similar historical service request types. The team used statistical regression models to predict future seasonal arrival patterns and adapted existing workload distributions into daily workload arrival patterns for the new transactions. Finally, the "As Is" model was simulated with the new workloads. The team provided analysis on expected response times, throughput, and resource utilization plus impacts anticipated to existing workloads.

Figure 7 below illustrates an example of resource utilization forecasted data.



**Figure 7: Simulation Model – Forecasted Server A Utilization**

Figure 8 below illustrates an example of forecasted system response times. Adherence to SLA response times were of critical importance to the customer and program.



**Figure 8: Simulation Model – Forecasted System Response Times**

### 3.4 Leveraging “As Is” to Develop “To Be” Models

Recently, the government system went through a massive modernization effort that upgraded both its hardware and software components. The customer expressed several concerns on how this would impact operations and most specifically SLA adherence. An M&S Tiger Team was therefore tasked to develop simulation models that would help forecast computational resource requirements to deliver needed capacity and to justify capital equipment acquisitions. Of additional concern were possible impacts to the front-end business processes and wide area network performance.

The M&S Tiger Team’s objective was to develop an end-to-end analysis solution that would provide an impact analysis on all three aspects of the business. On the back-end system, specific questions were raised on identifying impacts to

resource consumption and response times. For the latter, the back-end “To Be” system model was built leveraging the “As Is” simulation model described in Section 3.2 above. The resulting analysis assembled a comprehensive picture of the new system deployment impacts.

Performance analysis helped to proactively identify specific impacts and areas for operational improvement to ensure a smooth transition during system modernization. This was one of the most successful initiatives on the project demonstrating the critical insight that can be gleaned from using a combination of modeling techniques.

## 4. CONCLUSION

In conclusion, the development of an integrated modeling approach can significantly impact the success of the project’s overall capacity management efforts. The M&S implementation should encompass two or more predictive modeling techniques, complement each one’s respective strengths and weaknesses to support the validation of predicted results, and be tied directly to system performance and workload monitoring efforts.

The implementation should include evaluation of the “As Is” system as well as forecasting techniques. The models developed in support of the latter’s analysis should provide estimates for response times, throughput, and resource utilization for the “To Be” system. Furthermore, models should be designed to guide the project’s hardware acquisition and architecture validation efforts. From the beginning, the ITIL framework should be tailored to implement M&S within Capacity Management processes and relate to the following activities: Monitoring, Demand Management, Performance Tuning and Application Sizing activities.

Following these high level guidelines will establish and promote a successful Capacity Management Program for a broad array of enterprise IT application systems.

# Handling Emergency Management in Object Oriented Modelling Environment

Berna Eren Tokgoz; Volkan Cakir; Adrian V. Gheorghe

*Department of Engineering Management and Systems Engineering, Old Dominion University,  
Norfolk, VA, USA.  
e-mail: berentok@odu.edu*

**Abstract.** It has been understood that protection of a nation from extreme disasters is a challenging task. Impacts of extreme disasters on a nation's critical infrastructures, economy and society could be devastating. A protection plan itself would not be sufficient when a disaster strikes. Hence, there is a need for a holistic approach to establish more resilient infrastructures to withstand extreme disasters. A resilient infrastructure can be defined as a system or facility that is able to withstand damage, but if affected, can be readily and cost-effectively restored. The key issue to establish resilient infrastructures is to incorporate existing protection plans with comprehensive preparedness actions to respond, recover and restore as quickly as possible, and to minimize extreme disaster impacts. Although national organizations will respond to a disaster, extreme disasters need to be handled mostly by local emergency management departments. Since emergency management departments have to deal with complex systems, they have to have a manageable plan and efficient organizational structures to coordinate all these systems. A strong organizational structure is the key in responding fast before and during disasters, and recovering quickly after disasters. In this study, the entire emergency management is viewed as an enterprise and modelled through enterprise management approach. Managing an enterprise or a large complex system is a very challenging task. It is critical for an enterprise to respond to challenges in a timely manner with quick decision making. This study addresses the problem of handling emergency management at regional level in an object oriented modelling environment developed by use of TopEase® software. Emergency Operation Plan of the City of Hampton, Virginia, has been incorporated into TopEase® for analysis. The methodology used in this study has been supported by a case study on critical infrastructure resiliency in Hampton Roads.

## I. INTRODUCTION

Hampton Roads consists of sixteen city and county jurisdictions, and is home to 1.6 million people (the fifth largest metro area in the south eastern U.S. and the second largest metro area between Washington, D.C. and Atlanta). Hampton Roads is very critical for national security both militarily and economically, because it has the largest complex of military bases in the world and the second-largest port on the Atlantic coast, and it is the site of the world's largest shipbuilder of combat vessels. Hampton Roads is low-lying and thus prone to flooding, is vulnerable to the effects of hurricanes and occasionally tornadoes, and is a likely target for terrorist attacks. It is obvious that protection plans are not enough after a disaster strikes to the region. Thus, there is a need for a holistic approach to establish more resilient infrastructures to withstand extreme disasters. A resilient infrastructure can be defined as a component, system or facility that is able to withstand damage or disruption, but if affected, can be readily and cost-effectively restored. The key issue to achieve this is to

incorporate existing protection plans with comprehensive preparedness actions to respond, recover and restore as quickly as possible, and to minimize extreme disaster impacts [1]. Therefore, Critical Infrastructure Resilience of Hampton Roads Region (CIRHRR) project, which was funded by Department of Homeland Security, has been done to analyze the regional resiliency in terms of four critical infrastructures; namely electricity, transportation, communications and water sectors. Service interruption of any one or more of these interdependent infrastructures due to various threats could be catastrophic not only for the region but also for the entire nation.

Extreme disasters, both natural and manmade, must be handled by each city's emergency management departments. It is important to have a manageable plan which is prepared by emergency management departments before disasters. Since emergency management departments have to deal with other complex and large scale systems such as plant managements, public utilities, fire department and police

department, they have to have efficient and effective organizational structures to coordinate all these systems. A strong organizational structure is the key in responding fast before and during disasters, and recovering quickly after disasters. In order to establish strong emergency management and organizational structures for cities, an enterprise management approach can be useful. Managing an enterprise or large complex system is a very challenging task because of rapid technological changes, complex economic dynamics and adaptation to new markets, trends and opportunities. It is critical for enterprises to respond to these challenges in a timely manner with quick decision making. TopEase® is a software tool which provides managers the necessary critical information on an enterprise itself to visualize a holistic picture of a complex system such as an emergency service providing enterprise [2]. TopEase® has been developed by a Swiss Company, Pulinco Engineering AG, to provide a methodology for a holistic view of a system to manage its complexity, to get transparency, and to control the change and/or transformation processes for continuous improvement and success. It is designed to handle “business processes” and provides a desirable end state of an enterprise, business or application while highlighting the gap between the current “as is” and desired “to be” states. Therefore, an approach based on TopEase® has been used in the CIRHRR project. The goal of this application was to show that emergency management operations can be handled as enterprise management processes. In order to achieve this goal, Emergency Operation Plan (EOP) of the City of Hampton, Virginia has been implemented into TopEase®. In this paper, some examples and TopEase® implementation process are presented. TopEase® models emergency management at regional level in an object oriented environment. Hence, Object Oriented Programming (OOP) will be briefly explained in the next section.

## II. OBJECT ORIENTED PROGRAMMING

“Objects” are used to design applications and computer programs in OOP. OOP can use several concepts or techniques from previously established paradigms such as inheritance, modularity, polymorphism or encapsulation. These concepts give support to the development of efficient class structures. The aim of OOP is to approximate the behaviour of real world elements within software environment [3]. In OOP, each object

is capable of receiving messages, processing data and sending messages to other objects. Each object can be viewed as an independent unit with a distinct role or responsibility. Some of the key concepts of OOP are described next:

*Class* defines the abstract characteristics of a thing (its attributes or properties) and the things it can do (its behavior).

*Object* is a particular instance of a class, and is a software package that includes all the necessary data and procedures to represent a real world object for a specific set of purposes.

*Message Passing* signifies the objects interacting with each other by sending requests for services known as messages.

*Encapsulation* is the mechanism by which related data and procedures are bound together within an object. It conceals the exact details of how a particular class works from objects that use its code or send messages to it.

*Polymorphism* is the behavior that varies depending on the class in which the behavior is invoked, that is, two or more classes can react differently to the same message. The power of polymorphism is that it greatly simplifies the logic of programs by shortening and increasing the execution speed.

*Inheritance* is the mechanism that allows classes to be defined as special cases, or *subclasses*, of each other [2], [4].

The approach used by TopEase® is explained in the next section.

## III. TOPEASE® SOFTWARE AND ITS APPROACH

TopEase®, which has been used for more than 20 years as a *business application* tool, aims to provide solutions to the problems of variety of sectors such as economic, health and law enforcement. TopEase® helps manage the complexity of a system, see the holistic aspects of a system and control its processes at every step to achieve continuous improvements in the system. The idea behind developing this software was to establish balance between principles and pragmatism. TopEase® uses 1-3-5-7 axioms to achieve solutions for businesses (Figure 1).

*1 methodology* provides a common understanding based on an established terminology. TopEase® uses a single methodology, which pursues principals in a pragmatic and balanced manner to accomplish its targets.

3 layers assist in obtaining a target audience related business structure. 3 layers are named as definition, support and implementation layers.

5 models provide a system to be modelled, documented and elaborated. The system may be validated through value chains and questions, if all artifacts are modelled appropriately. 5 models are named as business, resource, information, delivery and change models.

7 questions help analyze and interpret connections between 3 layers and 5 models. It is important to ask 7 significant questions to determine interrelationships among nodes which are constructed models. These questions are about cost, benefit, risk, quality, feasibility [produce (how)], manageability [people (who)] and impact.

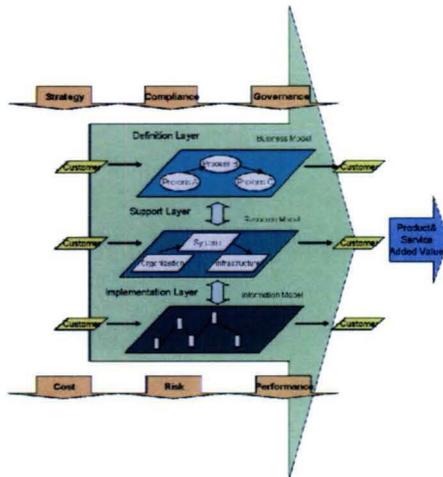


Figure 1: 1-3-5-7 approach used by TopEase® [2]

Problem definition and methodology used in the CIRHRR project are presented in the next section.

#### IV. PROBLEM DEFINITION AND METHODOLOGY

##### A. Critical Infrastructure Resilience of Hampton Roads Region Project

In order to enhance regional security and resiliency of Hampton Roads, a complex set of management and policy issues are required to be addressed. Diverse local jurisdictions and the range of federal, commonwealth and private facilities serving the region complicate the analysis and coordination of regional security and reliability. In order to establish a regional disaster mitigation, response and

recovery plan for Hampton Roads, there is a need for an integrated regional model for all branches of jurisdictions and private facilities involved. Facilities in all jurisdictions, relationships among them and their dependencies on private facilities have been analyzed to determine response and recovery capabilities of these jurisdictions during emergency situations.

##### B. Hampton City Emergency Model

Hampton Roads has unique characteristics as a region. Existence of multiple jurisdictions, privately owned utility companies and military facilities require an analytical solution through the application of the system of system technology. This seems to be adequate for addressing issues of emergency situations. In order to analyze the current state of emergency plan, the EOP of the City of Hampton has been implemented into TopEase® as part of the CIRHRR project. Critical infrastructures have been modelled as layers (Figure 2). Functions of these integrated critical infrastructures and emergency operations against different threats are cross cutting issues as can be seen in Figure 3. The purpose of this study was to put the EOP of the City of Hampton in a single model to see the interdependencies among critical infrastructures.

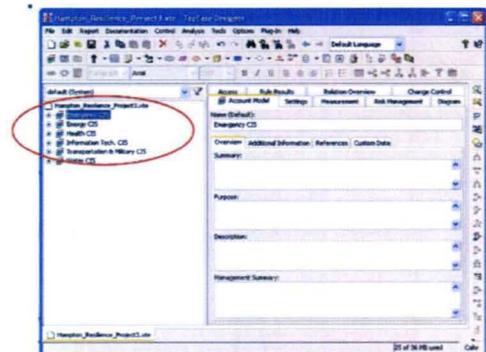


Figure 2: Critical infrastructures as layers

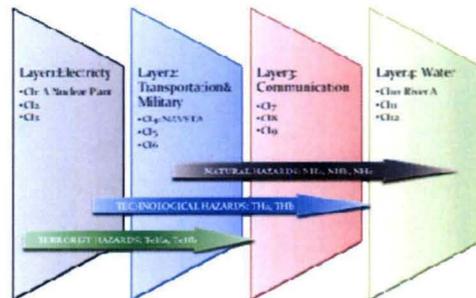


Figure 3: Critical infrastructures and threats associated with them

As the first step, the organizational chart of Hampton City Management has been incorporated into TopEase® RACI (Responsible, Accountable, Concerned and Informed) matrix as detailed as in the EOP of the City of Hampton. The RACI matrix evaluates and maps the characteristics and responsibilities of various positions in the organizational chart. Hence, RACI helps mapping of all artifacts which allow describing and analyzing the entire organizational mapping, influence, and most importantly, interdependencies. Part of the RACI matrix output can be seen in Figure 4.

Figure 4: Example of the RACI matrix output

Processes including detailed activities, lifecycle of emergency operations and their activities from origination to termination have been adapted directly from the EOP of the City of Hampton. Since full representation is impossible, partial representation of processes and activities along with assigned people, jobs and roles are visualized in Figure 5. TopEase® can be used to generate charts and diagrams that show complex and interconnected components of a system.

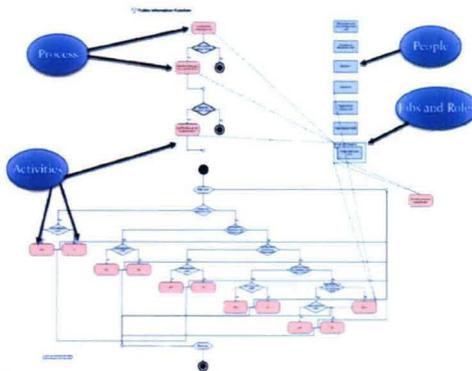


Figure 5: Visualization of processes, activities and roles

Risk catalogue function of TopEase® can handle all kinds of risk to the system. Risk is

defined in TopEase® by two parameters which are impact and likelihood. Categories of impact and likelihood are given in Tables 1 and 2, respectively. In our study, risk of having a disaster can be assigned to any operation or critical infrastructure to calculate the total risk of that disaster. This assigned risk can be represented in different ways such as using a risk matrix or interdependency diagram. For instance, likelihoods of having different types of disasters versus impacts of these disasters on a nuclear power plant are visualized as a risk map in Figure 6.

Table 1: Likelihood ranking categories

General	Frequency of Reoccurring Events	Probability of a One-off Event
Improbable	Once every 10,000 years	1 in 1,000
Remote	Once every 1,000 years	1 in 100
Occasional	Once every 100 years	1 in 10
Probable	Once every 10 years	More likely than not
Frequent	Once every year	Almost certain

Table 2: Impact severity categories

Descriptive	Safety	Security
Minor	Minor injuries	Minor breach regulations
Moderate	Major injuries	Reportable breach of regulations
Significant	Single fatality	Prosecution
Substantial	Multiple fatalities (tens)	
Mega	Multiple fatalities (hundreds)	

## V. CONCLUSION

As part of our efforts to obtain a holistic view of the EOP of the City of Hampton, this plan has been incorporated into TopEase® as a single model. TopEase® was developed for "business process" solutions based on OOP paradigm. The main idea behind implementing the EOP into TopEase® was to approach emergency management operations as a "business process" and to define critical

infrastructures as different layers in a single model. Other functions of TopEase® have also been used in the model in order to incorporate every detail of the EOP into a single model. For example, definitions in the EOP have been incorporated into the model via a glossary function. References and laws have also been included in the model. In addition, governmental partners like FEMA that are outside system boundaries have been identified and incorporated into the model as external agents. These functions altogether will provide ways of understanding the effects of interdependencies for determining vulnerable parts of the system.

As a future study, EOPs of other jurisdictions could be modeled to be able to make comparisons among jurisdictions to see their interdependencies. Such an analysis will provide the opportunity to evaluate vulnerable points of jurisdictions and to give decision makers an idea to use limited resources effectively. In addition, TopEase® has “as-is” and “to-be” functions which will be useful for comparative analysis and for coordinating the transition between “as-is” and “to-be” states.

As part of future study, impact analysis feature of TopEase® could be used to identify vulnerable points of critical infrastructures. It is possible to develop different threat scenarios in TopEase® to see how the system reacts under stress. As a powerful feature, any changes in an enterprise such as changes in system parameters, or presence or absence of a process can be tracked by the impact analysis. It is possible to analyze how an unexpected failure in part of a critical infrastructure can affect other infrastructures by using the impact analysis. This type of analysis can help emergency managers see all possible interdependencies among critical infrastructures.

Likelihood \ Impact	Very low	Low	Medium	High	Very High
Very high					
High	A nuclear plant Earthquake	A nuclear plant Cyber attack	A nuclear plant Hurricane		
Medium		A nuclear plant Industrial ac.			
Low					
Very low					

Figure 6: Risk map for a nuclear power plant

## VI. REFERENCES

1. Scalingi, P. L., "Moving beyond Critical Infrastructure Protection to Disaster Resilience," *Critical Thinking: Moving from Infrastructure Protection to Infrastructure Resilience, CIP Program Discussion Paper Series*, George Mason University, 2007.
2. Pulfer, R. & Schmid, U. (2006) "Control Your Business: The Balance between Principles and Pragmatism," Pulinco Engineering AG.
3. Agostini M.N., Decker, I.C, & Silveira e Silva, A. (2007) "A new approach for the design of electric power system software using object oriented modeling," *Electrical Power and Energy Systems*, vol. 29, pp. 505–513.
4. Gossain, S. & Anderson, D. B. (1999) "An object oriented approach to VLSI routing," *IEE Colloquium on Applications of OOP*, pp.6/1-6/4.



# Professor Created On-line General Biology Laboratory Course

**Arthur W. Bowman, Professor of Biology**  
*Biology Department & Science Education Center*  
Norfolk State University  
Norfolk, VA 23667  
[awbowman@nsu.edu](mailto:awbowman@nsu.edu)

**Abstract.** This paper will share the creation, implementation, and modification of an online college level general biology laboratory course offered for non-science majors as a part of a General Education Curriculum. The ability of professors to develop quality online laboratories will address a growing need in Higher Education as more institutions combine course sections and look for suitable alternative course delivery formats due to declining departmental budgets requiring reductions in staffing, equipment, and supplies. Also, there is an equal or greater need for more professors to develop the ability to create online laboratory experiences because many of the currently available online laboratory course packages from publishers do not always adequately parallel on-campus laboratory courses, or are not as aligned with the companion lecture sections. From a variety of scientific simulation and animation web sites, professors can easily identify material that closely fit the specific needs of their courses, instructional environment, and students that they serve. All too often, on-campus laboratory courses in the sciences provide what are termed confirmation experiences that do NOT allow students to experience science as would be carried out by scientists. Creatively developed online laboratory experiences can often provide the type of authentic investigative experiences that are not possible on-campus due to the time constraints of a typical two-hour, once-per-week-meeting laboratory course. In addition, online laboratory courses can address issues related to the need for students to more easily complete missing laboratory assignments, and to have opportunities to extend introductory exercises into more advanced undertakings where a greater sense of scientific discovery can be experienced. Professors are strongly encouraged to begin creating online laboratory exercises for their courses, and to consider issues regarding assessment, copyrights, and Intellectual Property concerns.

## 1.0 Introduction

In recent years, many institutions have begun to make more of their course offerings available online. This abundant availability of online courses is easily verifiable from even the most casual search for and perusal of such sites [1]. A few institutions have been offering such online courses and degrees for over twenty years, with one example being Phoenix Online. The majority of earlier online course and curricula offerings were designed for those seeking business degrees and career advancement training in technical fields. However, today there is hardly a discipline that does not have available online courses and degree programs, some even offering graduate and professional degrees. With this increased availability of online courses and degrees, nearly every intuition of Higher Learning must also provide online access to their academic programs so that they can try to meet the rapidly changing educational needs of both the non-traditional and traditional students. Resultantly, in order to remain competitive, the need to have an online availability appears to be an absolute requirement [2]. In this regard, many smaller institutions, along with those

that do not have what might be considered to be an optimum and/or robust infrastructure to support their educational technological need, must find expedient means of entering the online market that are not restricted by prohibitive cost and the need to have a large number of faculty members to develop and teach [3].

Typically, online offerings are organized and implemented through a college's office for distance learning or electronic learning. In that instruction of online courses can be very labor intensive, incongruence between the number of courses that are desired to be offered and the number of faculty members prepared to teach such courses can become a significant rate limiting factor for an institution's entry into and growth in the online market. As to be expected, most of the publishers of college textbooks have begun to provide a wider range of online ancillary materials, with some now offering versions of their curriculum materials and even course cartridges that approach what might be considered as a complete online version of the course(s) for which they publish text. Even though these materials are typically of high quality, they are also subject to limitations due to the fact that most of the authors have created materials that are

intended for the traditional print versions of their work. The author has noted that this is often addressed most immediately by having persons with expertise in creating educational technologies convert the more traditional formats into ones that are suitable for electronic distribution and engagement. Professors should be very careful not to have a too heavy reliance upon individuals having adequate technological capability but limited specific discipline background as they develop their online courses.

The author has found that course development considerations are driven in part by student demographics, the institutional setting, and the instructor's concept as to what the course's scope, sequence, and weighted emphasis areas should be. With a growing pressure to provide an increased number of online courses, there could be a natural tendency to opt for a utilitarian remedy by relegating the primary responsibility for creating and implementing online courses to entities other than those that are affiliated with the particular institution offering the online program. This approach raises questions of quality control, and the ability to create an educational environment that genuinely reflects the educational philosophy, mission, and objects of the institution, department, and faculty member. This paper addresses this particular circumstance by explaining how an online general biology laboratory course might be created via use of the instructor's existing technological skill set coupled with a relatively easy access to existing instructional resources such as animations, simulations, and video clips that are abundantly available without restrictions via the Internet

What this author has found is that two major logistical determinants in the delivery of an online laboratory course, or any other type of online course, are the faculty member's instructional philosophy or conceptual basis relative to how the entire online course will be approached, and the amount of time that can be dedicated to creating and implementing the course. In addition, it is also most helpful if the faculty member has at least a minimum familiarity with using an instructional support platform such as Blackboard. This paper will present the underpinning course philosophy, a description of the course's design, the process for identifying and selecting curriculum materials, a description as to how the web sites were identified and selected, a listing of useful web sites, mechanisms of enhancing the students' course engagement, how student performance and course effectiveness were assessed, and a conclusion.

## 2.0 COURSE PHILOSOPHY

This course that was developed, a general biological science laboratory course (BIO 100L-90) for non-science majors was established on the principle that students should be provided with an experience that would assess their knowledge, skills, and attitude regarding the course. In addition, it was also believed that their critical thinking skills should also be enhanced. Except for the S of KSA their knowledge (cognitive domain) and attitude (affective domain) were fairly easy to gauge. The skills component (psychomotor) was more difficult to assess because the majority of exercises use computer simulations and animations to provide the required laboratory experiences. Further, the author had as an additional objective the provision of opportunities for lessons to be extended past the classroom setting so that course information could be more connected to the students' lives. Other important aspects of the course philosophy are the belief that the course content has an importance to the students greater than just fulfillment of one science requirement within the University's General Education Requirements and that all of the students are capable of understanding the concepts to be taught. These last two philosophical considerations are felt to be important because they drive the inclusion of activities and information designed to make the students' reflect on the relationship of the course to their lives and develop the necessary confidence so that they can achieve at a high level. Confidence building and personal course relevancy are necessary because the majority of non-science majors frequently do have the necessary level of interest in the course that will allow them to learn maximally and to earn a good grade. It was also desired that the online course's activities would parallel, to the greatest extent possible, those exercises being conducted in the comparable on-campus general biology laboratory course. This was felt necessary because the on-campus laboratories are structured in such a manner that they accompany and complement the on-campus lecture component of the general biology course. An additional but somewhat less important consideration was the belief that the majority, if not all, of the students have the necessary computer access and skills to allow them to optimally navigate the Blackboard site used by the course and to successfully interact with the recommended Internet sites. This part of the course philosophy was relegated to a lower significance because any students seeking to enroll in any of the online courses are required to complete a survey that

indicates their level of preparedness to take the course from both computer access and skill capability standpoints. In addition, perspective students are directed to web sites that can serve as a guide for their entry into the online learning environment.

### 3.0 COURSE DEVELOPMENT & DESIGN

#### 3.1 Development

This particular course was not initially developed in the manner that is recommended by experts on Electronic Learning (E-Learning)/Distance Education because the author was recruited to create and teach this course nearly simultaneous with the beginning of the semester. The author was offered the opportunity to work with the course due to his history of teaching blended courses and frequent use of a variety of Internet sites to support the face-to-face courses that he taught on campus. Ideally, an online course would be developed well in advance of when it would be taught and not just prior to or during the course's implementation.

In an effort to facilitate the development of the course, those at the university responsible for E-Learning, working in consultation with the original professor that was to have taught the course, quickly reviewed a variety of commercially available virtual general biology laboratory course materials, selected one, and had its course cartridge installed on our Blackboard site. The materials selected were of high quality and were moderately user friendly in regard to site navigation and the clarity of instructions for their exercises; however, they were not topically aligned with our general biology curriculum. What is meant by not being aligned is that we were not able to find laboratory exercises in the virtual lab manual that cover the same topics covered in the on-campus laboratory. Not being user friendly meant that some of the exercises had insufficient directions and/or activities that did not work as designed. A third problem was that suitable worksheets to accompany the exercises were not always available. Therefore, the students could possibly successfully complete an exercise and yet still not grasp the full significance of what had been observed without adequate guided reflection. Due to the author's previous experience with curriculum design and engagement with students via the Internet, he was able to quickly develop a suitable curriculum for the online general biology course. The course syllabus is a blend of what is used for the on-campus laboratory, what is available from various Internet sites, and what the author has created.

#### 3.2 Design

The course was designed to provide three consistent components for all of the laboratory exercises: (1) background information (referencing the lecture course information); (2) simulations and animations; and (3) worksheets with questions that reinforced the concepts presented in the simulations and animations that connect the lesson to the students' everyday life experiences and that stimulated critical thinking. In that the author has taught the on-campus laboratory and lecture courses at Norfolk State University for approximately five years, he is very aware of how the Internet resources and other aspects of the laboratory should mirror the on-campus version of this course.

Usually, the first component of every laboratory exercise is the background, which is built upon the background information presented in the laboratory manual used for the on-campus course. The author modified the laboratory manual's introductory information so that it could provide additional insight about the exercise for students not present in the physical laboratory. Also, these modifications allowed for the inclusion of more information supporting the development of critical thinking skills. Table 1 presents the scope and sequence for the online laboratory exercises, all of which are aligned with those used for the on-campus laboratory course.

**Table 1: Schedule of Activities**

<b>Ex. 1 Scientific Articles assigned</b>
<b>Pre-Assessment Test</b>
<i>MLK Holiday (Monday)</i>
Scientific Articles' Summaries Due
<b>Ex. 2 Scientific Tools</b>
<b>Ex. 3 Scientific Measurements</b>
<b>Ex. 4 Illustrating Scientific Data</b>
<b>Ex. 5 Scientific Method</b>
<b>Ex. 6 Microscopy</b>
<b>Mid-Semester Assessment</b>
<b>Ex. 7 Cells and Tissues</b>
<b>Ex. 9 Cell Division (Mitosis)</b>
<b>Ex. 16 Principles of Heredity</b>
Body Systems Overview: <b>Ex. 11 Skeletal Ex.</b>
<b>12 Digestive and Respiratory</b>
<b>Ex. 13 Urogenital Ex. 14 Circulatory</b>
<b>Organisms and the Environment: Food Webs</b>
<b>Environmental Issues Part I</b>
<b>Environmental Issues Part II</b>
<b>Posttest / Final Exam</b>

This fall's version of the laboratory course will incorporate a set of hands-on activities that the online students can conduct in their homes. These activities will consist of observations and simple experiments that can be conducted with materials that are readily available in nearly every household. Previously, a few such activities have been used as a means to gauge students' willingness to carry out such inquiry activities. Examples of such activities was their using Metric and English measurement standards to determine how much water was used during a 5 minute shower as well as measuring the physical dimensions of various objects like compact discs, sandwiches, and articles of clothing. Another example was their investigation of enzyme activity by examining the influence of various physical and chemical variables on the activity of catalase, a ubiquitous enzyme that they can be easily obtained from raw white potatoes and different types of raw meats, especially liver. Catalase's substrate, hydrogen peroxide, is inexpensively obtained from any drug store. Many households typically have hydrogen peroxide in their medicine cabinets for use as an antiseptic on minor injuries to the skin. In that that no students expressed a lack of willingness or ability to carryout such simple hands-on activities, more will be included as a required part of selected exercises. Of course, any students that can not complete the hands-on activities will be provided with options that rely completely on computer access. One possible reason for a student's exemption from the requirement of conducting the hands-on activities might be their having special needs associated with a disability.

All exercises required the submission of completed worksheets within a set time frame. However, due to variety of reasons related to computer access and the students' schedules, the window for submitting completed exercises was usually extended until all of the students had submitted the assignments. In that the compilation of all completed exercises constituted a formal lab folder that was submitted at the end of the course, the ultimate deadline for completing the exercises was the due date for this lab folder. The primary impetus for their completing the exercises in a timely manner was created by their preparation for the online quizzes and exams that were required in a much more time stringent manner.

In some instances, after the successful completion of a few introductory assignments, students were given an opportunity to find, complete, and report on laboratory activities that they discovered for themselves on the Internet. Allowing students to find suitable web sites to accompany the assigned laboratory exercises proved to be another means of

getting the students to extend the assigned exercises. Having to critique the value of the site(s) was a critical thinking activity. Their selection and critiques of the web sites also provided another valuable means to assess student learning. In some instances, their recommended sites could be suitable for inclusion as a formal part of the course's exercises.

#### **4.0 SELECTING WEB SITES**

Web sites used for the exercises were located by using different search engines and were guided by key terminology and concepts associated with the weekly exercises. An abundance of web sites offering animations and simulations of varying degrees of sophistication were easily located. Information on the design of virtual courseware is also readily available [3]. The author, along with a few student volunteers, explored those web sites showing the most promise for use as course exercises. Once selected, the sites' accessibility via Blackboard was evaluated. Only those sites that presented the least amount navigational difficulty and the greatest amount of information were considered. Sites that would not be generally available to anyone searching the Internet or required and special permission for use were not selected. In that we did not bundle the exercises, or in any way use them for financial gain, issues of intellectual property violation were avoided. All sites referenced were presented as suitable recommendations for review in order for the students to complete the associated worksheets that were created by the author. Immediately upon visitation, the students could determine the authorship/development responsibility for the site(s). As mentioned previously, students were given the liberty to find other sites that supported fulfillment of a particular exercise's objectives. In that the quizzes and exams were focused primarily on the objective(s) of the exercises, the use of a wide range of animations and simulations could adequately prepare them for the assessments.

#### **5.0 ASSESSMENT**

The assessments consisted of graded submitted work sheets and the completion of quizzes and exams administered on Blackboard. The quizzes were made available following the completion of the weekly exercises with the two exams occurring at the midterm and end of the course. The quizzes consisted of 10 to 15 multiple choice questions with the exams given in a similar format and reflecting the same level of comprehension as the quizzes. As mentioned previously in regard to the submission of the completed worksheets, the

quizzes had some degree of flexibility in regard to their absolute completion deadlines. Due to the university's grade reporting requirements, students completed the two exams as scheduled. Fifteen minutes were allowed to complete each quiz, with exams having a one hour and fifty minute access time. These times were selected so that they would parallel those used for the on-campus laboratory courses. If the laboratory course was conducted as an independent experience without the need to be a mirror cohort, a different assessment protocol might have been used.

## **6.0 STUDENT PERFORMANCE**

Overall, the performance of the students in the online course exceeded that of the students in the on-campus course. Like other undergraduates [4], the students in this online course indicated on an end-of-course survey that they enjoyed the course and held the perception that it was effective. Somewhat to the students' surprise they reported that they spent considerably more time on the exercises that they would have for the on-campus course. In addition, students reported that they involved the family members in completing some of the activities. Also, in that the online assignments required more time to complete the assignments, the students tended to complete their work over several sessions. This is a good finding because it indicates the students were becoming more immersed in the course and extending their learning in such a manner that the input and thinking of others helped them establish their own authentic "home grown" community of learners. It was not unusual for some of the older and more "non-traditional" students to report that some of the class assignments were similar to those being carried out by their school-aged children. All reporting such a situation of family involvement spoke in very positive terms that seemed to indicate their interactions were benefiting both their children and themselves.

The relaxation of the deadlines for submitting completed exercises and assessments is thought to account for the apparent improved student performance because it allowed more time for course preparation. In contrast, the typical on-campus laboratory course finds students not fully engaged in the exercises but rather performing the exercises in a perfunctory manner. Also, the time limitation for the on-campus class period does not usually allow adequate time to fully complete all that would be desired for the laboratory period. For instance, the on-campus course must allow 10 to 15 minutes for the weekly quizzes, 10 to 15 minutes for introductory instruction, and approximately 90

minutes to complete the exercises. Because materials must be gathered and returned at the beginning and end of the laboratory period, little time is available for reflection and discussion of what has transpired during the period. If such reflection and discussions are deferred until the following laboratory period, that particular class meeting would have even less time for its required activities. Also, logistically, the on-campus course presents obstacles to the completion of missed assessments and exercise performance, especially for those labs where there is a moderate to extensive amount of laboratory preparation required. The on-line course provides many opportunities to complete missed assessments and exercises. Also, the on-line course exercises can easily and conveniently be repeated if a student lacks adequate understanding of the lesson or if problems arise as they attempt to complete the activities. Since students appear to be more willing to ask questions about the assignments there is a greater possibility that a more lasting understanding of the course content obtained. As is true with any course, there is a percentage of students that procrastinate and, in spite of extended opportunities, fail to complete the assignments. With the online course, the number of such delinquent students is significantly reduced due in part to the pre-qualifying that students attempting to enroll in the online course must undergo. We are currently conducting a quantitative analysis of assessment results and other course metrics that can provide a more objective comparison of the online and on-campus general laboratory biology courses. Findings will be used to make necessary modification to the online laboratory course that will be taught by the author this fall.

## **7.0 CONCLUSION**

The development and introduction of an online general biology laboratory course that parallels a comparable on-campus course has been successfully achieved by a professor in the NSU Biology Department working with assistance from the University's Office of E-Learning. The professor was not unusually adept with the use of educational technologies but his in-depth familiarity with the on-campus version of the course, along with his experience with curriculum development and use of Blackboard, has allowed him to create his department's first online biology course. By blending information from the currently adopted laboratory manual and from Internet web sites having ready access to scientific animation and simulations with the curriculum materials he has created a coherent course that closely mirrors the on-campus course already being offered. The

ready availability of technical support from the University's Office of E-Learning has allowed appropriate technological logistics to be applied and problems to be solved quickly. Formal The findings from this professor's experience can be of value to others seeking to develop quality, online laboratory courses that are tailored to their particular institutional setting. The course development was low cost to both the university and the students and was without many of the enrollment/access and operational imitations that are frequently experiences when trying to adopt published online curriculum materials. Though not optimal, and to some most objectionable [5], the development and implementation of this course by an individual professor was both effective and timely. Again, the cooperative relationship with the University's Office of E-Learning greatly benefited this effort. Also, interactions with professors in other disciplines that have experience in developing and teaching online courses were a positive influence.

assessments and anecdotal comments from the students indicate that an enjoyable and relevant high quality laboratory experience was provided.

## 8.0 REFERENCES

1. Stanford-Bowers, D. E. (2008) "Persistence in Online Classes: A Study of Perceptions among Community College Stakeholders," MERLOT Journal of Online Learning and Teaching, vol 4, no. 1.
2. Shepherd, C., A. Madelon, and M. Koeller. (2009). "The pedagogue & the Manatee: Danger of Extinction," Academic Exchange Quarterly, vol.13, no. 1.
3. Desharnais, R. A., and M. Limson. (2007), Designing and Implementing Virtual Courseware to Promote Inquiry-based Learning," MERLOT Journal of Online Learning and Teaching, vol. 3, no. 1.
4. Studkey-Mickell T. A., and B. D. Stuckey-Danner. (2007) "Virtual Labs in the Online Biology Course: Student Perceptions of Effectiveness and Usability," MERLOT Journal of Online Learning and Teaching, vol. 3, no. 2.
5. Oblinger, Diana G., and B. L. Hawkins.. (2006) "The Myth about Online Course Development: "A faculty Member Can Individually Develop and Deliver an Effective Online Course," EDUCAUSE Review, vol. 41, no. 1.

# The Umbra Simulation And Integration Framework Applied To Emergency Response Training

**Paul Lawrence Hamilton and Robert Brittain**

*ORION International Technologies, Inc., Albuquerque, NM, USA*

[phamilton@orionint.com](mailto:phamilton@orionint.com)

**Abstract.** The Mine Emergency Response Interactive Training Simulation (MERITS) is intended to prepare personnel to manage an emergency in an underground coal mine. The creation of an effective training environment required realistic emergent behavior in response to simulation events and trainee interventions, exploratory modification of miner behavior rules, realistic physics, and incorporation of legacy code. It also required the ability to add rich media to the simulation without conflicting with normal desktop security settings. Our Umbra Simulation and Integration Framework facilitated agent-based modeling of miners and rescuers and made it possible to work with subject matter experts to quickly adjust behavior through script editing, rather than through lengthy programming and recompilation. Integration of Umbra code with the WebKit browser engine allowed the use of JavaScript-enabled local web pages for media support. This project greatly extended the capabilities of Umbra in support of training simulations and has implications for simulations that combine human behavior, physics, and rich media.

## 1.0 INTRODUCTION

### 1.1 Problem and Significance

On January 2, 2006, there was an explosion at the Sago mine in Upshur County, West Virginia. Twelve miners were trapped. Eleven miners died. According to one mine rescue expert, in theory, they could have all walked out. The national outcry was overwhelming as the Sago story dominated all media for over a week. The Sago mine disaster demonstrated that mine accidents today have a potential for political and economic consequences far beyond their direct cost in money and lives.

Furthermore, although the US has the lowest fatality rate per ton produced of any major coal-producing nation, there are several factors that have the potential to make mining less safe in the future. The first is that much of the underground coal mining workforce is approaching retirement. [1] As they retire, the training and work habits that helped build this safety record will retire with them. The second is that underground miners must work under more challenging conditions as production expands. Closed mines have been reopened, like Sago, as the demand for coal has increased.

Coming after a half-year of dramatic increases in petroleum prices and rising concerns about energy and national security, the Sago disaster focused the nation's attention upon coal mining issues. Within weeks, the West Virginia Legislature passed sweeping new regulations affecting all underground coal mines in West Virginia and mandating a

number of new technologies for messaging and miner tracking. Dr. R. Larry Grayson, chair of the recently formed National Mining Association (NMA) Committee on Mine Safety, points out the limitations of this approach. "What miners really need", he says, "is a comprehensive system with scenario-based training and multiple options for survival"[2]

## 2.0 METHODS

### 2.1 Current Methods

"Table top" exercises are a common way to meet this need. People who could be in the command center role play the events of an emergency, often while sitting around a table. These exercises do a good job of helping participants visualize some of the implications of emergency procedures. However, they also miss much of the complexity inherent in mine emergencies and lack the impact that builds learning. Because only managers are involved in these exercises, they do not help miners understand how their actions must harmonize with those of mine management in order to mitigate the consequences of an emergency. "Mocks" provide more realistic training situations that do not have these shortcomings. Mocks are large-scale drills carried out at the mine itself or at a training and research facility. Not only does everyone involved with the mine participate, but in some cases, organizations from the nearby town even play roles in the simulation. Police, fire, and emergency

medical personnel all benefit from the opportunity to work with mine personnel. The mocks are realistic and inclusive, but they interfere with production and may not be practical at smaller mines.

## 2.2 Computer Simulation

Computer simulation has the potential to expand the realism of a “table top” exercise. Just as simulation has come to play a significant role in enhancing aviation safety, one would expect similar benefits in the field of mine safety. The Mine Emergency Response Interactive Training Simulation (MERITS), developed by the National Institute for Occupational Safety and Health (NIOSH) is a notable development [3]. In testimony before the U.S. Senate Appropriations Subcommittee on Labor, Health and Human Services, and Education, David E. Hess, Secretary of the Pennsylvania Department of Environmental Protection, stated that training, including MERITS, was a key factor in the successful rescue after a major accident at Quecreek, Pennsylvania. Hess testified that “the rescued miners have said the safety training they received helped them in several ways, first to warn the other miners to leave the rapidly flooding mine, how to share resources and protect themselves underground and to understand what rescuers above ground would be doing to rescue them.”

MERITS communicates events stemming from a simulated mine accident to a class through voice synthesis. The class may give instructions to miners and rescuers and order needed supplies. The class must piece together the nature of the emergency from the voices and execute a successful rescue of two trapped miners during a mine fire. The class can draw upon a multimedia library of information about the mine, including a personnel roster, emergency plan, and a detailed map of the mine and its contents. Although designed for supervisors who would direct a rescue operation, MERITS clearly gave these miners a useful understanding of the role their individual actions had to play in the rescue process.

Unfortunately, MERITS was designed to run on the computer systems of 10 years ago and is no longer compatible with current standard desktop security settings. It is a single scenario exercise that would require rewriting and recompiling portions of the code in order to change the scenario or reflect changes in Federal regulations. Furthermore, much of the code is written in MODSIM [4], a simulation language that is no longer widely used. We proposed to create a modernized, improved MERITS 2 to rectify these limitations.

## 2.3 Umbra Integration and Simulation Framework

The Umbra Simulation and Integration Framework, was a promising tool for a number of reasons. It allows program function to be easily changed, even during system execution. This is because, in an Umbra simulation, C++ modules call one another through easily edited Tcl scripts [4]; [5]. In Umbra, it is possible to “fine tune” behaviors without having to pause to recompile code. Standard interfaces facilitate the rapid construction of complex simulations. Scalability is enhanced through the use of the Worlds concept, by which each subprogram in the simulation is “aware” only of the information that the real entity would have access to.

Umbra was originally conceived at Sandia National Laboratories for the development of swarming and other cooperative behaviors and tactics for robots, including Unmanned Aerial Systems (UAS). Consequently, it lends itself to agent-based simulation and has a large library of routines that can be used for either embodied or non-embodied agents. However, it has not previously been applied to a training situation for a number of reasons. One is that there was no provision for integration with rich media, such as audio, video, and html documents. Another is that the lack of dockable windows made it impossible to cleanly page between simulation output and web pages. Finally, in the proposed mine rescue simulation, the computer would have to play the role of the mine telephone, communicating with the class through voice synthesis, a capability that had not been previously incorporated into Umbra.

## 2.4 Extending Umbra

The first requirement was to integrate the multiplatform Graphical User Interface toolkit, Qt 4.5 [8], into Umbra. By doing so, we would be able to not only quickly build effective user interfaces, but also to add the WebKit open source browser engine, the same code that powers Apple’s Safari browser. WebKit, unlike Explorer, provides a secure way to communicate between local and remote web content and local SQL databases using HTML5 database support. This would make it possible to replicate the media functionality of the original MERITS while complying with current security standards for networked desktops, thus solving one of the most important problems associated with the older version of MERITS. In order to use Qt 4.5, it was first necessary to recompile Umbra using VisualStudio 2008. The C-Space Toolkit [9], which provides detection of

geometric intersection of objects and plays a key role in path planning and the OpenSceneGraph output graphical libraries also had to be recompiled.

The existing libraries that were implemented into MERITS were the graphics interface for the visual elements, path planning, route placement interpolators and related generic utilities, collision detection, time-event generation, and event handling. New libraries generated were the Qt interface; GUI design, layout, and implementation; GUI event generation and handling; miner object behaviors; the voice interfaces; the wrapper for the existing NIOSH fire effects algorithm (GFire); and code to query GFire data based on position and object type.

## 2.5 Voice Synthesis

Because the current Umbra libraries had no provision for voice synthesis, it was necessary to add this capability. However, one of our constraints was that the resulting training application had to be distributable to noncommercial users without royalty payments. Consequently, we chose to use the open source code eSpeak [6]. However, the quality of the voices generated by this application was sufficiently low as to limit intelligibility. Consequently, we used phonemes generated by Mbrola [7], developed by the TCTS Lab of the Faculté Polytechnique de Mons (Belgium). Mbrola performs concatenation of diphones. Starting with a file of phonemes, durations, and a piecewise linear description of pitch, it uses a phoneme database to produce synthetic speech. We configured eSpeak to create appropriate phoneme files for Mbrola to convert to .wav output.

## 2.6 Agent-Based Simulation

Miners and rescuers were represented by autonomous agents based on preexisting robot simulation codes. We programmed these agents to carry out the following commands:

- Check (smoke, gasses, status)
- Close (door, regulator, etc.)
- Open (door, regulator, etc.)
- Evacuate (section or mine)
- Monitor (phone, fan, etc.)
- Stop monitoring (phone, fan, etc.)
- Move object(s)
- Move people (includes self)
- Pick up phone
- Turn on (belts, etc.)
- Shut off/turn off (belts, etc.)

Figure 1 shows how the menu selects telling a miner to pass on a command to make a gas check at a specified location and send the results back by mine phone.

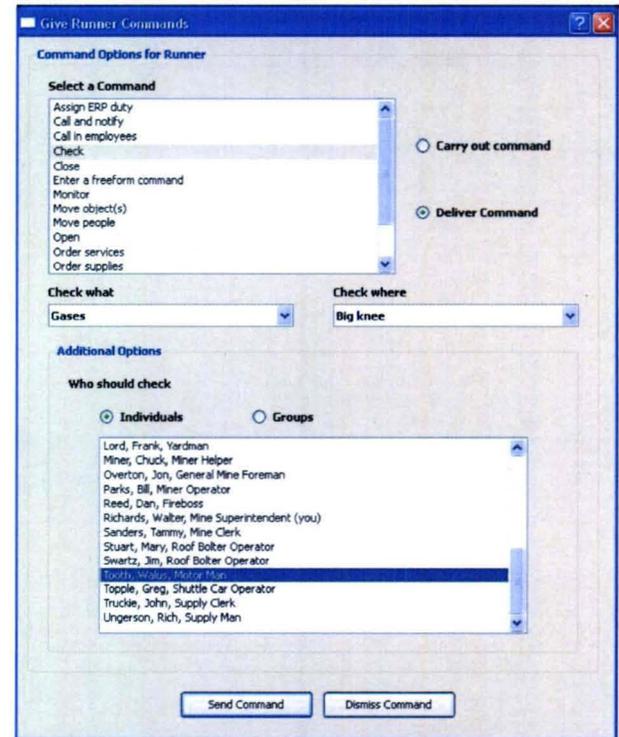


Figure 1: Command Menu

Rescuer agents are miner agents that are scripted to establish a Fresh Air Base for communications and enter the mine to find stranded miners using approved procedures. A third class of agents is the mantrip. These are trolley-like vehicles that travel in and out of the mine on command, carrying miners and supplies.

## 2.7 Path Planner

We used Umbra's existing RoadPathPlanner class, which implements the GBS, or graph-based search, algorithm. This algorithm was created to plan a path using city streets on irregular terrain. However, we found that appropriate reformatting of the files describing mine geometry allowed the code to simultaneously plan paths using both the underground tunnels and surface road. Also, the ability to randomize the speed taken on a given route is very useful for naturally separating miners on the same route – otherwise they would all be on top of each other and it would be impossible to determine who was there. It also prevents multiple "arrival" events from occurring simultaneously. In the event of a group evacuating the mine, the miner

agents can be instructed to travel together at the speed of the slowest member of the group.

### 3.0 RESULTS

#### 3.1 Extending Umbra

The combination of Qt and WebKit worked well, allowing the full integration of local and remote web media into the Umbra environment. JavaScript code not only made dynamic web pages possible, but also allowed the incorporation of persistent data storage. Simulation variables could be easily transferred to and from web pages, enhancing realism. One is having each miner tag in and out of the mine on a simulated tag board visible to the trainees. The persistent data storage capability allowed users to insert, edit, and delete notes labeled with the simulation time into the record of the session.

Figure 2 shows a JavaScript-enabled graphic that simulates a fan pressure chart. It is refreshed every minute from the simulation. As the simulation progresses, there is a roof collapse that is accompanied with a sudden pressure spike at the intake fan. The trainees can quickly page between a map of the mine and the fan pressure chart as they try to interpret the available data to determine what has happened.

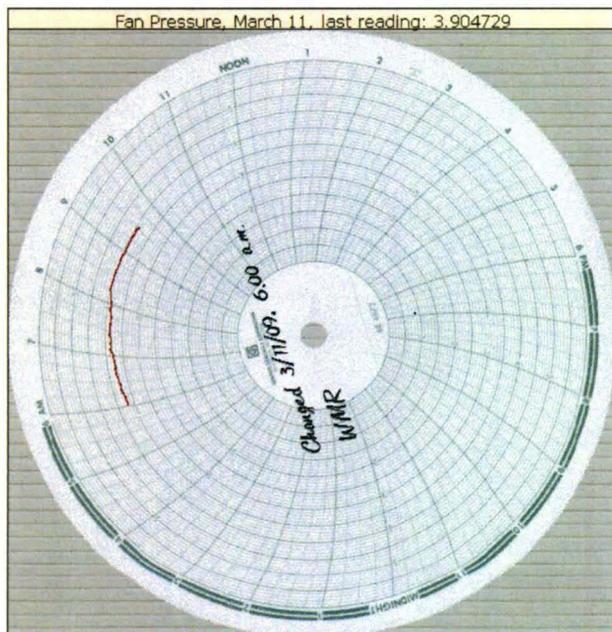


Figure 2: Web Image of Fan Chart

#### 3.2 Voice Synthesis

The combination of eSpeak and Mbrola produced voices that were highly intelligible. Depending upon the text supplied to the synthesis code, the resulting

voices vary in naturalness, primarily because there is no way to code syllable stress. Nevertheless, the voices were judged suitable for this application. For commercial applications, proprietary products such as AT&T Natural Voices could greatly enhance realism as they overcome the limitations of diaphone concatenation [11].

One problem we encountered was that the voices had to run in the same thread as the simulation. This led to the voices breaking up under some circumstances. The MERITS 2 system employs a stream of background "chatter" to simulate the effect of typical conversations over the mine phone system. When an event-driven simulation response produced speech during the playing of "chatter", the result was two voices speaking at the same time. In most circumstances, the voices simply played together. However, there were times when one of the voices was blocked and replayed, causing an unnatural "echo" effect. Both of these problems were mitigated by allowing the Umbra process to use more of the CPU. They might well persist in a more graphics-intensive environment than MERITS 2, however, and it would be important to investigate ways of multithreading such simulations.

#### 3.3 Agent-based Training

We demonstrated that by equipping miner, mantrip, and rescuer agents with a relatively small number of scripted behaviors that could be elicited by either simulation events or user commands, we could build a flexible simulation in which scenarios could be easily modified by editing Tcl scripts. We plan to extend the behavior by adding some more rules such as detection and reporting of smoke and fire, mortality if unprotected in a toxic atmosphere, and incorporation of fire, smoke, and gas concentrations into the path planner. We anticipate that these additions will further increase the training value of MERITS scenarios by fostering more complex emergent behavior for the trainees to deal with.

In Figure 3, three miner agents have noticed smoke in the passageways. The routine for the calculation of smoke density and gas concentrations is an adaptation of one developed by NIOSH. This example is from a developmental run and shows how a legacy program can be incorporated into an Umbra simulation. During a MERITS 2 exercise, the trainees are not able to see this view of the miners and smoke except when the instructor replays the exercise with "god mode" turned on. However, the trainees can call the miners on the nearest mine phone and ask them to report back airflow, smoke, and gas data.

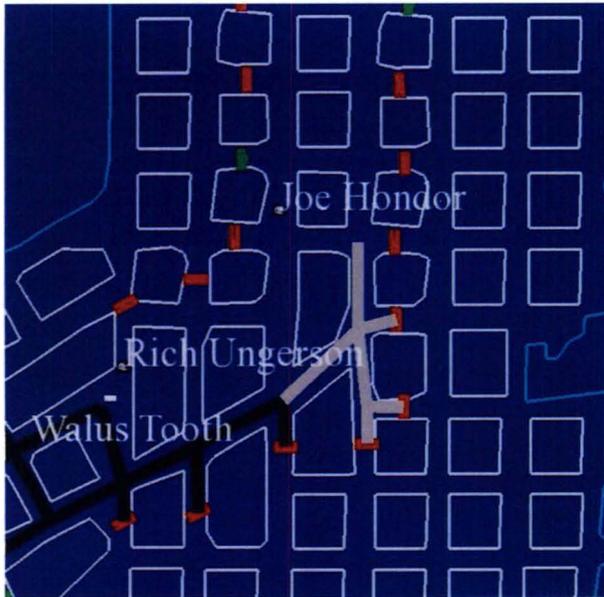


Figure 3: Miners Discover Smoke and Fire

### 3.4 Path planner

Once the C-Space Toolkit was incorporated into MERITS 2, it was found to work well. The original MERITS had scripted miners who followed a specified path regardless of whether or not a real miner could use it. With MERITS 2, miner and rescuer agents were able to discover a route between locations either on the surface or inside the mine whenever such a path existed. In fact, it was discovered that one of the work locations specified in the original MERITS was inappropriate when the path planner determined that there was no exit from it.

### 4.0 DISCUSSION AND CONCLUSIONS

As a result of this work, Umbra became much more useful in a scenario-based training environment. The ability to modify agent actions through the use of Tcl scripts turned out to be useful in the refinement of miner and rescuer behavior. The integration of dynamic web pages and audio into the Umbra simulation made for a more immersive training experience while also allowing users to research topics relevant to a better understanding of the simulated mine emergency. The potential uses of the Qt 4.5 integration include the incorporation of event-triggered video into Umbra simulations and the triggering of simulation actions from HTML links.

In a broader sense, our work with MERITS 2 shows how the benefits of simulator training, long recognized in aviation, can be extended into a variety of industrial environments to address the

more general problem of workplace and other civil accidents and emergencies. These situations include firefighting, ship damage control, and response to acts of terrorism. Such training simulations must incorporate accurate physics to work at all yet also deal with fragmentary knowledge of the situation, a potential for complex emergent behavior leading to unexpected consequences, and even a lack of definition as to what the objective should be (for example, should one try a risky plan to rescue all or a safer one that will only rescue some). These problems fall on a continuum stretching from linear "tame" problems to ill-defined "wicked" ones. Umbra, with its flexibility and low computational overhead, has been effective even at the "wicked" end of this continuum, and is now suitable in a training context.

### 5.0 REFERENCES

- 1 Fotta, B., and Bockosh, G., "The Aging Workforce: An Emerging Issue in the Mining Industry". Proceedings of the Thirty-First Annual Institute of Mining Health, Safety, and Research, 28-30 Aug 2000, Virginia Polytechnic Institute and State University, Department of Mining and Minerals Engineering, Blacksburg, VA, 2000, 16 Mar 2006, [http://www.cdc.gov/niosh/mining/pubs/pdfs/ta\\_wae.pdf](http://www.cdc.gov/niosh/mining/pubs/pdfs/ta_wae.pdf).
- 2 Smith, V, "Technology may not offer quick fixes to coal mine dangers", TMC NET, 24 Feb 2006, 16 Mar 2006 <http://news.tmcnet.com/news/2006/02/24/1405329.htm>.
- 3 Reinke, Dana, Launa Mallett, Michael J. Brnich, and Charles Vaught, "MERITS Volume 2: Trainer's Guide (Draft)", 2002, 16 Mar 2006, [http://merits.niosh.cdc.gov/merits/misc\\_files/MERITS%20Trainer's%20Guide.pdf](http://merits.niosh.cdc.gov/merits/misc_files/MERITS%20Trainer's%20Guide.pdf).
- 4 Mullarney, A. and J. West. (1987) Mullarney, A. and J. West. "ModSim: A Language for Object-Oriented Simulation. Design Specification." ACI Products, La Jolla, California. 1987
- 5 Gottlieb, E. J, R. Harrigan, M. McDonald, F. Oppel., and P. Xavier (2001) "The Umbra Simulation Framework," SAND2001-1533,

Sandia National Laboratories, Albuquerque,  
NM

- 6 Gottlieb, E. J., M. J. McDonald, F. J. Oppel, J. B. Rigdon, and P. G. Xavier (2002), "The Umbra Simulation Framework as Applied to Building HLA Federates," *Informs*.
- 7 Sourceforge, eSpeak text to speech  
<http://espeak.sourceforge.net/docindex.html>
- 8 MBROLA Project Development Team  
Faculte Polytechnique de Mons, MULTITEL-  
TCTS Lab,  
Initialis Scientific Park, 1, Copernic Ave, B-  
7000 Mons, Belgium.  
<http://tcts.fpms.ac.be/synthesis/mbrola.html>
- 9 Nokia Corporation  
Qt - A cross-platform application and UI  
framework  
<http://www.qtsoftware.com/>
- 10 Xavier, P.G., "Fast swept-volume distance for  
robust collision detection", (1997) IEEE  
International Conference on Robotics and  
Automation, 1997. Proceedings., Volume 2,  
Issue , 20-25 Apr 1997 Page(s):1162 - 1169  
vol.2 Digital Object Identifier  
10.1109/ROBOT.1997.614295
- 11 AT&T (2002) "AT&T Natural Voices Text-To-  
Speech Engines System Developer's Guide  
Server, Server-Lite, and Desktop Editions US  
English, Latin American Spanish, UK English,  
German, and French, Release 1.4"  
<http://www.naturalvoices.att.com>

# Joint Composable Object Model and LVC Methodology

**Richard Rheinsmith**  
Senior Enterprise Architect  
Novonics Corporation  
[rrheinsmith@novonics.com](mailto:rrheinsmith@novonics.com)

**Jeffrey Wallace**  
Chief Technology Officer  
Intelligent Integrations  
[jwallace@cots-llc.com](mailto:jwallace@cots-llc.com)

**Warren Bizub**  
Director, Joint Advanced Concepts  
Joint Warfighting Center, USJFCOM  
[warren.bizub@jfc.com](mailto:warren.bizub@jfc.com)

**Dr. Andy Ceranowicz**  
Technical lead,  
Joint Semi-Automated Forces (JSAF)  
Alion

**Dannie Cutts**  
Senior Computer Scientist,  
AEGIS Technologies Group Inc

**Dr. Edward T. Powell**  
Lead Architect Test and Training Enabling Architecture  
SAIC

**Paul Gustavson**  
Chief Scientist  
SimVentions, Inc

**Robert Lutz**  
Principal Staff Scientist  
Johns Hopkins University Applied Physics Laboratory

**Terrell McCloud**  
Chief Architect Common  
Training and Instrumentation Architecture (CTIA),  
Lockheed Martin

**Abstract.** Within the Department of Defense, multiple architectures are created to serve and fulfill one or several specific service or mission related LVC training goals. Multiple Object Models exist across and within those architectures and it is there that those disparate object models are a major source of interoperability problems when developing and constructing the training scenarios. The two most commonly used architectures are; HLA and TENA, with DIS and CTIA following close behind in terms of the number of users. Although these multiple architectures can share and exchange data the underlying meta-models for runtime data exchange are quite different, requiring gateways/translators to bridge between the different object model representations; while the Department of Defense's use of gateways are generally effective in performing these functions, as the LVC environment increases so too does the cost and complexity of these gateways. Coupled with the wide range of different object models across the various user communities we increase the propensity for run time errors, increased programmer stop gap measures during coordinated exercises, or failure of the system as a whole due to unknown or unforeseen incompatibilities. The Joint Composable Object Model (JCOM) project was established under an M&S Steering Committee (MSSC)-sponsored effort with oversight and control placed under the Joint Forces Command J7 Advanced Concepts Program Directorate. The purpose of this paper is to address the initial and the current progress that has been made in the following areas; the Conceptual Model Development Format, the Common Object Model, the Architecture Neutral Data Exchange Model (ANDEM), and the association methodology to allow the re-use of multiple architecture object models and the development of the prototype persistent reusable library.

## 1. INTRODUCTION

One of the fundamental difficulties involved with mixed architecture live, virtual, and constructive environments is the coordination and correlation of the data exchange models that enable state sharing and interoperability. The Joint Composable Object Model (JCOM) project was chartered to address this problem, and its progress to date is described in this paper. The principal results are the design of an eight phase process for data exchange model composition, and the creation and integration of the infrastructure required for its implementation. This paper will cover: the JCOM concept of operation including the composition process, application of conceptual modeling, the Architecture Neutral Data Exchange Model (ANDEM), and a discussion of the enabling metadata.

A quick detour into terminology is needed at this point to identify and define the key terms used in this paper, as they are interpreted broadly in the community. 'Data exchange model' (DEM) refers to the structure of the data used to communicate state and state changes between cooperating simulations. We use DEM instead of the more common term 'Object Model', such as the Federation Object Model (FOM) used by the HLA, to avoid confusion with software object models which include functional aspects. The term 'object' in distributed simulation originates from the fact that many messages used in a DEM are updates for the state of a real or simulated object such as a person or vehicle.

As such, the term 'object' is used in LVC environments in the common sense rather than

the software sense. Heavy use is made of the term 'component' in the general sense, indicating units that can be composed to create larger units, essentially reusable piece parts. DEMs are primarily composed of messages, and messages are composed of attributes and all of these are components of a DEM. Simulations are components of LVC federations. The LVC and distributed simulation community often refers to messages as classes and allows the use of inheritance to extend messages. We use 'conceptual models' to refer to abstractions of real or synthetic worlds that we want to include in our LVC environment. These abstractions include entities, processes, events, and states. 'Model' and 'representation' are used as equivalent terms; thus 'data exchange model' is equivalent to 'data exchange representation'.

In order to reuse DEMs efficiently, an easy way to find and retrieve them is necessary. An intelligent, searchable repository for DEMs must be built; allowing many new DEMs to be composed from existing ones. This should be a repository rather than a registry, because for efficiency the engineer should be able to retrieve the DEMs that match his search criteria immediately as opposed to a registry that tells him who to call to get the DEM.

This repository needs to contain the links between conceptual models of the domain and data exchange model components. Standard repository development techniques employing simplistic metadata descriptions are not sufficient to support semantic, concept-based queries. While the project intends to improve conventional metadata description initially, for the long term it will rely on

the open standards, methods, and technologies that have been developed for application areas such as the Semantic Web<sup>1</sup> to support semantically rich repositories and queries.

The essence of the JCOM project is to show how conceptual models of the domain can be used to organize and select data exchange model components which can be rapidly composed to create new LVC environments for training, experimentation, and other purposes. While this approach can be used to augment current federation building processes, only by leveraging semantic technologies can long-term breakthroughs in speed and accuracy of composition be achieved.

The basic JCOM concept of operation is for existing object models from the different LVC interoperability architectures to be parsed into an architecture neutral data exchange model format and stored in a repository.

## 2. COMPOSITION PROCESS

In this compositional development environment, LVC federation creation may be viewed as a constructive activity. A simulation of the desired functionality is composed from a set of existing LVC components. The LVC components are interfaced together via DEM components and the composition process produces a composite DEM that can connect all the LVC components required to implement the desired composite LVC federation.

In this Compositional Model of DEM development, the Accumulation, Evaluation, and Adaptation activities can be conceptually grouped into the process of Reuse. Feedback occurs between the Conceptualization and Reuse processes when conceptualization is influenced by the availability of components. This influence can be either in the form of repartitioning within the parameters of the original design, or of relaxing design constraints. If no candidate artifacts are found to satisfy the requirements, the designer may revise the conceptualization under a different design strategy to increase the opportunity for reuse, or may elect to implement the needed component (Prieto-Diaz 1987).

Feedback also occurs between the Reuse and Composition step when interface requirements dictate certain adaptations that may not be feasible with a particular artifact.

Standard development methodologies fail to support the compositional development model in three important ways. The compositional development processes of Accumulation and Evaluation are most tractable when object model definitions are independent, but this is often not the case. Most data and object modeling approaches lack support for representing the inter-object relationships that can capture this dependence. They only support two kinds of inter-object references, inheritance (IS-A) and client (HAS-A) relationships. From the standpoint of reuse, this is insufficient, because coupled components cannot be evaluated independently and the accumulation and evaluation processes take on a combinatorial aspect.

The second problem involves methodologies based on class reuse. Class-level reuse often occurs at too fine a granularity to be effective. It has been noted by other researchers that the advantages involved in reusing a component increase super-linearly as the component grows in size (Biggerstaff 1987). Thus a methodology that allows the reuse of larger components is more effective.

The third criticism of reuse support observes that object-oriented design methodologies only offer the developer *syntactic* support and only after the conceptualization, accumulation, and evaluation process has produced a candidate object for adaptation. Object-oriented methodologies offer this support through inheritance allowing the developer to "design by difference," adapting a chosen component through inheriting the candidate object into another class and specializing its structure. However, there is considerable intellectual challenge in the compositional processes of conceptualization, accumulation, and evaluation which need support.

This type of inheritance makes object and data model maintenance and evolution harder because the inheritance relationships violate the semantic model of the system. In recognition of the fact that object-level approaches are inherently insufficient to facilitate large-scale improvements in reuse, researchers have begun to look at higher-level abstractions and compositions; in the object-oriented community, these abstractions are referred to as design patterns and frameworks (Johnson 1988, Gamma et al. 1994, Whitehurst 1997), while non-object oriented systems research refers to these abstractions as reusable architectures.

---

<sup>1</sup> <http://www.w3.org/2001/sw/>

To begin development some method is needed to capture a conceptual model description that represents the training objectives in a format that can be algorithmically processed to support discovery and selection. The Joint Capability Areas (JCA) and Mission Essential Task List (METL) are good resources for building these conceptual models. Discovery and selection require that the DEM components (DEMCs) represented in ANDEM and stored in the repository are indexed by the same conceptual models that are employed to describe the training or experimentation tasks. A search mechanism that is capable of utilizing the conceptual model's semantically rich metadata developed in Phase One is required to match up the training objectives to the DEMCs.

Composition requires the ability to quickly and easily manipulate the inheritance and composition relationships of and between DEMCs is important. The ability to merge the graph structure representation underlying the DEMC is required, in addition to the ability to join, and potentially re-label the schemas of the DEMCs. This activity and capability is at the core of object and data exchange model composition.

Implementation of new OMCs/DEMs requires the ability to create new artifacts. As such, authoring and editing tools such as those commonly found in standard data modeling or object modeling tool environments would be desirable. The issue is the integration of such a capability into this composition environment and process. Typically such tools are stand alone and have limited import and export capabilities that permit the interoperation with other tools and processes. The principle capability required for the expansion process is the ability to quickly and easily send new and adapted DEMCs back into the repository to fulfill future task or mission requirements. The final phase of adding architecture specific information requires the ability to manipulate and augment the ANDEM data structures in a flexible manner.

The notion of an end-to-end Integrated Development Environment (IDE) for an object model composition process needs to be developed. Support tools, such as Protégé, GraphML, and Xerlin (for XML editing) should be combined through open source IDEs, which provide the necessary flexibility, through plug in creation and implementation.

### 3. CONCEPTUAL MODELING

**What this means for the war-fighter: Rapid and efficient federated simulation development.** Current technologies require considerable time to create a complex multi-architecture training and experimentation environment. As a result, a few established federated LVC environments are relied upon, where users are forced to make do with what exists, which means their requirements are not necessarily met. Conceptual modeling has been found to be a key part of the Object Model composition process. Conceptual modeling describes what is to be represented, the assumptions limiting those representations, and other capabilities needed to satisfy the user's requirements (IEEE P1730). In general, the conceptual model must identify the distinct entities or phenomena involved in the mission thread under consideration. It must also identify the actions of entities and the collaborative actions or activities that take place between them. These actors and common behavioral patterns are captured in a machine understandable form capable of triggering a semantic search.

Without a structured method for conceptual modeling, automating, or even semi-automating, the process of mapping between training and experimental objectives and the DEMs supporting them is challenging. As such, the ad-hoc processes for building federations will continue. In addition, the problem of finding and integrating LVC environment resources is made more difficult by the presence of multiple LVC integration architectures. There are frequently separate assets, subject matter experts (SMEs), DEMs, and repositories. Conceptual models are necessary to organize all these resources under a uniform schema that allows reuse independent of interoperability affiliation. Conceptual modeling can also help the LVC community to move away from the specialized terminology of M&S to that of the War fighter and live ranges. This will make M&S more understandable and useful to the War fighter.

Typically upfront conceptual modeling and analysis is limited and sometimes non-existent. Defining the scope of a project; understanding requirements and the way forward – pairing with what is needed to what is to be built and used is critical. Projects are often limited in applying conceptual modeling because of budget/personnel/resource constraints. Another obstacle is that using Object Model design for discussing capabilities with stakeholders may be “too big of a leap”. Without knowing where to go “to mine” (defining / integrating) reusable

conceptual models each project is left with the overwhelming task of defining everything from scratch.. Contracts rarely include contractual obligations to support Conceptual Model development, delivery and reuse. All these problems can be helped with the creation of a structured methodology for reusable conceptual modeling and sharing conceptual models can make better use of limited resources for conceptual model development. In general, Conceptual Modeling needs to be emphasized more fundamentally as an activity that not only assists in implementation, but also helps programmatic judgment.

In the initial phase of the JCOM project techniques such as the Object Modeling Groups (OMGs) Unified Modeling Language (UML) were employed to represent the conceptual models of a sampling of authoritative mission threads that could realistically be required as a part of a training exercise, experiment, or test and evaluation event. Additionally the Base Object Model (BOM) template specification (which employs UML sequence and activity diagrams) has been studied as an example of conceptual modeling. Some of the questions considered are:

- How do we extract the "piece parts" of existing object models that correspond to conceptual model components?
- How do we define the mapping from a conceptual model component to a corresponding object model component?
- How do we compose whole object models from a set of object model components?

Conceptual models for the JCOM effort have been captured using sequence diagrams The sequence diagrams provide the opportunity to identify common patterns, where these patterns can be extracted and potentially reused. Using the sequence diagram an entire mission thread can be examined and understood at the high level. The sequence can then be reviewed and decomposed to further explore the layers of sub-patterns that compose the mission thread. As the mission thread is further decomposed, the patterns may expose more details and variations such breadth of entity types (e.g. HQ at the mission thread layer includes Division, Brigade, and Company at the lower sub-pattern layers).

In addition to capturing the patterns of interplay, the conceptual model also identify types of conceptual entities required and their states providing a means to understand entity behavior that would need to be represented by a system or simulation. For example, in our original pattern of

interplay, three conceptual entities were identified: Target, Observer, and HQ. For the Observer, there are three states associated to this entity: Observe, Decide, and Communicate. These are states are reflected in the figure above.

#### 4. ARCHITECTURE NEUTRAL DATA EXCHANGE MODEL

**What this means for the war-fighter: The effective and efficient reusing of multiple architecture products regardless of service, component, or development tool.** The independent format allows mapping any interoperability architecture DEM to a common language. Once mapped, it will support reuse in multiple interoperability environments.

The question is not whether one object model can be mapped to another. The use of gateways to bridge the multiple LVC architectures is prima facie evidence that architecture specific DEMs can be mapped to each other. That problem is solvable by developers familiar with the models involved. The problem at hand is to accelerate and automate as much of the mapping process as possible. There is strong agreement that an Architecture Neutral Data Exchange Model (ANDEM) format for data exchange models can simplify the problem both for humans and machines. Humans can handle the problem for specific federations since the number of OMs that need to be translated between in a particular exercise environment is small. However, once the general problem is attempted, the large number of potential OMs necessitates a many to one approach rather than a many to many approach.

To create the ANDEM, JCOM started with the goal of extracting a single data exchange metamodel from the metamodels for TENA, HLA, DIS, and CTIA. This metamodel should be able to express the same data exchange capabilities as any TENA, HLA, or DIS object model. In the process there was disagreement as to whether ANDEM should be the intersection or union of these architecture specific metamodels. The intersection produces abstraction which is necessary for recognizing equivalence between different data exchange. For example, if transmission reliability were a necessary parameter of ANDEM, then there would never be equivalence between any HLA FOM component that uses reliable data transfer and the DIS Protocol Data Units – even though they may describe exactly the same world state.

However, once equivalence between two data exchange models has been established, there is the requirement for synthesis and implementation, which cannot be automated without capturing the specific implementation options of each protocol. Thus it was decided that in addition to the ANDEM, an architecture specific extension, or appendix, would need to be kept for each data model for use in building the translation between the formats. Yet, even then it was not easy to separate the conceptual pieces from the implementation on pieces.

architectures and are orthogonal to the common core.

Figure 2 depicts the Architecture Neutral Data Exchange Model (ANDEM) core metamodel. As stated previously the goal is to create a structure into which all of the constructs present in the four major LVC architectures can map into. This includes the three variations of HLA, the 1.3NG,

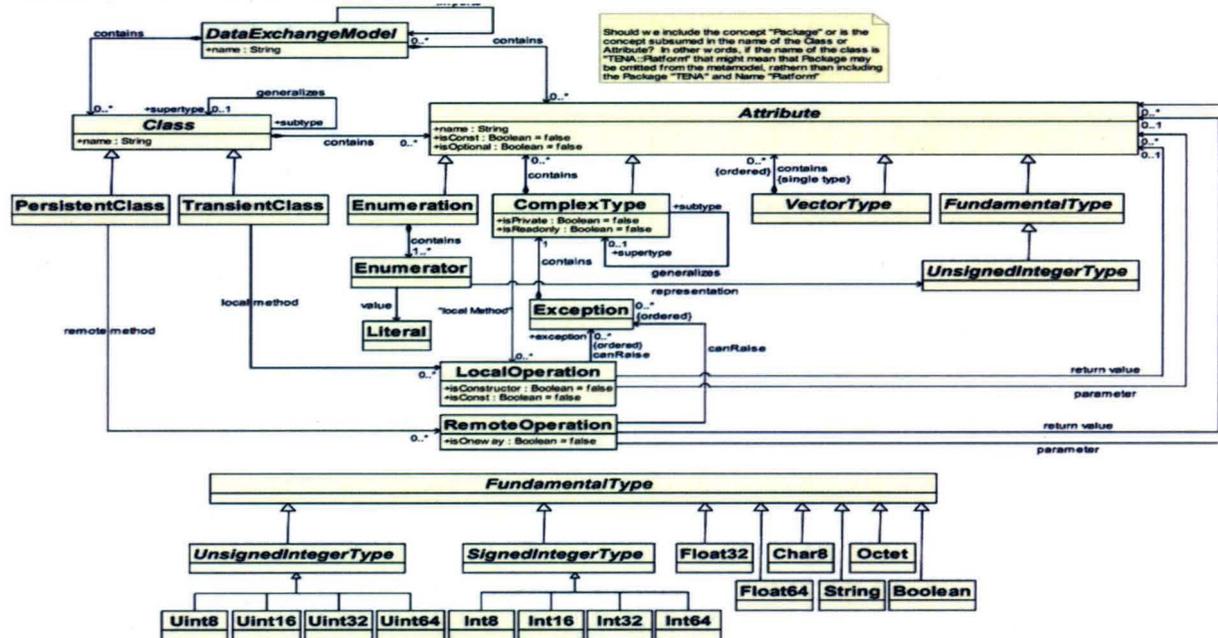


Figure 2. ANDEM Metamodel Prototype

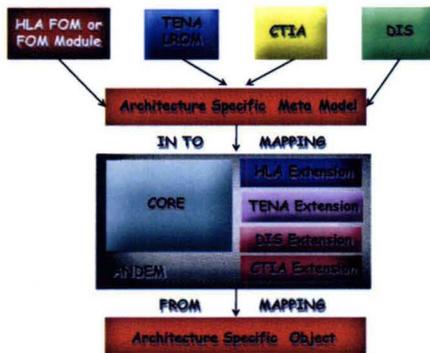


Figure 1. ANDEM Concept of Operation

Figure 1 illustrates the concept of operation of the Architecture Neutral Data Exchange Model (ANDEM). The ANDEM core represents the constructs that are common between each of the LVC architectures plus those constructs that materially affect a useful metamodel structure. This means including constructs that are not shared by all of the LVC architectures. The ANDEM architecture specific extensions represent those constructs that are unique to one of the LVC

IEEE 1516, and HLA Evolved. For example, the current ANDEM includes primitive data types, which lies in the intersection of all our three prototypical data exchange metamodels. Another question that arose was whether Live Architecture / Data models are adequately represented in the current set of the four LVC architectures under consideration.

The central feature of the ANDEM metamodel is the notion of a Class, which is the fundamental unit of representation. This concept exists in all four LVC architectures. The notion of inheritance is also present, even though it is not strictly present in all four, as is composition by inclusion (HAS-A relationships). The data exchange model being in several separate files is a construct that presently exists only in HLA Evolved and TENA. The notion is central enough that it is included in the ANDEM core metamodel, as such a construct would be difficult to retrofit.

The class construct has two sub-types, the persistent class and the transient class. The

distinction is made between classes that represent entities whose state persists over time (e.g., a platform or a sensor) and those that do not, such as weapon firing events or communication. The main feature of both types of classes is the ability to contain other classes or an attribute. As is indicated in the figure above the attribute construct has four variations: Enumerations, fundamental type, Vector type, and Complex types

The specialization of the fundamental type is standard across all of the LVC architectures. Notice that the vector type is configured to accommodate a single type of any attribute. Strictly speaking, the construct is not present in all of the architectures but its inclusion here is most natural.

## **5. METADATA AND COMPOSABILITY SERVICES What this means for the war-fighter: A simple but robust method for categorizing everything from a handgun to the newest air superiority jet.**

Making previously created artifacts easy to find and retrievable should help alleviate reimplementations due to the common expedient of "I can't find it so I'll just create a new one". In conjunction with a structured conceptual model and rapid reuse of multiple architectures in the LVC community, this technology will allow commanders at all levels to better understand and apply their tools.

There is agreement that ontologies as metadata, and related tools to create and maintain them offer great promise for the future in terms of composability support. Ontologies enable reduction in ambiguity of specification, and will reduce the current labor intensive processes required to create data exchange models. They will also permit and facilitate archiving and maintaining interoperability knowledge that is typically lost, or kept only by original designers.

## **6. SUMMARY**

This paper summarizes the JCOM project, along with the strategy and supporting technologies needed to achieve those goals. JCOM is considered just the first step in a longer and more extensive process to promote convergence and improve LVC interoperability. While object modeling is just one aspect of the broader LVC interoperability problem, the products and lessons learned from this project will provide a solid foundation for follow-on initiatives.

**Questions or comments** related to the conduct of this effort may be directed to the JCOM Program Manager, Mr. Warren Bizub ([warren.bizub@jifcom.mil](mailto:warren.bizub@jifcom.mil)).

## **ACKNOWLEDGEMENTS**

The authors thank Barbara and Jaclyn Hannibal for their support in the editing, graphics arts, and production support of this paper.

## **REFERENCES**

- Biggerstaff, Ted J. and Charles Richter. *Reusability framework, assessment and directions*. IEEE Software, 4(3):41-49, March 1987.
- Daniel Elenius, et al, 2007. "Purpose-Aware Reasoning about Interoperability of Heterogeneous Training Systems," In the *Proceedings of the 2007 International Semantic Web Conference*, Semantic Web Science Association, Busan, Korea, August 20, 2007
- Gamma, Erich, et al., 1994. *Design Patterns: Elements of Reusable Object-Oriented Software*, Addison-Wesley.
- IEEE P1730, "Distributed Simulation Engineering and Execution Process", *IEEE Draft Recommended Practice (V3)*, 2009.
- Johnson, Ralph and Brian Foote. *Designing reusable classes*. Journal of Object-Oriented Programming, 1(2):22-35, June 1988.
- Prieto-Diaz, Reuben and Peter Freeman. *Classifying software for reusability*. IEEE Software, 4(1):6-16, January 1987.
- Sowa, John, 2000. *Knowledge representation: logical, philosophical, and computational foundations*, MIT Press.
- Tolk, Andreas et al., 2008. "Implied Ontological Representation within the Levels of Conceptual Interoperability Model," *Intelligent Decision Technologies*, IOS Press, Vol 2, Number 1, 2008.
- Wallace, Jeffrey and Barbara Hannibal, 2008. "Software and Hardware System Integration and Intelligent Automation using Ontology-based Knowledge Representation Technology," In the *Proceedings of the 2008 International Conference on Artificial Intelligence*, World Academy of Sciences, Las Vegas, NV, July 14-16.

Whitehurst, R. Alan, 1997. "Using IMPORT to Implement Complex Behaviors in Simulations," In the *Proceedings of the 1997 Object-Oriented*

*Simulation Conference*, The Society for Computer Simulation, Phoenix, AZ, January 12-15, 1997.

# Using Game Theory to Qualitatively Analyze Global Relations

Corliss A. Tacosa, Visiting Assistant Professor  
Old Dominion University  
[ctacosa@odu.edu](mailto:ctacosa@odu.edu)

Taryn Cuper  
MYMIC, LLC  
[taryn.cuper@mymic.net](mailto:taryn.cuper@mymic.net)

**Abstract.** As the world becomes smaller, the importance of global education increases. Technological advancements have made news instantaneous. The advent of the Internet has made contact with another part of the world just a click away. Such achievements have put pressure on educational institutions to prepare students for the global community in which they already live. They must not only be taught about the world and the interactions among countries, but also, how the decisions are made which facilitate these interactions and what factors might contribute to the chain reactions which might erupt from such decisions. It is critical that students learn the skills of decision-making and analysis in order to be able to operate within a world where threats and crises abound. As tomorrow's leaders, today's students must be exposed to the global world and its mechanisms and learn the tools to navigate within it.

## 1.0 INTRODUCTION

In teaching international relations/politics, teachers must create an interest in their students for the global community and the issues that plague it. To do this, creative approaches must be incorporated in the classroom curricula to foster enthusiasm for the subject matter, encouraging the students to effectively learn the analytical skills necessary to become knowledgeable global citizens. Political theory offers such tools, but game theory specifically provides a methodology for developing an understanding of why entities execute certain courses of action and the costs and/or benefits associated with such decisions. Furthermore, through such theoretical modeling, insights are acquired as to the results of choices made when addressing an international issue or crisis. This new awareness provides students with the ability to develop astute and perceptive decision-making skills as well as enhances their knowledge and perceptions of global interactions. In the primary author's experience, this vehicle has resulted in great success, generating a deep-seated interest in the "ins and outs" of international relations/politics.

This paper outlines an approach to teaching international relations/politics, utilizing a qualitative approach to game theory. Student learn to model international incidents in the classroom for in-depth evaluation and appreciation. In this particular case, learners consist of American undergraduates enrolled in an International Relations or International Politics course. This

method can be used in any political science academic setting to promote understanding and awareness of interactions in the global community. In addition, it provides learners with several tools to further investigate what is happening in the world around them.

## 2.0 BACKGROUND

Game theory has been recognized as a tool used by academics to make sense of what is happening around them. In political science, it specifically allows for a simplification of actual world phenomena so that students can grasp the core reasons behind decision-making and interactions concerning international events as well as what precipitated such events. In many political science courses that use game theory as a tool for instruction, a quantitative approach is used, measuring the mathematical probability of actions to occur. However, this paper postulates that for undergraduates in basic political science courses, the meaning and purpose behind global relations is more easily understood through a qualitative discussion and modeling. In this way all students can participate whether mathematically inclined or not.

Through game theory, critical factors of a decision making process can be illustrated as well as isolated by scrutinizing the actions and reactions of several world players, weighing each dynamically in terms of costs and benefits. This is done by creating a hypothetical environment using real world issues and by simulating the decision-making process to create awareness and

understanding. R. Hrair Dekmejian labels such an analytical framework as “political physics.” [1] While specifically relating his approach through game theory to political violence, the characteristics of his “political physics” can be applied to any international issue analyzed through the game theory approach:

First, people [entities] behave rationally, with individuals and unitary groups acting in their respective perceived self interest. Second, relations between two individuals or groups are usually interactive, in that each party reacts to the other’s actions. Third, neither party is likely to “turn the other cheek” to what it sees as a wrongfully committed act by the other side. [1]

Through the enactment and use of various games, students learn the intricacies of interactions on the international level and the importance of making the right decision when addressing international conflictions. Students also learn the value of perception, as this particular attribute can mean the difference between a peaceful resolution and a devastating conflict.

This newfound awareness also introduces the concept of bias. Students learn that decision makers, while appearing objective, often have their own prejudices and agendas. When using game theory to analyze a given action, this knowledge about known decision makers can be determined and taken into consideration. It is more difficult to make conclusions about unknown decision-makers, but the skills used in playing the games can help students to become more astute in analyzing the actions of such people or states.

### 3.0 GAMES

With the international community in a state of anarchy, any number of adversarial situations can arise. Since all do not have the same characteristics, different ways to approach them must be studied. Thus, the following sections illustrate how the international environment can be modeled through three types of game theory, each with its own characteristics and purpose. The three are: Prisoner’s Dilemma; Zero-Sum; and Non-Zero Sum.

#### 3.1 Prisoner’s Dilemma

The first game introduced to students is patterned after the ageless Prisoner’s Dilemma Game. This

game demonstrates the key role of strategy and trust in making a decision when one’s survival is at stake. In the actual game, there are two prisoners accused of robbery. Interrogated in separate rooms, each prisoner must make a decision which will affect his respective future. This decision involves whether he can trust his fellow cohort to keep silent or whether his cohort will confess (defect) to his role in the respective crime. [2] Both prisoners’ futures depend upon the decision each will make after weighing the costs and benefits of either confessing or defecting. [2] The game is illustrated by a four square diagram with each quadrant representing a combination of the two possible actions, as shown in Table 1.

**Table 1:** Illustration of the Prisoner’s Dilemma quadrants

<u>Prisoner A</u> Confesses <u>Prisoner B</u> Confesses	<u>Prisoner A</u> Keeps Silent <u>Prisoner B</u> Confesses
<u>Prisoner A</u> Confesses <u>Prisoner B</u> Keeps Silent	<u>Prisoner A</u> Keeps Silent <u>Prisoner B</u> Keeps Silent

Students play the game by strategizing what each prisoner will do and why. This exercise begins the process of learning to consider the consequences of an action and how to come up with the best action with the least amount of punishment. In explaining this example to students, this type of strategic thinking is emphasized.

Extending this example to the international level, the Prisoner’s Dilemma type game is demonstrated by using the same four square illustration. To demonstrate how this game can enhance learning in international relations/politics, a historical topic is chosen where the actual outcome is already known. This approach is helpful because students may already know the result, but what they do not know is what might have happened if a major decision had been made differently. Usually, the example used concerns the relations between the United States and Japan during World War II with the actions being to fight or negotiate. Each quadrant has its respective attributes, i.e. fight/fight, negotiate/negotiate, fight/negotiate, or negotiate/fight. The quadrant representing the

actual occurrence (fight/fight) is crossed out, as the decision and its consequences are already known. Then each of the remaining quadrants is discussed, a process which is facilitated by the instructor. The goal is to force students to come up with attributes of what might have resulted if a given quadrant occurred. Based on brainstorming and analysis, alternative consequences for a decision within a given quadrant are eventually arrived at, often amazing students and expanding their recognition of the value of prudent decision-making in a crisis situation. Also, students realize that if a decision concerning this issue had been made differently, even if it was only one decision, their lives today might be very different. In this example, it is a given that the actors are rational entities; therefore, it is easier to predict their respective actions. This, however, is not always the case, especially when the game involves many of the actors on the international stage today.

A more current situation is then presented to the students; this involves a crisis in which the outcome is not known. For example, relations between Iran and Israel with the actions being attack or negotiate can be used. The four quadrant illustration is used with each quadrant representing attack/attack, attack/negotiate, negotiate/attack, and negotiate/negotiate. Through this game, with the outcome unknown, students become aware of consequences of certain actions of which they had not previously thought. They begin to see the key role of perception and the key role of leadership rationality. Predictive skills are honed as they begin to "connect the dots," seeing what could result on an international level or a regional level if two opposing actors engage in violent conflict. They also ascertain the key role of allies for both entities and how certain decisions would widen the scope of the original crisis, regardless of the entities' intentions. Dismissing their own biases as Americans, students learn that in the realm of international politics, the entire world feels the consequence of interactions between two actors and one decision can make all the difference.

### 3.2 Zero-Sum

The second game presented to the students is referred to as a Zero-Sum Game. In this game, there is one loser and one winner. The most easily understood example is the result of a war; one side wins and the other side loses. In World War II, the Allies won and the Axis Powers lost. However, there were ramifications of this outcome

of which most students are not aware, even within this game example. There is no question that the Axis Powers lost, but were the results of this positive or negative? Germany lost, was occupied, and then permanently divided. Previous German-occupied countries received their independence only to be "swallowed up" by the Soviet menace. However, West Germany, through western occupation was able to have its political culture changed from authoritarianism to democracy. So the question of one winner and one loser in this instance is relative.

Another example can be the end of the Cold War. Using an outline map of Europe, the various changes are illustrated. The United States and its allies defeated the Soviet Union and its allies. However, further ramifications occurred; the Soviet Union broke up with Russia emerging as its recognized successor; Eastern Europe was no longer under Russian control as each country became a truly independent state; Russia lost its influence over these countries and, in many cases, its economic ties with them. The loss of the Cold War not only cost the Russians politically and militarily, but also economically. It also caused a security dilemma as the Russians perceived a threat, even though there was no defined threat against them. As a result, Russia has continued to build advanced weaponry which in turn can be perceived as a threat to its neighbors. As with the previous exam, the terms winner and loser are relative.

Students are then assigned a current conflict, unresolved but in which there could be one winner and one loser. They then brainstorm and create a scenario of the ramifications of whomever wins and loses. Through use of this game, students' awareness of international issues and their ramifications are heightened. They begin to become aware that the international landscape that they take for granted has a history behind it. Winning or losing a war sets up new criteria for changing this landscape, perhaps to their detriment.

### 3.3 Non-Zero-Sum

The third game example is a Non-Zero-Sum game. This game does not have a winner or a loser, but it can illustrate the most dangerous of all situations on the international level. It is often equated with the well-known game of "chicken," where two cars race toward each other to see which one will veer first, thereby becoming the "chicken." The most common way to play this

game, on an international level, is through iterations with each iteration building on the previous iteration.

The students are first given an historic example. The nuclear arms race between the United States and the Soviet Union is a perfect example. Both countries are labeled as rational actors as both participated in the race out of fear for their respective security. When one country achieved one threshold in building a nuclear arm, the other matched it and raised the stakes. The race is illustrated through the dates and respective weapon explosions until the pinnacle was reached. With each side's nuclear arms becoming more and more dangerous and destructive; the pinnacle was reached with the explosion of the Soviet "Tsar Bomba" in October 1961, a 60 megaton explosion in the atmosphere that was the largest nuclear bomb ever exploded in the atmosphere. At this point, each side had to weigh the cost and benefits of continuing such testing. To the credit of the leaders involved, the race was partially stopped with the signing of the Limited Test Ban Treaty in 1963 and further harnessed with the signing of the Non-Proliferation Treaty in 1968.

With a Non-Zero-Sum game, both sides agree to stop or else they continue on until complete catastrophe prevents further continuation. This game shows students how unchecked build-ups from opposing sides can result in complete devastation or cooperation of rational minds, as happened in the nuclear arms race.

After a thorough discussion of the precariousness of build ups, attacks, or actions, students are assigned to develop their own non-zero sum game from current international issues. This provides awareness in them that if there are at least two negative actions of the same type concerning international entities at different times, then a third iteration may occur and how is the world suppose to react to it. As with the other types of games, students learn just how dangerous the world can be, heightening their perception of decisions made that might affect their own futures.

#### 4.0 Conclusion

Students today must be made aware of what is happening in the world. The instantaneousness of news makes it vital to have the rational and analytical foundations upon which to make and understand decisions. Game theory gives students a tool with which to test their respective

perceptions and further their awareness of the global world that is just at their doorstep.

#### References

- [1] Dekmejian, R.Hrair. *Spectrum of Terror*. Washington, D.C.: CQ Press, 2007.
- [2] Spruyt, Hendrik, *Global Horizons: An Introduction to International Relations*. Toronto, Canada: University of Toronto Press, 2009.

# Tuning Parameters In Heuristics By Using Design Of Experiments Methods

Arif Arin; Ghaith Rabadi; Resit Unal

*Department of Engineering Management and Systems Engineering, Old Dominion University  
aarin@odu.edu*

**Abstract.** With the growing complexity of today's large scale problems, it has become more difficult to find optimal solutions by using exact mathematical methods. The need to find near-optimal solutions in an acceptable time frame requires heuristic approaches. In many cases, however, most heuristics have several parameters that need to be "tuned" before they can reach good results. The problem then turns into "finding best parameter setting" for the heuristics to solve the problems efficiently and timely. One-Factor-At-a-Time (OFAT) approach for parameter tuning neglects the interactions between parameters. Design of Experiments (DOE) tools can be instead employed to tune the parameters more effectively. In this paper, we seek the best parameter setting for a Genetic Algorithm (GA) to solve the single machine total weighted tardiness problem in which  $n$  jobs must be scheduled on a single machine without preemption, and the objective is to minimize the total weighted tardiness. Benchmark instances for the problem are available in the literature. To fine tune the GA parameters in the most efficient way, we compare multiple DOE models including 2-level ( $2^k$ ) full factorial design, orthogonal array design, central composite design, D-optimal design and signal-to-noise (S/N) ratios. In each DOE method, a mathematical model is created using regression analysis, and solved to obtain the best parameter setting. After verification runs using the tuned parameter setting, the preliminary results for optimal solutions of multiple instances were found efficiently.

## 1. INTRODUCTION

One of the most important effects of the improving modern sciences and technologies is to enable us understand and model real life problems realistically and in more details. The natural outcome of this fact is the rapid increase of dimensions and complexity of the problems. With the growing complexity of today's large scale problems, it has become more difficult to find optimal solutions by using only exact mathematical methods. Due to the concern of efficiency in terms of the solution quality, the need to find near-optimal solutions in an acceptable time frame requires using heuristic approaches.

Heuristics are quite new approaches in the field of combinatorial optimization. A heuristic can be defined as "a generic algorithmic template that can be used for finding high quality solutions of hard combinatorial optimization problems" [1]. Heuristic approaches have already proved themselves in many large scale optimization problems by offering near-optimal solutions where there is no optimal solution found by other approaches. In many cases, however, most heuristics have several parameters that need to be "tuned" before they can reach good results. The accepted values of the parameters to be employed in the heuristics have considerably significant impact on both solution process and the solution itself. To obtain the best results, the problem then turns into "finding the best parameter setting" for the heuristics to solve the

problems efficiently and timely, which becomes an optimization problem by itself.

There are various methods used to find the best parameter setting in the literature. One-Factor-At-a-Time (OFAT) approach for parameter tuning is one of them; however, it neglects the interactions between the parameters that might change the whole solution process and quality of solution. Particularly, in terms of the interactions, Design of Experiments (DOE) methods are promising approaches and can be easily employed to tune the parameters more effectively.

In this paper, we seek the best parameter setting for a genetic algorithm to solve the single machine total weighted tardiness problem in which  $n$  jobs must be scheduled on a single machine without preemption, and the objective is to minimize the total weighted tardiness. Benchmark instances for the single machine total weighted tardiness problem are available in the literature.

## 2. DESIGN OF EXPERIMENTS (DOE)

To fine tune the genetic algorithm parameters in the most efficient way, we compare multiple DOE tools including 2-level ( $2^k$ ) full factorial design, orthogonal array design, central composite design, D-optimal design and signal-to-noise (S/N) ratios method. In each DOE method, a mathematical model is created using regression analysis, and solved to obtain the best parameter setting. After verification runs for other benchmark instances by using the tuned parameter setting,

DOE methods presented will be compared in terms of their solution qualities.

The single machine total weighted tardiness problem is used in this paper as a difficult problem to demonstrate the use of DOE for setting the optimization Genetic Algorithm (GA) parameters. In this problem,  $n$  jobs must be scheduled on a single machine where each job  $j$  has a given processing time  $P_j$  and a due date  $d_j$ . The tardiness  $T_j$  is defined as  $\max(0, C_j - d_j)$  where  $C_j$  is the job's completion time – a decision variable that is based on the job sequence. The objective function then becomes to minimize  $\sum_{j=1}^n w_j T_j$ . This is a well known problem to which benchmark problems are available. In seeking best parameter setting for the GA, we will be using a MS-excel Add-in called Evolver from Palisade [6].

We first implemented the problem in Excel spreadsheet, and used the first instance of 40-job benchmark problem to compare different DOE methods that are discussed below. The upper and lower levels for the GA parameters are given in the Table 1.

**Table 1:** Upper and lower levels for parameters

Level	Crossover Prob. (A)	Mutation Prob. (B)	Population Size (C)
Lower	0.01	0.06	30
Upper	1	0.2	100

The GA stopping criteria are to run for 10 minutes or to stop whenever the percent deviation of the solution from the optimal solution/best solution found so far becomes 0. In the following sections, we discuss and compare five DOE methods to see which method performs best.

### 2.1. 2-Level ( $2^k$ ) Full Factorial Design

2-Level ( $2^k$ ) full factorial design is the one of the most widely used DOE tools. In  $2^k$  full factorial design,  $k$  is the number of factors. After the lower and upper levels of the factors are determined, all combinations of these factor levels are studied simultaneously. In order to analyze the design, each factor should be linearly independent, which means the covariance of the factors should be equal to zero. The covariance is a measure of linear relationship between two random variables [5], and can be calculated by using the following equation where  $E(x)$  stands for the expected value of  $x$ .

$$\text{Cov}(x,y) = E(x,y) - E(x)E(y)$$

To calculate the covariance of the design, a transformation is needed from the lower and upper levels to (-1) and (+1), respectively. After these substitutions, because  $E(x,y) = 0$ ,  $E(x) = 0$ , and  $E(y) = 0$ ,  $\text{Cov}(x,y)$  is equal to zero. In orthogonal designs, the covariance is always equal to zero.

The  $2^k$  full factorial design is generated by using Yates algorithm. According to this algorithm; for the first factor, a column of (-1) and (+1) is written down with the signs alternating each time. For the second factor, the signs alternate in pairs, for the third factor they alternate in triple, and so on. To create the interactions columns, the levels of the each factor forming the interactions are simply multiplied.

In an experimental design, the number of experiments (rows) must at least be equal to the total degrees of freedom (DF) required for the study, as shown in Table 2.

**Table 2:** DF for  $2^k$  full factorial design with  $k=3$

Factors/Interactions	DF
Overall Mean	1
A,B,C	3 (2-1)
AB, AC, BC	3(2-1)(2-1)
ABC	1(2-1)(2-1)(2-1)
<b>Total</b>	<b>8</b>

One drawback of  $2^k$  full factorial design is rapid increase of the number of experiments while increasing the number of the factors ( $2^5=32$ ,  $2^8=256$ ,  $2^{10}=1024$ ). In 1940's, Fisher showed that meaningful results can be obtained by conducting a selected fraction of full factorial design which is called fractional factorial design,  $2^{k-p}$ , where  $p$  stands for the fraction portion.

Since there are 3 factors ( $k=3$ ) in our problem,  $2^3=8$  experiments are needed to run for 2-level full factorial design in Table 3.

**Table 3:** The  $2^k$  full factorial design with  $k = 3$

	A	B	C	AB	AC	BC	ABC
1	-1	-1	-1	1	1	1	-1
2	1	-1	-1	-1	-1	1	1
3	-1	1	-1	-1	1	-1	1
4	1	1	-1	1	-1	-1	-1
5	-1	-1	1	1	-1	-1	1
6	1	-1	1	-1	1	-1	-1
7	-1	1	1	-1	-1	1	-1
8	1	1	1	1	1	1	1

In each experiment, the factors, or parameters, are set and run according to the design. After the solutions  $Y$  obtained from the experiments are analyzed by implementing regression analysis, the mathematical model is derived. However, because  $R^2$  value of the model is 1.00, the term that has minimum effect (AB) is removed, and after running the regression analysis again, the following model with  $R^2 = 0.96$  is obtained.

$$Y = 954.37 + 1.63A + 4.88B + 9.13C - 9.13AC + 15.63BC - 15.63ABC$$

When this model is solved by employing Excel Solver to minimize  $Y$ , the parameter setting is found by using  $2^k$  full factorial design as

"Crossover = 0.01, Mutation = 0.2, Population = 30".

## 2.2. Orthogonal Array Design

The fact that effects of 3 or higher interactions tend to be insignificant, and therefore may be ignored, bring us to the fractional factorial design type named orthogonal array (OA) design where only main factors and 2-factor interactions are considered. A typical OA tabulation is in the form of  $L_a(b^c)$ , where a is the number of experiments, b is the number of levels, and c is the number of columns. Taguchi has formulated 18 standard OA designs [7], however they can also be modified by using various methods. To select the appropriate OA, first, number of factors and levels for each factor, and 2-factor interactions to be estimated must be defined. After calculating the DF, the OA with the closest number of the experiments to DF is selected. Interaction tables, or linear graphs developed by Taguchi are then utilized to follow the confounding pattern.

The DF of our problem for OA is 7 due to the absence of 3-factor interactions. The most appropriate OA for 3 factors, 2 levels and 7 experiments is  $L_8(2^7)$  which is created in Table 4.

Table 4: The OA design with k = 3

	C	B	BC	A	AC	AB
1	-1	-1	1	-1	1	1
2	-1	-1	1	1	-1	-1
3	-1	1	-1	-1	1	-1
4	-1	1	-1	1	-1	1
5	1	-1	-1	-1	-1	1
6	1	-1	-1	1	1	-1
7	1	1	1	-1	-1	-1
8	1	1	1	1	1	1

Because there are only 3 factors in the problem, all 2-factor interactions are included. As you notice, the  $2^k$  full factorial and OA designs with k=3 are about the same. The reason is that the number of factors is quite small, and increasing this number will clearly bring out the advantages of OA designs in terms of the number of experiments needed to study.

After implementing regression analysis for the OA design, the same mathematical model with  $2^k$  full factorial design is derived, except for the ABC term. This model has  $R^2$  value of 0.65. As in  $2^k$  full factorial design, Excel Solver gives the same solution set for A, B, and C, respectively, namely, the parameter setting for the OA design is again "Crossover = 0.01, Mutation = 0.2, Population = 30".

## 2.3. Central Composite Design

In  $2^k$  full factorial and OA designs it is assumed that the relationship between the 2-level factors is

linear. It is possible to increase the number of levels to 3 to capture the nonlinearity, however, it would be a bit controversial and none of the rules for the 2-levels would apply in those designs. Also, this would not be the best candidate for continuous factors like parameters used in heuristics. A better approach to cope with the nonlinearity and continuous factors could be Response Surface Method using the Central Composite Design (CCD) developed by Box & Wilson in 1950's [4].

CCD is a first-order design augmented by additional points that allow the estimation of the second-order mathematical model. CCD consists of a full factorial or fractional factorial design ( $2^k$  or  $2^{k-p}$ ), a center point (a row of zero's), and two points on axes for each factor at a distance  $\alpha$  from the design center which result  $2^k+2k+1$  or  $2^{k-p}+2k+1$  experiments in total. The distance  $\alpha$  is calculated as (number of experiments in fractional portion)<sup>1/4</sup>. It is possible to choose  $\alpha = \pm 1$ , which is then called face-centered design.

In our problem,  $2^3 = 8$  experiments for the fractional portion,  $2(3) = 6$  experiments for axial portion, and 1 experiment for center portion, total 15 experiments are needed. The distance  $\alpha$  is equal to  $(8)^{1/4} \cong 1.4$ . To be able to set the parameters for each experiment, the levels of the parameters must be coded for the values (-1.4, -1, 0, 1, 1.4). The complete CCD with k = 3 is shown in Table 5.

Table 5: Central Composite Design with k = 3

	A	B	C	AB	AC	BC	A <sup>2</sup>	B <sup>2</sup>	C <sup>2</sup>
1	-1	-1	-1	1	1	1	1	1	1
2	1	-1	-1	-1	-1	1	1	1	1
3	-1	1	-1	-1	1	-1	1	1	1
4	1	1	-1	1	-1	-1	1	1	1
5	-1	-1	1	1	-1	-1	1	1	1
6	1	-1	1	-1	1	-1	1	1	1
7	-1	1	1	-1	-1	1	1	1	1
8	1	1	1	1	1	1	1	1	1
9	-1.4	0	0	0	0	0	2	0	0
10	1.4	0	0	0	0	0	2	0	0
11	0	-1.4	0	0	0	0	0	2	0
12	0	1.4	0	0	0	0	0	2	0
13	0	0	-1.4	0	0	0	0	0	2
14	0	0	1.4	0	0	0	0	0	2
15	0	0	0	0	0	0	0	0	0

After implementing regression analysis for outcomes of the experiments, the following mathematical model with  $R^2 = 0.90$  is derived:

$$Y = 939.42 + 3.58A + 0.75B - 5.58C - 1.13AB + 5.38AC - 11.88BC + 5.94A^2 + 5.94B^2 - 11.31C^2$$

The solution set produced by Excel Solver is back coded to their real values, and the parameter setting found by CCD is "Crossover = 0.218, Mutation = 0.193, Population = 100".

## 2.4. D-Optimal Design

CCD is quite an efficient design especially due to adding the second-order nonlinearity; however, in some cases it may not be enough to understand the relationships between factors. And also, the number of experiences must be kept to an absolute minimum. If a design has an absolute minimum number of experiments, such design is called "saturated design". The minimum number of experiments can be calculated as  $(n+1)(n+2)/2$  where  $n$  is number of factors. Besides these advantages, if some experiments are infeasible, saturated designs can be still used by extracting these experiments from the design.

As some of the interesting features of saturated designs, unlike the previous DOE methods, they are not orthogonal and there are no degrees of freedom to test the accuracy of the model.

Saturated designs are constructed by applied D-optimality criterion. The following equation is the estimator of simple linear regression:

$$\hat{Y} = b_0 + \sum b_i x_i$$

where  $b_0$  is the intercept,  $b_i$  are the slopes. If this equation is written in matrix form, we have:

$$Y = XB + \varepsilon.$$

The set of design  $B$  can be estimated in the following form by applying the Least Square Regression method.

$$B = (X^T X)^{-1} X^T Y$$

A statistical measure of accuracy of  $B$  is the variance-covariance matrix:

$$V(B) = \sigma^2 (X^T X)^{-1}$$

where  $\sigma^2$  is the variance of the error.  $V(B)$  is a function of  $(X^T X)^{-1}$  and to increase the accuracy,  $(X^T X)^{-1}$  should be minimized. Statistically, minimizing  $(X^T X)^{-1}$  is equal to maximizing the determinant of  $(X^T X)$ . "D" in the term of D-optimal comes from the first letter of the word "determinant". There are some heuristics [2], and software [3] to come up with a design that maximizes the determinant of  $(X^T X)$ . To obtain more accurate results, D-optimal designs can be augmented by adding more experiments.

The absolute minimum of experiments for our problem is 10  $[=(3+1)(3+2)/2]$ , and the D-optimal design displayed in Table 6 is created by augmenting the design by 2 experiments.

Like CCD, the levels of the parameters must be coded for the values (-1, 0, 1). With the help of regression analysis, the following mathematical model is acquired:

$$Y = 92.48 - 0.63A + 2.62B + 8.37C - 6.38AB - 9.13AC + 15.63BC + 19.86A^2 + 23.11B^2 - 18.83C^2$$

After the solution set given by Excel Solver is back coded to their real values, and the parameter setting found by D-Optimal is "Crossover = 0.420, Mutation = 0.148, Population = 30".

Table 6: D-Optimal Design with  $k = 3$

	A	B	C	AB	AC	BC	A <sup>2</sup>	B <sup>2</sup>	C <sup>2</sup>
1	-1	-1	-1	1	1	1	1	1	1
2	-1	-1	1	1	-1	-1	1	1	1
3	-1	0	0	0	0	0	1	0	0
4	-1	1	-1	-1	1	-1	1	1	1
5	-1	1	1	-1	-1	1	1	1	1
6	0	-1	0	0	0	0	0	1	0
7	0	0	1	0	0	0	0	0	1
8	1	-1	-1	-1	-1	1	1	1	1
9	1	-1	1	-1	1	-1	1	1	1
10	1	1	-1	1	-1	-1	1	1	1
11	1	1	0	1	0	0	1	1	0
12	1	1	1	1	1	1	1	1	1

## 2.5. Signal-To-Noise (S/N) Ratio

DOE methods until this section are only based on one instance of our problem, and do not consider any information of other instances. The method of signal-to-noise (S/N) ratio can be defined as a performance measure that takes the mean and the variability into account, and give the ability to use information of other instances in seeking the best parameter setting. It involves two types of factors: control factors and noise factors. Noise factors cause variability which leads to loss of quality. There are three kinds of noise; outer noise, inner noise, and between product noise, or here can be defined as "between instance noise" is the main reason in applying S/N ratio method in our problem.

Generally, data analysis using S/N ratio ( $\eta$ ) can be performed to achieve three types of purposes: smaller-the-better, larger-the-better and nominal-the-best. Since our target is to minimize the total weighted tardiness for the single machine, the appropriate type of  $\eta$  is smaller-the-better. To minimize the sensitivity to noise factors, we maximize  $\eta$  which is calculated by the following equation [4].

$$\eta = -10 \log_{10} (\bar{Y}^2 + \sigma^2)$$

In addition to first instance, fourth and ninth instance are randomly selected as different "products". Unlike in other methods, instead of OA, D-optimal design in Table 6 is used in creating the experiments for each instance because of its advantages, and  $\eta$  is calculated as the outcome for each experiment. Three replications of D-optimal design for three instances increase the total number of experiments by 36  $(=3 \times 12)$ .

After applying the steps of D-optimal design for each instance, the regression analysis is run for to obtain the following mathematical model:

$$Y = 2.36 - 0.83A + 1.17B - 1.62C - 0.36AB + 0.52AC - 4.49BC - 9.27A^2 - 7.30B^2 + 6.07C^2$$

After back coding the findings in Excel Solver to their real values, the parameter setting found by D-Optimal are "Crossover = 0.465, Mutation = 0.157, Population = 30".

### 3. COMPARISON OF DOE RESULTS

After applying five DOE methods to find the best parameter setting for the single machine total weighted tardiness problem, the findings are summarized in Table 7. To test which method is most effective with this problem, these parameter settings are used in solving the first 20 instances for both 40-job problems in Table 8 and 50-job problems in Table 9 respectively [8].

Table 7: Parameter settings of DOE methods

DOE Type	Crossover Prob.(A)	Mutation Prob. (B)	Population Size (C)
2 <sup>k</sup> FF	0.010	0.200	30
OA	0.010	0.200	30
CCD	0.218	0.193	100
D-Opt.	0.420	0.148	30
S/N	0.465	0.157	30

To be able to compare the solutions for different instances, the percent deviation of the solution from the optimal solution/best known solution is used instead of the real outcomes of the experiments.

Because 2<sup>k</sup> full factorial and orthogonal array designs give same parameter settings for 3 factors, their common results share the first three columns.

Table 8: Comparison of Parameter settings for 40-job problem

Inst.	Orthogonal Array & 2 <sup>k</sup> Full Factorial Designs			Central Composite Design			D-Optimal Design			S/N Ratios Design		
	%Dev	Iteration	Time	%Dev	Iteration	Time	%Dev	Iteration	Time	%Dev	Iteration	Time
1	4.71	4918	00:01:22	1.86	18863	00:04:44	1.86	1895	00:00:16	1.86	31781	00:17:00
2	4.65	23416	00:03:55	0.08	22391	00:05:09	0.08	23344	00:02:19	4.65	3789	00:01:16
3	6.70	726	00:00:09	6.70	2850	00:00:38	6.70	2260	00:00:18	0	2826	00:00:38
4	1.29	5016	00:01:02	0	6853	00:01:11	0	4966	00:00:37	1.29	2607	00:00:33
5	0	5213	00:00:51	0	4780	00:00:51	0	15052	00:01:41	0	961	00:00:13
6	0	35635	00:04:50	0	34295	00:04:23	0	6226	00:00:50	0	14237	00:02:18
7	3.91	10038	00:01:39	0	85445	00:13:50	0	15282	00:01:37	3.91	2536	00:00:34
8	0	12210	00:01:48	0	26492	00:03:32	0	8461	00:01:06	0	4203	00:01:03
9	0	39811	00:05:15	0.59	46315	00:06:28	0.65	4881	00:00:40	0	35208	00:04:39
10	1.40	24753	00:03:25	0.04	88492	00:12:38	0.04	51283	00:08:53	0	51208	00:06:26
11	1.94	72938	00:11:37	0.01	96244	00:18:17	0	23267	00:03:58	0	23501	00:03:51
12	1.40	93314	00:14:41	0.68	96335	00:21:41	0	26996	00:05:14	0	29814	00:05:48
13	1.00	99337	00:15:36	0.74	98275	00:29:59	0.64	25285	00:03:31	0.64	29204	00:05:11
14	0.35	94223	00:19:43	0.77	88256	00:17:54	0.33	24304	00:03:11	0.33	29969	00:07:08
15	0.95	94728	00:34:23	0.91	89698	00:16:42	0.84	36549	00:12:00	0.09	76972	00:14:47
16	2.96	85614	00:22:42	3.55	94294	00:11:51	0.69	94079	00:30:19	0.82	77826	00:14:46
17	1.96	82399	00:19:53	2.94	96468	00:30:05	0.49	82127	00:29:51	0.37	94687	00:17:20
18	2.53	99367	00:28:07	2.61	95935	00:22:10	0.85	89000	00:26:04	0.66	93455	00:15:24
19	2.26	93600	00:11:32	3.25	50709	00:10:12	0.94	99959	00:31:45	0.82	97005	00:17:52
20	3.29	94366	00:10:28	3.91	79570	00:40:05	1.04	92765	00:17:28	1.54	90153	00:32:35
$\bar{Y}$	2.07	53581	00:10:39	1.43	61128	00:13:37	0.76	36399	00:09:05	0.85	39597	00:08:28
$\sigma$	1.85	39953	00:10:04	1.84	36776	00:11:03	1.49	34984	00:11:22	1.31	35674	00:08:37
0%	4			5			7			8		

According to data from the 40-job and 50-job problems, the S/N ratios and D-optimal designs seem to be the best two methods of the five DOE methods. While S/N ratios design could reach optimum solutions/best known in 8 instances for 40-job and 6 instances 50-job problems, D-optimal design could obtain them in 7 instances for the 40-job problems, and 6 instances for the 50-job problems. In terms of average percentage deviation, the number of iteration and running

time, they are also better than the other three methods. We might accept that S/N ratios design is slightly better than D-optimal design, but it needs three times more experiments than D-optimal design. Even though all DOE methods are completed based on the first instance of 40-job problem, the parameter settings found in these processes produce very close results to the 50-job problems which gives an idea about the robustness of the parameter settings.

**Table 9: Comparison of Parameter settings for 50-job problem**

Inst.	Orthogonal Array & 2 <sup>k</sup> Full Factorial Designs			Central Composite Design			D-Optimal Design			S/N Ratios Design		
	%Dev	Iteration	Time	%Dev	Iteration	Time	%Dev	Iteration	Time	%Dev	Iteration	Time
1	0	2198	00:00:43	0	4320	00:00:34	0	1095	00:00:20	0	3787	00:00:35
2	0.10	3836	00:01:00	0.10	4320	00:00:34	0.75	1079	00:00:14	0.75	2963	00:00:31
3	0	55007	00:12:09	1.39	2517	00:00:29	1.39	469	00:00:07	0	3500	00:00:34
4	0	2977	00:00:59	0	4306	00:00:34	0	1339	00:00:19	0	2802	00:00:30
5	5.67	7730	00:02:51	5.67	7645	00:00:57	5.67	5460	00:01:02	5.67	2313	00:00:27
6	0.57	83204	00:37:36	0.63	98371	00:09:48	0.09	8665	00:01:39	1.9	15136	00:01:52
7	0.53	86284	00:19:19	0.10	98266	00:10:33	0	16295	00:03:46	0	23888	00:03:10
8	1.59	45109	00:12:23	1.52	89622	00:10:42	0	10580	00:02:14	0.53	18688	00:02:14
9	0.34	15032	00:03:37	0.39	25102	00:02:39	0.39	8665	00:00:25	0.39	6699	00:01:53
10	0	69766	00:11:03	0	84602	00:12:46	0	7054	00:01:17	0	19074	00:02:38
11	3.37	92353	00:14:54	4.92	87583	00:13:10	0	26931	00:03:34	0.51	93839	00:13:46
12	4.15	85150	00:14:19	5.74	98291	00:14:20	0.93	17253	00:02:25	0.79	93465	00:12:35
13	3.49	89551	00:24:01	5.00	78979	00:11:31	0.26	23561	00:03:44	0.16	97849	00:11:58
14	1.12	86613	00:21:19	2.06	92867	00:14:09	0.52	14498	00:02:18	0.62	65048	00:10:35
15	1.16	96432	00:11:46	2.59	93136	00:13:10	0.92	27107	00:04:37	0	82041	00:12:49
16	5.50	95136	00:16:26	7.11	71351	00:15:37	0.02	85596	00:08:51	1.41	90719	00:14:29
17	6.67	97738	00:09:44	7.61	91351	00:13:28	0.15	87644	00:09:17	1.2	99878	00:15:23
18	5.58	96171	00:09:25	7.97	65694	00:10:25	0.53	97533	00:09:33	2.16	99272	00:11:57
19	3.51	95654	00:10:56	5.38	80849	00:12:35	0.19	86540	00:08:52	0.92	93995	00:18:40
20	5.31	78924	00:08:26	6.60	91293	00:16:13	0.48	88186	00:08:31	1.92	99114	00:12:08
$\bar{Y}$	2.43	64243	00:12:09	3.24	63523	00:09:13	0.61	30777	00:03:39	0.95	50704	00:07:26
$\sigma$	2.37	36918	00:08:57	2.94	38429	00:05:47	1.25	35523	00:03:26	1.31	42911	00:06:24
0%	4			3			6			6		

**4. CONCLUSIONS**

DOE offers a practical way to tune the heuristic parameters. Because the number of parameters, or factors, is not the same for all heuristics, it is important to select the right DOE method. Table 10 shows how fast the number of experiments increases for a small amount of increase in the number of factors with three levels. Other important issues include selecting the number of levels, values of the levels, the type of relationships between factors, and the cost of running of an experiment.

**Table 10: Number of experiments for 3-levels**

Factor (k)	3 <sup>k</sup> FF	OA	CCD	D-Opt.
4	81	27	25	15
5	243	81	43	21
7	2187	-	143	36

It should be noted that the same parameter setting produces different solutions for different instances although all instances are created from the same distributions. For the total weighted tardiness problem, the most effective methods turned out to be the D-Optimal and S/N Ratios Design, with the D-Optimal design requiring less runs.

This paper presented a structured framework on using DOE to tune optimization algorithm parameters. The weighted tardiness scheduling problem was used as a vehicle to demonstrate the

approach. The same approach can be applied to other problems.

**REFERENCES**

1. Birattari, M. (2009) "Tuning Metaheuristics: A Machine Learning Perspective", *Studies in Computational Intelligence*, Volume 197, Springer-Verlag Berlin Heidelberg.
2. Box, J.M., Drapper, N.R. (1974) "On minimum point second order designs", *Technometrics*, 16, 613-616.
3. JMP, Design of Experiments Guide (2007), SAS Institute Inc, Cary, North Carolina.
4. Montgomery, D.C. (2001) "Design and Analysis of Experiments", Fifth Edition, John Wiley & Sons, Inc., USA.
5. Montgomery, D.C., Runger G.C. (2003) "Applied Statistics Applied Statistics for Engineers", Third Edition, John Wiley & Sons, Inc., USA.
6. Palisade Corporation (2008), "Evolver, The Genetic Algorithm Solver for Microsoft Excel", Version 5.0, Ithaca, NY USA.
7. Phadke, M.S. (1989) "Quality Engineering Using Robust Design", AT&T Bell Laboratories, PTR Prentice-Hall, Inc., New Jersey.
8. <http://people.brunel.ac.uk/~mastjib/jeb/orlib/wtinfo.html>.

# An Exploratory Study of the Butterfly Effect Using Agent-Based Modeling

Mahmoud T. Khasawneh, Ph.D. Student, [mkhasawn@odu.edu](mailto:mkhasawn@odu.edu)

Jun Zhang, Ph.D. Student, [jxzhang@odu.edu](mailto:jxzhang@odu.edu)

Nevan E. N. Shearer, Ph.D. Student, [nesheare@odu.edu](mailto:nesheare@odu.edu)

Raed M. Jaradat, Ph.D. Student, [rjaradat@odu.edu](mailto:rjaradat@odu.edu)

Elkin Rodriguez-Velasquez, Ph.D. Student, [erodrig@odu.edu](mailto:erodrig@odu.edu)

Shannon R. Bowling, Assistant Professor, [sbowling@odu.edu](mailto:sbowling@odu.edu)

Department of Engineering Management and Systems Engineering  
Old Dominion University, Norfolk, Virginia 23529, USA

**Abstract.** This paper provides insights about the behavior of chaotic complex systems, and the sensitive dependence of the system on the initial starting conditions. How much does a small change in the initial conditions of a complex system affect it in the long term? Do complex systems exhibit what is called the "Butterfly Effect"? This paper uses an agent-based modeling approach to address these questions. An existing model from NetLogo© library was extended in order to compare chaotic complex systems with near-identical initial conditions. Results show that small changes in initial starting conditions can have a huge impact on the behavior of chaotic complex systems.

## 1. INTRODUCTION

The term the "butterfly effect" is attributed to the work of Edward Lorenz [1]. It is used to describe the sensitive dependence of the behavior of chaotic complex systems on the initial conditions of these systems. The metaphor refers to the notion that a butterfly flapping its wings somewhere may cause extreme changes in the ecological system's behavior in the future, such as a hurricane.

## 2. LITERATURE REVIEW

Lorenz is major contributor to the concept of the butterfly effect. He concluded that slight differing initial states can evolve into considerably different states. Bewley [2] talked about the high sensitivity observed in nonlinear complex systems, such as fluid convection, to very small levels of external force. Wang et al. [3] explored an approach for identifying chaotic phenomena in demands, and studied how a small drift in predicting an initial demand ultimately may cause a significant difference to real demand. Palmer [4] argued that a hypothetical dynamically-unconstrained perturbation to a small-scale variable, leaving all other large-scale variables unchanged, would take the system in a completely different direction, off the attractor. Yugay and Yashkevich [5] mentioned that the butterfly effect occurs in Long Josephson Junctions (LJJs) as described by a time dependent nonlinear sine-Gordon equation. This equation states that any alteration within the initial perturbation fundamentally changes the asymptotic state of the system. Social systems can also exhibit the butterfly effect

phenomenon. Several studies were dedicated to examine the butterfly effect which resulted from the format of the ballots in Palm Beach County, Florida during the presidential elections in the year 2000 [6, 7, 8]. The chaos emerging from the confusing configuration of the dual-column ballot is said to have caused 2,000 Democratic voters, a number larger than then Texas Governor George W. Bush's certified margin of victory in Florida, to cast their vote for another candidate instead of then Vice President Al Gore, which effectively made George W. Bush the 43rd President of the United States.

## 3. METHODOLOGY AND MODEL DEVELOPMENT

A modified version of the GasLab© model from the chemistry and physics library in NetLogo© was used as a basis for our analysis of chaotic complex systems. The following are the assumptions of the modified GasLab© model:

- A random seed sets the initial conditions (x-y coordinates, speed, heading).
- Two types of agents: particles and diablos (the two agents are identical, with the exception of name and color).
- The two types of agents only interact with their own type. They do not interact with each other.
- For the complete duration of the simulation, particles are in blue, while diablos are in red.
- Agents move in a random heading and certain speed until they collide with another agent of

the same kind. Upon collision, a new speed and heading for the participating particles/diablos are set.

- Particles and diablos bounce off a wall and continue moving in the box.
- A collision occurs if two particles or two diablos are on the same patch.

The criterion this paper adopted to test the existence of a butterfly effect is the average distance between particles and diablos at each tick. The formula for average distance (D) is shown below:

$$D = \frac{\sum_{i=1}^N \sqrt{(X_{Ai} - X_{Bi})^2 + (Y_{Ai} - Y_{Bi})^2}}{N}$$

(1)

- N: number of particles/diablos in the system (for N = 10, 20, 30, ..., 100, 200, 300, ..., 1000) thus obtaining 19 configurations in total
- $X_{Ai}$ : the x-coordinate of particle i
- $X_{Bi}$ : the x-coordinate of diablo i
- $Y_{Ai}$ : the y-coordinate of particle i
- $Y_{Bi}$ : the y-coordinate of diablo i

The reason behind using different numbers of particles/diablos is to examine the effect of the size of the population on the speed at which the butterfly effect emerges in the model.

The modeling methodology was divided into five phases:

1. Creating a model with two random systems: The original GasLab© model had only one agent; particles. Another agent, diablos, was added to the model with identical behavior patterns to those of particles. Because two random systems are created, particles use a different random seed than diablos (for speed, positioning, and heading).
2. Creating a model with same settings: After the establishment of a model with two random systems, we then modified the model again so that particles and diablos use the same random seed, thus sharing the same speed, initial positioning, and heading, creating a model with same initial settings. This model was the basis to test the hypothesis of the butterfly effect. The rationale this paper used to have a slider bar to incorporate extremely small changes to the heading of a single agent, which we randomly chose to be a

diablo. Our assumption is that this small change is an equivalent to a butterfly “flapping its wings.”

3. Automation setup for data collection: the code was adjusted to avoid the need for doing manual runs and to enable collection of sufficient data to test the existence, or lack thereof, of a butterfly effect in the system. All data points were exported to a text file.
4. Statistical analysis: a macro was developed to organize the data into an excel spreadsheet in order to make the graphs and plot confidence intervals.
5. Visual demonstration of divergence: a separate model was created to visually demonstrate the point at which particle i and diablo i diverge after starting in the same position. For the purpose of visual demonstration, when the distance between particle i and diablo i is equal to half-patch, their colors are changed to black to symbolize the transition from identical systems to random systems.

In recognition of the importance of systems' complexity in determining the existence of a butterfly effect, we ran our model(s) for different configurations of particles/diablos as mentioned earlier. Moreover, to reduce the effect of randomness and obtain confidence intervals for our results, each configuration was run for 30 times, each run consisting of 10,000 ticks.

#### 4. DISCUSSION AND CONCLUSIONS

Figures 1, 2, and 3 show the average distance between two random systems for 10, 500, and 1000 particles. The graphs illustrate that regardless of the number of agents we have in the model, the average distance tends to fluctuate around 38. It is evident that the variance decreases as the number of agents increases.

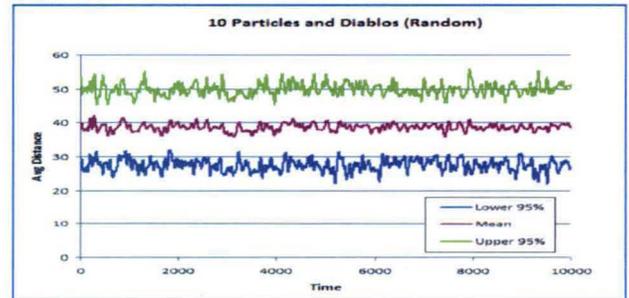


Figure 1: Average Distance for 10 Particles/Diablos

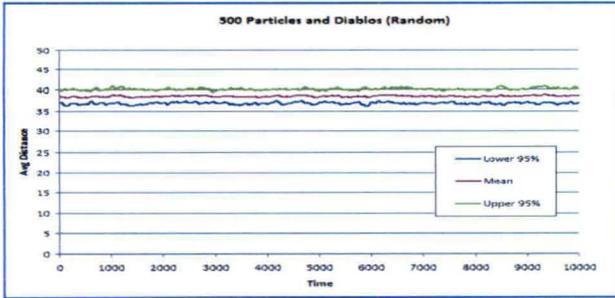


Figure 2: Average Distance for 500 Particles/Diablos

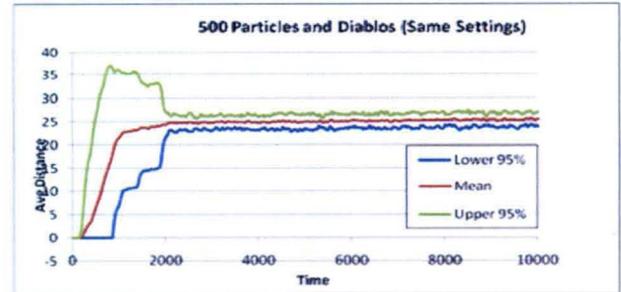


Figure 5: Average Distance for 500 Particles/Diablos (Same Settings)

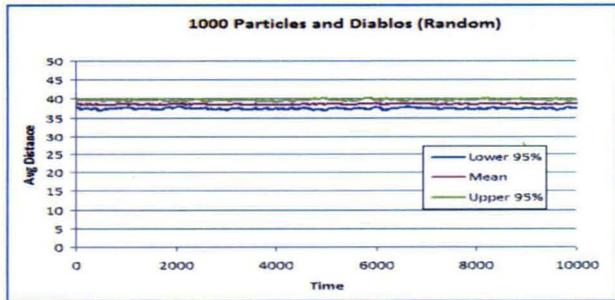


Figure 3: Average Distance for 1000 Particles/Diablos

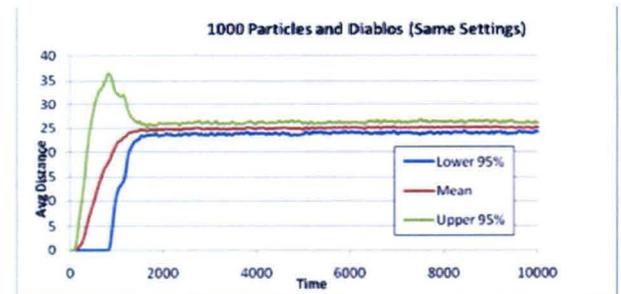


Figure 6: Average Distance for 1000 Particles/Diablos (Same Settings)

Figures 4, 5, and 6 show the average distance between two systems with same settings for 10, 500, and 1000 particles/diablos, with the exception of making a change to the heading of one diablo to examine the butterfly effect. An observation is that as the number of particles/diablos is increased in the model, the system diverges quicker. Similarly to the random systems, as the number of particles/diablos is increased, the variance decreases.

Figures 7, 8, and 9 show the difference in average distance between the model with random systems and the model with same settings. In all cases, the model with same settings will approach the same conditions as the model with random systems. Although the model with same settings quickly approaches the behavior of the model with random systems, it takes longer to actually reach the same average distance of 38. Moreover, as the number of particles/diablos increases, it takes longer to reach the same average distance for model of random systems.

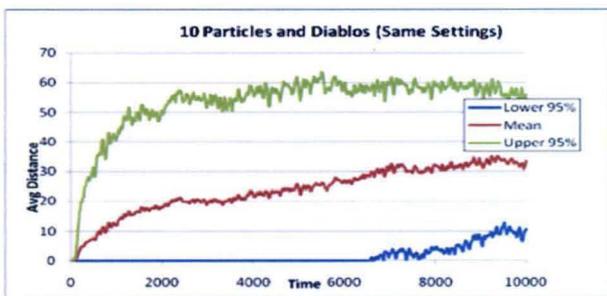


Figure 4: Average Distance for 10 Particles/Diablos (Same Settings)

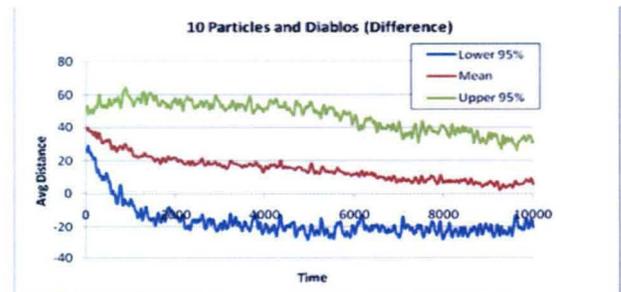


Figure 7: Average Distance for 10 Particles/Diablos (Difference)

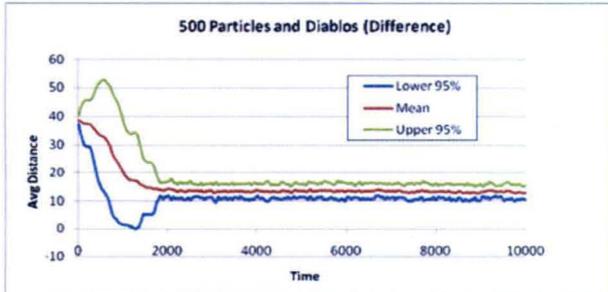


Figure 8: Average Distance for 500 Particles/Diablos (Difference)

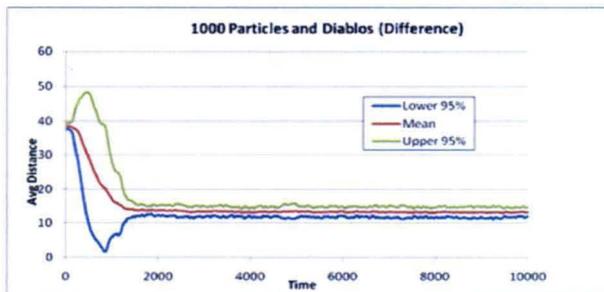


Figure 9: Average Distance for 1000 Particles/Diablos (Difference)

Collectively, the results of this paper demonstrate that there is a butterfly effect in chaotic complex systems. In fact, as complexity increases, the butterfly effect emerges quicker but takes a longer time to completely replicate the model with random systems. Therefore, an additional experiment was run to determine how long it takes for the model with same settings to completely replicate the model with random systems. As evident in Figure 10, the results of the model indicate that it actually takes about 2 million ticks to completely replicate the model of random systems, for the setting of 1000 particles/diablos.

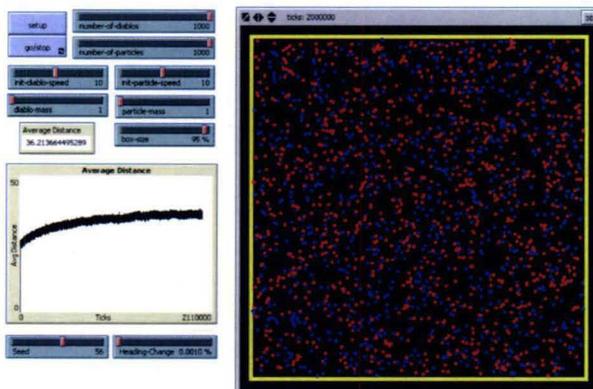


Figure 10: Simulation Results for 1000 Particles/Diablos After 2 Million Ticks

The most important implication of this study is that chaotic complex systems can actually exhibit the butterfly effect. Scientists, from all disciplines, should acknowledge that when studying complex systems and complex phenomena, reaching an understanding of the current state of the systems can be traced back to a small perturbation earlier in the system's life cycle.

## REFERENCES

1. Lorenz, E., 1963, "Deterministic Nonperiodic Flow", *Journal of The Atmospheric Sciences*, 20, 130-141.
2. Bewley, T. R., 1999, "Linear control and estimation of nonlinear chaotic convection: Harnessing the butterfly effect", *Physics of Fluid*, 11(5), 1169-1186.
3. Wang, K. J., Wee, H., M., Gao, S., F., and Chung, S., L., 2004, "Production and inventory control with chaotic demands", *Omega Journal of Management Science*, 33, 97- 106.
4. Palmer, T. N., 2005, "Quantum Reality, Complex Numbers, and The Meteorological Butterfly Effect", *Bulletin of the American Meteorological Society*, 86(4), 519-530.
5. Yugay, K. N. and Yashkevich, E. A. 2006, "The Bradbury Butterfly Effect in Long Josephson Junctions", *Journal of Superconductivity and Novel Magnetism*, 19(1/2), 1-9.
6. Sinclair, R. C., Mark, M. M., Moore, S. E., Lavis, C. A., and Soldat, A. S., 2000, "Psychology: An electoral butterfly effect", *Nature*, 408, 665-666.
7. Wand, J., Shotts, K., Sekhon, J., Mebane Jr., W., Herron, M., Brady, H., 2001, "The Butterfly Did it: Aberrant Vote for Buchanan in Palm Beach County, Florida", *American Political Science Review*. 95 (4), 793-810.
8. Wu, D.W., 2002, "Regression Analyses On the Butterfly Ballot Effect: A Statistical Perspective of the US 2000 Election", *International Journal of Mathematical Education in Science and Technology*, 33, 309-317.

# Random vs. Combinatorial Methods for Discrete Event Simulation of a Grid Computer Network

D. Richard Kuhn\*, Raghu Kacker\*, Yu Lei\*\*

\*National Institute of Standards and Technology, \*\* University of Texas, Arlington  
kuhn@nist.gov, raghu.kacker@nist.gov, ylei@uta.edu

**Abstract:** This study compared random and  $t$ -way combinatorial inputs of a network simulator, to determine if these two approaches produce significantly different deadlock detection for varying network configurations. Modeling deadlock detection is important for analyzing configuration changes that could inadvertently degrade network operations, or to determine modifications that could be made by attackers to deliberately induce deadlock. Discrete event simulation of a network may be conducted using random generation of inputs. In this study, we compare random with combinatorial generation of inputs. Combinatorial (or  $t$ -way) testing requires every combination of any  $t$  parameter values to be covered by at least one test. Combinatorial methods can be highly effective because empirical data suggest that nearly all failures involve the interaction of a small number of parameters (1 to 6). Thus, for example, if all deadlocks involve at most 5-way interactions between  $n$  parameters, then exhaustive testing of all  $n$ -way interactions adds no additional information that would not be obtained by testing all 5-way interactions. While the maximum degree of interaction between parameters involved in the deadlocks clearly cannot be known in advance, covering all  $t$ -way interactions may be more efficient than using random generation of inputs. In this study we tested this hypothesis for  $t = 2, 3,$  and  $4$  for deadlock detection in a network simulation. Achieving the same degree of coverage provided by 4-way tests would have required approximately 3.2 times as many random tests; thus combinatorial methods were more efficient for detecting deadlocks involving a higher degree of interactions. The paper reviews explanations for these results and implications for modeling and simulation.

## 1 Background

A number of studies have shown combinatorial methods to be highly effective for software testing (e.g., [3], [6], [16], [8]). The effectiveness of combinatorial test methods rests on the observation that a significant number of events in software are triggered only by the interaction of two or more variable values. By including tests for all 2-way, 3-way, etc., interactions, the test set should be able to detect events that occur only with complex interactions. The complexity of discrete event simulation suggests that, as with software testing, combinatorial methods may be effective for finding events triggered only by rare multi-way interactions of input values. In this paper, we compare the effectiveness of combinatorial versus random generation of inputs in a grid computer network simulation for finding configurations that lead to deadlock.

The key enabler in combinatorial testing is a *covering array* that covers all  $t$ -way combinations of parameter values, for a desired strength  $t$ . Covering arrays are combinatorial

objects that represent interaction test suites. A covering array,  $CA(N;t,k,v)$ , is an  $N \times k$  array, where  $k$  is the number of variables, and  $v$  is the number of possible values for each variable such that in every  $N \times t$  subarray, each  $t$ -tuple occurs at least once, then  $t$  is the *strength* of the coverage of interactions. Each row of a covering array represents a test, with one column for each parameter that is varied in testing. Collectively, the rows of the array include every  $t$ -way combination of parameter values at least once. For example, Figure 1 shows a covering array that includes all 3-way combinations of binary values for 10 parameters. Each row corresponds to one test, and each column gives the values for a particular parameter. It can be seen that any three columns in any order contain all eight possible combinations (000, 001, 010, 011, 100, 101, 110, 111) of the parameter values. Collectively, this set of tests will exercise all 3-way combinations of input values in only 13 tests, as compared with 1,024 for exhaustive coverage.

The primary goal in the simulation is to study the behavior of the system with different input configurations. For example, a network simulation

may investigate the effect of configurations on packet rate, delay, or potential for deadlock in the network, just as a production line simulation may study the effects of changing line speed, interconnection between workstations, and buffer size on the number of items that can be produced per hour.

	Parameters									
Test 1	0	0	0	0	0	0	0	0	0	0
Test 2	1	1	1	1	1	1	1	1	1	1
	1	1	1	0	1	0	0	0	0	1
	1	0	1	1	0	1	0	1	0	0
▪	1	0	0	0	1	1	1	0	0	0
▪	0	1	1	0	0	1	0	0	1	0
▪	0	0	1	0	1	0	1	1	1	0
	1	1	0	1	0	0	1	0	1	0
	0	0	0	1	1	1	0	0	1	1
	0	0	1	1	0	0	1	0	0	1
	0	1	0	1	1	0	0	1	0	0
	1	0	0	0	0	0	0	1	1	1
Test 13	0	1	0	0	0	1	1	1	0	1

Figure 1: 3-way covering array for 10 parameters with 2 values each

In this study we compare random and combinatorial testing of a network simulator, to determine if these two test approaches produce significantly different deadlock detection in the simulation. Using deadlocks as events of interest makes evaluating program responses straightforward and unambiguous. Numerical results such as packet rates or delays are not considered, but could be the subject of a future investigation. The two test modes – random or combinatorial – are compared using a standard two-tailed t-test for statistical significance.

## 2 Experimental Evaluation

This work investigates the hypothesis that combinatorial test suites will detect significantly more deadlocks than random test suites of the same size, for interaction strengths of  $t = 2, 3, 4$ .

**Independent and Dependent Variables:** The independent variable in this study is the type of testing used, either  $t$ -way combinatorial or random. The dependent variable is the number of deadlocks detected.

**Subject Application and Test Suites:** Software under test for the experiment was Simured [13], a multicomputer network simulator

developed at the University of Valencia. The software is available in C++ and Java versions, for both Linux and Windows. The core command line code (not including user interface or graphical display) consists of 2,131 lines of C++. Simured provides a simulation of the switching and routing layers for a multicomputer, allowing the user to study grid computer configurations to investigate the effect of topologies and configurable parameters on routing, timing, and other variables of interest. We used the C++ command line version of this software, compiled with gcc and run on 64-bit processors under Red Hat Enterprise Linux V4. No modifications were made to the Simured software.

Table 1: Simured configuration parameters and test values used

Parameter	Values
1	DIMENSIONS 1,2,4,6,8
2	NODOSDIM 2,4,6
3	NUMVIRT 1,2,3,8
4	NUMVIRTINJ 1,2,3,8
5	NUMVIRTEJE 1,2,3,8
6	LONBUFFER 1,2,4,6
7	NUMDIR 1,2
8	FORWARDING 0,1
9	PHYSICAL t, f
10	ROUTING 0,1,2,3
11	DELFIPO 1,2,4,6
12	DELCROSS 1,2,4,6
13	DELCHANNEL 1,2,4,6
14	DELSWITCH 1,2,4,6

Simured provides a set of 14 parameters that can be set to a variety of values in a configuration file that is read by the simulator. Parameters and possible values used are shown in Table 1. The total number of possible configurations with these parameter values is  $3.1 \times 10^7$ . Larger values are possible for a number of parameters, but would require extensive run time on a large system.

**Evaluation Metrics:** Test suites were evaluated according to the number of deadlocks detected. We also compare the percentage of  $t$ -way combinations covered for the random test suites of equal size, and determine the number of random tests needed to provide 100% coverage of the respective  $t$ -way combinations. (By definition, a covering array provides 100% coverage of  $t$ -way combinations.)

**Threats to Validity:** Clearly there are limitation on the extent to which these results can be generalized to other applications. While previous comparisons of combinatorial and random testing focused on fault detection, this study evaluates these methods with respect to deadlock detection in a simulation. Some implications of this difference are discussed in the analysis of results, in Section 4.2. A second difference is the nature of the software under test. Simured is a small but complex program that is not assumed to have characteristics similar to other application domains. Network simulation requires extensive calculations for statistics such as packet transmission rates and delays, and is not directly comparable to other types of software.

While the issues raised above should be considered in evaluating results, we believe that the experiment has identified a number of factors that can be usefully considered when deciding whether to use random or combinatorial testing for a particular problem.

### 3 Testing Procedure

Covering arrays that include all  $t$ -way combinations for  $t = 2, 3,$  and  $4$  were generated using the IPOG algorithm [11], which produces compact test suites. Test suites for the configuration shown in 0 included 28, 161, and 752 tests for  $t = 2, 3,$  and  $4$  respectively. Random test suites matching the sizes of the 2, 3, and 4-way combinatorial test suites were produced using the standard C library *rand()* function, producing one test at a time with a call to *rand()* for each variable value. In generating random test sets, the *rand()* function was initialized with a call to *srand()* to seed the pseudo-random number generator from the system clock. From these tests, configuration files were generated for Simured and the command line version of Simured invoked with each configuration file.

Each test set was executed for 500, 1000, 2000, 4000, and 8000-packet simulation runs. For combinatorial testing, one test suite run was conducted for each of the five packet counts and three interaction levels (28, 161, and 752 tests, for a total of 4,705 simulations). Random generation produces a different test set with each test generation run. For random testing, eight runs at each combination of packet count and interaction level were conducted (37,640

simulations), and the average deadlock detection calculated.

## 4 Results and Analysis

### 4.1 Test Results

Results for the two test modes were compared with a standard t-test for paired samples. Table 2 shows the number of deadlocks detected using tests produced from IPOG versus the average number of deadlocks detected with an equal number of randomly generated tests. Values for random test detection represent the average of eight runs with randomly generated tests at each combination of interaction level and packet count. Table 3 gives the two-tailed probability of a difference between the numbers of deadlocks detected by combinatorial and random testing.

**Table 2:** Deadlocks, combinatorial vs. random

Deadlocks Detected – combinatorial						
$t$	Tests	Packets				
		500	1000	2000	4000	8000
2	28	0	0	0	0	0
3	161	2	3	2	3	3
4	752	14	14	14	14	14
Average Deadlocks Detected – random						
$t$	Tests	Packets				
		500	1000	2000	4000	8000
2	28	0.63	0.25	0.75	0.50	0.75
3	161	3.00	3.00	3.00	3.00	3.00
4	752	10.13	11.75	10.38	13.00	13.25

**Table 3:** t-test results for difference between random and IPOG generated tests

Interaction strength	Two-tailed probability
2	.0035
3	.1778
4	.0235

For pairwise testing ( $t = 2$ ), combinatorial testing detected slightly fewer deadlocks than an equal number of random tests, and the difference is statistically significant. At interaction strength  $t = 3$  the difference between the two test methods is not statistically significant. At  $t = 4$ , however, the covering arrays produced by IPOG detected significantly more deadlocks than an equal number of random tests. In the next section we consider some possible reasons for the variation in effectiveness of these two test methods. Two important considerations should be noted about the difference in deadlocks detected:

combinatorial methods found more deadlock configurations, but also consistently found 14 deadlocks for the most complex (4-way) interactions, while there was a great degree of variation among the random configurations.

## 4.2 Analysis of Results

In considering explanations for the results, we first note that there can be a number of differences between the simulations conducted in this work and software testing in other application domains. In many applications, such as databases or web applications, different parameter values may result in different execution paths within an application, but the amount and complexity of processing is often similar for many different inputs. Network simulation, by contrast, may exhibit wide variations in processing depending on whether the input configuration is a small network of simple topology, or a large, complex one. This difference was observed in widely varying run times (not reported in this paper), and may also contribute to the distribution of deadlocks detected at the three interaction levels. Previous work (see Section 1) has found that increasing values of  $t$  detect progressively fewer faults, even in cases where combinatorial testing performed no better than random tests. Pairwise testing ( $t=2$ ) often detected 70% to more than 90% of faults, while 3-way tests found roughly 10% to 20% of faults, and 4-way to 6-way tests typically detected less than 5%. This distribution is essentially reversed for the Simured testing (see Table 2), with 0%, 18%, and 82% of deadlocks detected at  $t=2, 3$ , and 4 respectively. This result is not unexpected. Faults can be triggered by combinations of any of the variables in a program. Even though a large set of variables may be directly or indirectly involved in triggering deadlocks, the set can be expected to be much smaller than the total number of variables in a program. With deadlocks occurring in roughly 2% of simulation runs, larger test sets would be expected to locate more deadlocks.

In addition to the "reverse" relationship between deadlock detection and interaction strength, another interesting finding was that pairwise tests detected slightly fewer deadlocks than the same number of random tests. Careful analysis shows that there is in fact a combinatorial explanation for this result, which we discuss in the remainder of this section.

Because a significant percentage of events can only be triggered by the interaction of two or more variables, one consideration in comparing random and combinatorial testing is the degree to which random testing covers particular  $t$ -way combinations. Any test set will also cover a certain proportion of possible  $(t+1)$ -way,  $(t+2)$ -way, etc. combinations as well. Tables 4 and 5 compare this coverage for the Simured test inputs.

We also analyzed the average percentage of  $t$ -way combinations covered by 100 randomly generated test sets of the same size as a  $t$ -way covering array generated by IPOG, for various combinations of  $k = \text{number of variables}$  and  $v = \text{number of values per variable}$ . Table 6 shows the combination coverage of an equivalent number of randomly generated tests for  $t=2,3,4$ . For example, row 2 shows that a covering array with 30 tests covers all 2-way combinations for 10 variables with 4 values each, but 30 randomly generated tests cover only 84.6% of all 2-way combinations.

The coverage provided by a covering array versus a random test suite of the same size varies considerably with different configurations. An important practical consideration in comparing combinatorial with random testing is the effectiveness of the covering array generator. Algorithms have a wide range in the size of covering arrays they produce, but all are designed to produce the smallest array possible that covers all  $t$ -way combinations. It is not uncommon for the better algorithms to produce arrays that are more than 50% smaller than other algorithms. Comparisons show that there is no uniformly "best" covering array algorithm [10]. Algorithms vary greatly in the size of combinatorial test suites they produce, so the comparable random test suites will also vary in the number of tests. Random testing may produce results similar to combinatorial tests produced by an algorithm that generates a larger, sub-optimal covering array, because the correspondingly larger random test set has a greater probability of covering the  $t$ -way combinations.

A covering array algorithm that produces a compact array, i.e., a minimal number of tests, for  $t$ -way combinations may also include fewer  $(t+1)$ -way combinations because there are fewer tests. Note that at  $t=2$  (pairwise), an equal sized random test set covers more 4-way and 5-way combinations, which may explain why the random tests detected more deadlocks than the  $t=2$

covering array. Almost paradoxically, a sub-optimal algorithm that produces a larger covering array may be more effective because the larger array is statistically more likely to include  $t+1$ ,  $t+2$ , and higher degree interaction tests as a byproduct of the test generation. This result demonstrates that the smallest possible array is not necessarily best for testing purposes if higher strength interactions are not also tested. It also suggests that covering array generation algorithms that fill "don't care" values (those for which all combinations have already been covered) with random values may provide better test results by covering a larger number of  $t+1$ ,  $t+2$ , and higher degree combinations.

**Table 4:** Combination coverage of IPOG t-way tests

t	2-way	3-way	4-way	5-way	Avg
2	1.00	.758	.429	.217	0.601
3	1.00	1.00	.924	.709	0.908
4	1.00	1.00	1.00	.974	0.994

**Table 5:** Combination coverage, random tests

size = t-way	2-way	3-way	4-way	5-way	Avg
2	.940	.735	.499	.306	0.620
3	1.00	.942	.917	.767	0.906
4	1.00	1.00	.965	.974	0.985

**Table 6:** Combination coverage of an equivalent number of random tests

Vars	Vals / Var	IPOG tests t=2	Rand 2-way covg	IPOG tests t=3	Rand 3-way covg	IPOG tests t=4	Rand 4-way covg
10	2	10	94.1	20	94.3	42	93.2
10	4	30	84.6	151	90.6	657	92.3
10	6	66	85.6	532	91.6	3843	94.8
10	8	117	83.8	1214	90.6	12010	94.7
10	10	172	82.1	2367	90.6	29231	94.6
15	2	10	93.9	24	96.2	58	97.5
15	4	33	88.1	179	94.1	940	97.5
15	6	77	88.6	663	95.4	5243	98.2
15	8	125	86.1	1551	95.2	16554	98.2
15	10	199	86.4	3000	95.0	40233	98.2
20	2	12	96.5	27	97.3	66	98.6
20	4	37	90.9	209	96.2	1126	98.8
20	6	86	91.3	757	97.0	6291	99.2
20	8	142	91.3	1785	96.9	19882	99.2
20	10	215	88.4	3463	96.9	48374	99.2
25	2	12	95.9	30	98.5	74	99.2
25	4	39	92.1	233	97.5	1320	99.4
25	6	89	91.8	839	97.9	7126	99.6
25	8	148	90.3	1971	97.9	22529	99.6
25	10	229	90.0	3823	97.8	54856	99.6

Now consider the size of a random test set required to provide 100% combination coverage. Table 7 gives the ratio of randomly generated tests to combinatorial tests for the variable/value combinations. For example, for 10 variables with 2 values each, random generation requires 1.80, 3.05, and 3.57 times as many tests as a covering array to cover all combinations at  $t=2$ , 3, and 4 respectively. For most covering array algorithms, the difficulty of finding tests with high coverage increases as tests are generated. Thus even if a randomly generated test set provides better than 99% of the coverage of an equal sized covering array, it should not be concluded that only a few more tests are needed for the random set to provide 100% coverage. Table 7 shows that the ratio of random to combinatorial test set size for 100% coverage exceeds 3 in most cases, with average ratios of 3.9, 3.8, and 3.2 at  $t = 2, 3$ , and 4 respectively. In other words, using random tests to obtain coverage of all  $t$ -way combinations required more than three times as many tests as were needed when using a covering array. Thus combinatorial testing offers a significant efficiency advantage over random testing if the goal is 100% combination coverage.

**Table 7:** Ratio of random to combinatorial tests for 100% combination coverage

Var	Vals/ var	2-way Tests		3-way Tests		4-way Tests	
		IPOG Tests	Ratio	IPOG Tests	Ratio	IPOG Tests	Ratio
10	2	10	1.80	20	3.05	42	3.57
10	4	30	4.83	151	6.05	657	3.43
10	6	66	5.80	532	3.73	3843	3.48
10	8	117	4.26	1214	4.46	12010	4.39
10	10	172	4.70	2367	4.94	29231	4.71
15	2	10	2.00	24	2.17	58	2.24
15	4	33	3.67	179	3.75	940	2.73
15	6	77	3.82	663	3.79	5243	3.26
15	8	125	4.41	1551	4.36	16554	3.66
15	10	199	4.72	3000	5.08	40233	3.97
20	2	12	1.92	27	2.59	66	2.12
20	4	37	3.78	209	2.98	1126	3.35
20	6	86	3.35	757	3.39	6291	2.99
20	8	142	4.44	1785	4.73	19882	3.00
20	10	215	4.78	3463	4.04	48374	3.25
25	2	12	2.83	30	2.33	74	2.35
25	4	39	3.08	233	3.39	1320	2.67
25	6	89	3.67	839	3.44	7126	2.75
25	8	148	5.71	1971	3.76	22529	2.72
25	10	229	4.50	3823	4.32	54856	3.50
Ratio Avg.			3.90		3.82		3.21

The analysis suggests two significant advantages for combinatorial methods in simulations where interactions between input variables are likely to be important:

Significantly fewer tests required to provide 100% combination coverage for a particular interaction strength. Depending on problem size, random generation requires approximately 2 to 6 times as many test inputs as a covering array to cover all combinations (Table 7). While random generation will cover a significant portion of the data space, sometimes 99% or more (Table 6), this may often not be adequate in practice. The network simulation described in previous sections illustrates that combinatorial methods can detect rare interactions that may be missed with an equal number of random inputs.

*Better coverage of higher strength interactions.* As shown in Table 4, a covering array for interaction strength  $t$  is likely to provide better coverage of  $t+1$ ,  $t+2$ , etc. combinations than an equal number of random tests. This characteristic provides a greater chance of detecting events triggered by rare combinations.

## 5. Conclusions

For the simulation program tested in this study, pairwise tests detected slightly fewer deadlocks than an equal number of random tests, but 4-way combinatorial testing produced better results than an equal number of random tests. Analyzing the random test sets suggests a number of reasons for these results. Although pairwise tests covered all 2-way combinations and an equal number of random tests covered fewer, the random tests covered more 4-way and 5-way combinations, and thus had a greater probability of triggering deadlocks that depended on 4-way or 5-way interactions. However, the 4-way combinatorial tests covered significantly more 4-way combinations (100% vs. 96%) and also provided equal 5-way coverage compared with the corresponding random test set, and found more deadlocks as well.

This result demonstrated that the smallest possible array is not necessarily best for testing purposes if higher strength interactions are not also tested. When using  $t$ -way combinatorial testing, it can be helpful to evaluate the test set for coverage of  $t+1$  and higher interaction strengths. Methods of combining combinatorial and random tests may also be effective, as proposed in [2], [1]. These results also suggest that covering array algorithms may provide better test results by filling "don't care" values

with random (rather than constant, sequential, or other non-random) values.

Note: Reference to commercial products or trademarks does not imply endorsement by NIST, nor that such products are necessarily best suited to any purpose.

## References

1. Bell, K.Z. and M.A. Vouk. On effectiveness of pairwise methodology for testing network-centric software. Proc. ITI Third IEEE Int Conf on Information & Communications Technology, pages 221–235, Cairo, Egypt, Dec. 2005.
2. Bell, K.Z., Optimizing Effectiveness and Efficiency of Software Testing: a Hybrid Approach, PhD Dissertation, North Carolina State University, 2006.
3. Burr K., and W. Young, Combinatorial Test Techniques: Table-Based Automation, Test Generation, and Test Coverage, Int Conf on Software Testing, Analysis, and Review (STAR), San Diego, CA, October, 1998
4. Burroughs, K., A. Jain, and R. L. Erickson. Improved quality of protocol testing through techniques of experimental design. In Proceedings of the IEEE Intl Conf on Comm. (Supercomm/ICC'94), May 1-5, New Orleans, Louisiana, USA. IEEE, May 1994, pp. 745-752
5. Cohen, D. M., S. R. Dalal, J. Parelius, G. C. Patton The Combinatorial Design Approach to Automatic Test Generation IEEE Software, 13, n. 5, pp. 83-87, Sept 1996
6. Dunietz, S., W. K. Ehrlich, B. D. Szablak, C. L. Mallows, A. Iannino. Applying design of experiments to software testing Proceedings of the Intl. Conf. on Software Engineering, (ICSE '97), 1997, pp. 205-215, New York
7. Kuhn, D. R., D. Wallace, and A. Gallo, "Software Fault Interactions and Implications for Software Testing," IEEE Transactions Software Engineering, 30(6):418-421, 2004.
8. Kuhn, D. R. and V. Okun, "Pseudo-exhaustive Testing for Software," Proceedings of 30th NASA/IEEE Software Engineering Workshop, pp. 153-158, 2006.
9. Kuhn, D.R., M.J. Reilly, An Investigation of the Applicability of Design of Experiments to Software Testing, 27th NASA/IEEE Software Eng. Workshop, NASA Goddard Space Flight Center, 4-6 Dec. ,2002 .
10. Lei, Y., R. Kacker, D.R. Kuhn, V. Okun, J. Lawrence, "IPOG/IPOG-D: Efficient Test Generation for Multi-Way Combinatorial Testing", Software Testing, Verification, and Reliability. (Nov 29 2007, DOI: 10.1002/stvr.381)
11. Lei, Y., R. Kacker, D.R. Kuhn, V. Okun, J. Lawrence, "IPOG - a General Strategy for t-way Testing", IEEE Engineering of Computer Based Systems Conf, 2007.
12. Kobayashi, N., T. Tsuchiya, T. Kikuno, "Applicability of Non-Specification Based Approaches to Logic Testing for Software", Proc. 2001 International Conference on Dependable Systems and Networks, IEEE, pp. 337 – 346.
13. Pardo, F., JSimured - Simulador de Redes de Multicomputadores Paralelo, University of Valencia, May, 2005. <http://simured.uv.es/doc/memoria.pdf>
14. Pretschner, A., Tejjeddine Mouelhi, Yves Le Traon. Model Based Tests for Access Control Policies, 2008 International Conference on Software Testing, Verification, and Validation pp. 338-347
15. Wallace, D. R. and D. R. Kuhn, "Failure Modes in Medical Device Software: An Analysis of 15 Years of Recall Data," International Journal of Reliability, Quality and Safety Engineering, 8(4):351-371, 2001.
16. Williams, A.W., R.L. Probert. A practical strategy for testing pair-wise coverage of network interfaces The Seventh International Symposium on Software Reliability Engineering (ISSRE '96) p. 246

# Assessing Requirements Volatility and Risk using Bayesian Networks

Michael S. Russell

Booz | Allen | Hamilton

[russell\\_michael@bah.com](mailto:russell_michael@bah.com)

## Abstract

There are many factors that affect the level of requirements volatility a system experiences over its lifecycle and the risk that volatility imparts. Improper requirements generation, undocumented user expectations, conflicting design decisions, and anticipated / unanticipated world states are representative of these volatility factors. Combined, these volatility factors can increase programmatic risk and adversely affect successful system development. This paper proposes that a Bayesian Network can be used to support reasonable judgments concerning the most likely sources and types of requirements volatility a developing system will experience prior to starting development; and by doing so it is possible to predict the level of requirements volatility the system will experience over its lifecycle. This assessment offers valuable insight to the system's developers, particularly by providing a starting point for risk mitigation planning and execution.

## Introduction

When a new system is being considered for development, the system's users, developers, and other stakeholders establish a set of requirements to be implemented by the system. These requirements run the gamut from high-level concepts to design-level implementation. Over time, these requirements change as the system concept matures, user needs change, technology advances, or in response to a host of other

factors (Armour, 2000). Requirements volatility is one term that describes this change. Volatility, the inverse of stability, is not necessarily a bad thing. Some program managers would like to have a volatile schedule – as long as it is always being extended. However, changing requirements is generally viewed as detrimental to the program.

Requirements volatility makes its presence known in development projects of all sizes and types. Jones (Jones, 1994) noted that more than 70% of large software application development programs experience volatility; and this volatility, combined with poor requirements development processes and inadequate risk management, contributes to poor system quality, schedule slips and cost overruns. Jones also found that of the 60 projects surveyed, over 35% experienced scope or purpose related requirements volatility. A recent analysis of 44 different system development efforts (Stark, 2002) found that volatility affected about 63% of the system's initial requirements, and 6% of these directly impacted the system's scope. Additionally, it was found that requirement additions were much more likely than requirement deletions during the development phase of the lifecycle. Additionally, Stark found that customer changes accounted for 36% of overall volatility, system requirement developers accounted for 28%, and 36% were attributable to the system's developers.

Obviously, quantifying the effect of requirements volatility on a system would

benefit program managers and other stakeholders most notably supporting the identification system development risks. If the stakeholders were able to look at each factor that causes volatility and derive a quantifiable indicator of its impact prior to program initiation, they would have the information needed to plan for the mitigation or avoidance of each volatility factor. Additionally, process standards such as Carnegie Mellon's CMMI<sup>SM</sup> (CMMI, 2001) require programs to track and assess the impact of requirements volatility as part of their program metrics. This paper proposes a method for identifying volatility factors, assessing them, and providing useful information to the decision maker concerning the likely impact volatility will have on the program.

Methods to quantify volatility have been previously proposed. Quality Goal Modeling (Myers, 1988) is a rules-based approach designed for software developers using software quality metrics to validate requirements and to identify potentially volatile requirements. Quality Goal Modeling judges the relative volatility risk of each system requirement in terms of imprecision, conflict, and multiplicity. Additionally, it presupposes the program manager is knowledgeable enough of new system to list its essential characteristics and rank those characteristics in importance. This ranking is essential to determining the impact each requirement may have should it change. The method proposed in the paper used system characteristics as a volatility risk indicator, employing Bayesian analysis rather than rules based analysis. Using Bayesian analysis results in a more scalable analysis approach that can benefit from volatility analysis performed in previous programs.

York (York, 2001) proposed the Volatility of Requirements Assessment Method (VRAM) to uncover potentially volatile requirements early in a system's lifecycle. VRAM uses the Analytic Hierarchy Process (AHP) to compare system requirements against historical causes of volatility, such as "User Needs Change." The results of this process are used as a decision aide to help program managers determine if additional requirements analysis should be conducted prior to beginning system development. York considered his research inconclusive and found that even experts were not able to accurately predict volatility. He emphasized the importance of future research, especially concerning enhanced support aids to engineers in assessing volatility. This paper proposes using a Bayesian Network to provide a decision support aid.

The first challenge to developing this method is determining the situations indicative of future requirements volatility, followed by determining the impact these situations have on the system's development. As it is impractical to quantify the entire set of situations that might impact requirements volatility, a representative set must be chosen. Then the relationship between this representative set and the anticipated level of requirements volatility must be established. Often, this relationship is expressed in terms of experience-based belief, rather than with hard data, which tends to complicate any attempt to quantify the effects of volatility.

The Bayesian approach to statistical modeling relies on prior evidence to provide a rational basis for design making (Lee, 1989), and the idea of using Bayesian Analysis as a decision tool was noted in (von Winterfeldt, 1996). Bayesian statisticians are well versed in using limited or incomplete data, unquantifiable beliefs,

and other “soft” evidence to derive useful information for decision makers. By understanding the impact specific volatility factors have had on past system development efforts the program manager can use Bayesian analysis to turn this previous information into a reasonable and defensible prediction as to the level of requirements volatility the new system may experience. Other more commonly used statistical methods rely on vast quantities of hard data to make an inference as to what the data might mean. Using Bayesian analysis allows the program manager to make a reasonable judgment about volatility early in the program when the large amounts of data needed to support other statistical methods is not available.

### **Requirement Volatility Factors**

There are as many factors for requirements volatility as there are people who write requirements, with each person having his/her own understanding why requirements change and the effect of specific volatility factors. As it is not feasible to evaluate every potential source of requirements volatility and quantify its effect on a developing system, a representative set of volatility factors that most directly effect system development should be evaluated. Ideally, this representative set would be general enough to be domain insensitive and applicable to a wide assortment of development systems. By adhering to this ideal; the method, and any tools developed using it, could be applied to many different programs. Additionally, the lessons learned in each program can be retained and used to provide better estimates of requirements volatility in future programs.

The following set of volatility factors was derived from literature (Sommerville, 1992;

Brooks, 1987; Christel 1992), interviews with program managers, systems engineers, and examinations of previous system development efforts, as appears below.

- **Schedule Stability:** Measures the anticipated stability of the project’s schedule. A shorted schedule can affect requirements development through the elimination of requirements engineering time, resulting in missed or poorly specified requirements. A shorter the development schedule may mean that some requirements initially specified will have to be dropped, which affects the overall system design. While shortening a schedule could be considered to be detrimental, lengthening a schedule may not necessary be good. Sometimes a longer schedule gives the systems stakeholders more opportunities to change requirements. The effects of any schedule change should be carefully considered.
- **Budget Stability:** Measures the anticipated stability of the project’s budget. Increases in project budget often come with additional, unplanned requirements. These late arriving requirements pose integration challenges. Decreases in budget may cause non-core functional requirements to be dropped, which affects overall requirements stability and integration.
- **Scope Stability:** Measures the anticipated stability of the project’s scope. Changes in the project’s scope may have a serious impact on the requirements defined for the system. In the worst case, the purpose for the system may be completely changed, leading to a

whole new set of system requirements.

- **System Need:** Measures the level at which the user's need for the new system has been established. Without clearly defining the user's need for the system, the requirements that are critical to making the system useful for its intended audience may not be documented and subsequently built into the system.
- **Changing Priorities:** Measures whether or not, and how often, the system's customer's priorities change. Changing priorities are related to changing needs; however, where system need deals with how that system will solve a problem the customer has priorities measure how critical that need is. A system that starts out as a high priority will be provided with plentiful resources and development time.
- **Changing Expectations:** Measures how often or to what extent the customer's expectations for the system change. Expectations are hard to quantify, as they are rarely documented and may not appear as defined requirements. Expectations not only drive how a project is perceived, but also its future success and how individuals react to it. Many times the system's customer may anticipate the system will meet a specific need while the testable requirements that would enable that expectation to be met are never documented.
- **Operational Concept Stability:** Measures the stability of the systems operational concept. The operational concept defines the system's place in the world and how it fits into the overall enterprise. It also describes how the ultimate user, who may not

be the system's customer, intends to use the system to accomplish a mission resulting in a shared vision for the system (Wheatcraft, 2003).

- **System Interface Plan:** Measures whether or not a system interface document is scheduled for development. The system interface document lays out physical and functional designs for how the developing system will interface with other systems.
- **System Design Plan:** Measures whether or not a system design document is scheduled for development. Often this document represents the first time all system requirements are identified and documented (Wilson 1997). This level of detail is normally not included in system scope and operational need documents.
- **System Test Plan:** Measures whether or not a system test plan will be produced and, if so, how formal the system test process will be. Without a clearly defined system testing approach, it is impossible to know whether or not the system requirements have been met or whether the documented requirements are the right ones.
- **Technical Change:** Many systems, especially those in the information technology domain, are required to incorporate the leading edge of technology. Unfortunately, technology constantly changes, and this change is rapid (Armour, 2000). Many times a newly developed system is obsolete when fielded due to rapid technology change. Even when technology changes don't have a direct impact on system is development it may have an impact

on the customer's expectations for the system.

- **Requirement Traceability:** Measures the extent to which requirements are traceable to user needs, expectations, and the system's scope. Requirements not directly traceable to one of these or to another requirement are prime candidates to be modified or dropped.
- **Requirement Conflicts:** Measures the expected number of requirement conflicts. Conflicts can occur in many different, sometimes unanticipated places within the system's design and can be difficult to predict during system's planning stages. A conflict, such as a messaging protocol not matching the communications network that it must be transmitted on, must be adjudicated with the customer prior to system design finalization. There are many requirements engineering software tools on the market, such as DOORs®, that can be used to support the projects requirement development effort. By implementing these tools, the number of conflicting requirements is generally reduced; and traceability between requirements and from requirements to systems concepts and objectives is increased.
- **Implied Requirements:** Measures the expected number of implied requirements. An implied requirement is not specifically stated by the customer, but must be implemented in order to realize the customer's original requirement. One root cause of implied requirements is unstructured, natural language in requirements development which leads to ambiguity, inaccuracy, and assumed

requirements (Stokes, 1991). Additionally, what seems like a simple requirement or requirement change to the customer brings with it costly implied requirements. For instance, adding an additional antenna to an aircraft seems like a minor change; however, any new antenna would mandate a new hole in the aircraft's pressure hull requiring extensive FAA mandated pressure testing and hull recertification. These tests are time intensive and more expensive than the antenna itself.

- **Interoperability Requirements:** Measures the anticipated amount of interoperability requirements. These requirements may deal with external systems or system subcomponents. These requirements are implemented and given structure by the system interface plan. Without this plan, these requirements lack context, and potential overlaps or conflicts between them are hard to uncover. Changes to systems that must interoperate with the new system can greatly impact requirements in unanticipated ways. When evaluating the potential impact of interoperability requirements, both their use within the system and as conduits to other systems must be considered.
- **Environment Change:** Measures the impact a change in the physical environment that the system operates in may have on requirements. This can work in two ways. First, the user's need may now require the system to work in an Arctic or other extreme environment. Second, the expected environment may change in some manner.

- **System Complexity:** Measures the relative complexity of the system. This measure can vary depending on the domain. An assembly line for coat hangers isn't complex from a technology perspective; however, developing this system requires close synchronization of many mechanical subsystems. Conversely, an embedded operating system for a cellular telephone is a relatively small piece of software that represents technical complexity and reliance on interoperability standards to function correctly. Evaluating complexity as it relates to requirements volatility requires sound judgments concerning how likely the complexity of the system is to drive requirements change. In many cases, higher systems complexity is more likely to exhibit requirements volatility.
- **Reuse Requirements:** Measures the relative level of reusable component integration desired by the customer or required based on technical standards within the domain. Reusing components of existing systems, hardware and/or software, to support new system development is an increasingly common requirement. Component reuse allows for greater built-in interoperability within a domain and may lead to decreased costs, but only if interfaces to the reusable components are accessible. System complexity will increase if it is known or anticipated that the interfaces to the reuse component will be difficult to decipher.
- **Subject Matter Expert (SME) Availability:** Measures how available SMEs will be to assist in the requirements engineering process. SMEs, either from the customer or domain, are key to successfully generating stable requirements. The developer may have a good bit of domain experience; however, the best judge of how well requirements have been identified are the customer's SMEs
- **Analyst Skill:** Measures the experience level of the analysts who are working with the customer and SMEs to facilitate and document requirements.
- **Program Management Skill:** Measures the experience level of the program management team. Skill is needed in two areas to mitigate volatile requirements. First, managing the development team and the requirements development process. Second, managing the customer. The first skill is much easier to judge than the second, the second being the most vital.
- **Developer Skill:** Measures the ability of the developers to interpret and transform system requirements into system design correctly and recognize the impact that requirement change will have on the system. The developers have the best understanding of functional dependencies and can usually provide the best impact estimate for a changing requirement.
- **Defined Processes:** Measures the existence and institution of requirements engineering processes. Process standards set out a defined and repeatable process to support requirements work. When followed, they help ensure requirements are derived, documented, and changed in a reasonable manner. When an organization fails to implement a

consistent requirements engineering process, volatility will follow.

- **Project Turnover:** Measures the amount of employee turnover expected during the project. Project employee turnover can greatly affect the way requirements are documented and interpreted even with a well-defined configuration management process.
- **Customer Turnover:** Measures the amount of customer turnover expected during the project. Customer employee turnover can be a serious issue for a development team, as the new customer representative may have a completely different idea as to what the system is to do. Customer management is essential to success.
- **Company Domain Experience:** Measures the amount of experience the development organization has in the systems domain. Companies with lots of domain experience should be able to lean on that experience to produce less volatile requirements. Simply being technically able to build the product is not enough. A company that builds financial planning software may have the technical expertise to build a military command and control system, but a lack of domain knowledge will result in more volatile requirements. Of course a lack of domain experience is not always a bad thing. A developer with limited domain experience will be forced to ask many questions to fill in gaps that the customer just assumed everyone knew and didn't bother defining. With a more experienced company, these gaps are filled by implied, experienced-based

requirements that may or may not meet the user's expectations.

## Volatility Measurement

A note on requirements; a requirement that is not verifiable is not a requirement. In the same way, in order to make a judgment about the potential impact of a requirement volatility factor, some way to measure its impact must be established. The difficulty in measuring the impact each factor may have on overall requirements volatility is that these impacts are inherently unquantifiable. Value judgments, prior beliefs, and "gut-feelings" tend to color evaluations of factors such as "will the customers expectations change" or "how many requirement conflicts will appear."

Typical systems engineering methods, from the waterfall to the spiral, all consider volatility risk (Sommerville, 1992). In particular, the spiral was designed specifically to take volatility into account throughout the lifecycle (Boehm, 2000); however these methods do not provide a systematic method to measure potential volatility. Without a way to measure and then relate volatility factors, there will be gaps in a program's volatility analysis. Here Bayesian analysis becomes a valuable resource for judging the impact of each factor while programming planning has yet to be completed. By combining this method with the spiral development lifecycle, volatility measurement will be more rigorous and cause – effect relationships between factors will be maintained.

## Building the Requirements Volatility Model

At its most fundamental level, making a judgment about how different volatility factors will affect a program during the

initial planning stages is a decision problem; one that could potentially be solved in many ways. Hopefully, the program manager will use his/her experience, or that of others, to make a reasonable and informed decision concerning the extent to which the program's requirements will be subject to change, and to characterize that change. In any case, the program manager will make assumptions about the program, its customer, and world states both within and outside his control. Based on this knowledge, the program manager can take many actions, each with its own consequences. The challenge to the program manager is to make sound judgments or inferences based on prior knowledge or the experience of others, while oftentimes not knowing all the consequences of potential actions.

In order to decide on the best actions, those that minimize or mitigate the effects of volatility on the program, the program manager must implement a method that represents his/her beliefs about each volatility factor and make an inference about their impact to the overall program. Armed with this information, he can build volatility mitigation planning into the program plan.

Bayesian analysis is a statistical method for supporting the decision making process by representing beliefs about the world as probabilities. These probabilities are not definitive, meaning reasonable people might disagree about the validity and applicability of the resulting data. However, given informed prior information, a reasonable and defensible inference about new data based on previous data can be made (Laskey, 2003). For program managers, this means factors such as budget shortfalls and documentation problems that have caused requirements volatility on past programs can reasonably be used to predict the same

problems on the current project given similar development environments.

In order to apply the Bayesian approach, a method for combining information about a project's perceived level of volatility with Bayesian reasoning must be established. This can be accomplished through the development and application of a Bayesian network. A Bayesian Network (BN), based on probability theory, is a knowledge representation that effectively captures the uncertainties and conditional independences present in a given domain. As such, it can be used to make reasonable inferences with limited data (Jensen, 1997).

BN's are drawn as directed graphs comprised of nodes and arcs. The nodes represent variables whose value is uncertain and the arcs represent dependency relationships between the variables. As a computational architecture, a BN allows the user or application to declare "evidence" on some of the nodes and, through a process called "evidence accumulation," compute revised probabilities for all other nodes in the network. (Laskey, 2002)

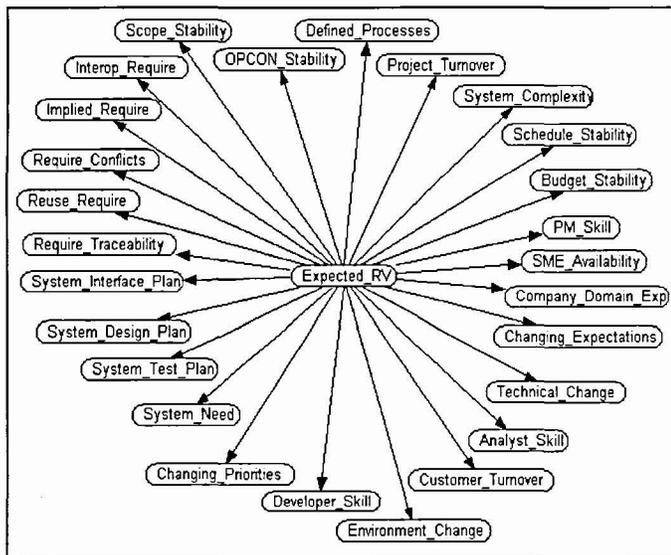
It has been postulated that BNs can be inadequate as a general knowledge representation language for large and complex domains (Mahoney, 1996). As noted previously, it is not reasonable to try to quantify every factor that could impact requirement volatility for a system. This is what makes the Bayesian approach so valuable. It is also important to choose volatility factors that are not domain specific, so the resulting BN can be used repeatedly and among many domains. Staying as domain generic as practical means that BN tools would not have to be customized for each project and a repository of volatility information can be created (Koller, 1997).

Unless a specialized decision support tool incorporating a Bayesian network model is available, developing this tool for a single application will consume significant resources during the beginning stages of a program. So the cost of developing the model must be weighed against the benefit the model provides to the program manager. The Bayesian network model developed to support this research provides a good starting point for program managers seeking to incorporate Bayesian inference into their decision making process.

Using BNs to help solve decision problems or derive useful information is not a panacea for every situation; however, it has been found to be very useful in solving a variety of real life problems such as quickly identifying friendly from enemy aircraft (Laskey, G, 2002). Using the volatility factors identified in the previous section as the nodes of the BN, a model was constructed using Norsys Corporation's Netica software tool.

that the features of the BN, in this case the volatility factors, are conditionally independent from each other given the expected volatility. Another type of BN is an "Optimal Classifier," which takes dependencies and other types of relationships between nodes, states, and other model elements into account. (Laskey, 2003)

The naïve BN was used for this model for two reasons. First, the optimal classifier method is more accurate; however, it carries with it a greater computational load and need for more complete information concerning the relationships between network nodes (Laskey, 2003). As the requirements volatility model is to be used during the early stages of program development and is intended to support rough order of magnitude predictions, it was felt that the data necessary to support a fully optimized model might not be present. Also, the time needed to enter and run the optimized model might limit its use by busy



**Figure 1: Requirements Volatility BN**

The BN was constructed as a "Naïve Bays" Network. A naïve Bays network assumes

program managers. A method such as using a BN to predict requirements volatility must be used to be useful. This was taken into consideration. Second, research concerning interrelationships between volatility factors

is immature, and there is not general agreement within the industry as to which factors influence other factors. While some relationships are easy to infer, such as the relationship between customer expectations and a changing scope, others such as the relationship between schedule slips and requirement changes are harder to quantify. As more research in this area is completed, it would be wise to revisit the type of BN used to support the volatility model at a later date.

### **Developing Priors**

In order to be a useful tool, the BN must be seeded with information describing the various volatility factors that have impacted systems in the past. The prior information for the requirements volatility model was developed based on reviews of pertinent literature and by surveying requirements engineers to uncover their beliefs about the root causes of requirements volatility.

The first survey was web-based and consisted of three parts. The part 1 contained 27 questions in 3 categories: (1) technical and program management skills, (2) requirement and design related, and (3) project environment and prior planning. The questions covered a range of requirements engineering and project environment challenges with the idea that each one represented a root cause of requirements volatility.

To verify that the questions asked were clearly written and appropriate to the issue at hand, a group of experienced engineers was polled to validate the survey's questions. Most possessed 10+ years of project management and requirements engineering experience. As an additional verification step, each survey respondent was asked to list the top three reasons why they felt requirements were subject to change in part

2. It was felt that part 2's "free text" entry style would be conducive to eliciting the respondent's true beliefs concerning requirements volatility. As will be noted later, the volatility factors listed in part 2 closely mirrored the more structured questioning found in part 1.

Part 3 elicited demographic information from each respondent, covering academic and work experience background as well as experience in the requirements generation process. The demographic information would be used to determine if a significant variation occurred in the answers given by respondents from different demographic groups.

Although the survey was anonymous, demographic information indicated variation among respondents covering academic, industry, and government perspectives, and encompassing a variety of experience levels. From these responses, a probability distribution for each volatility factor was developed. Based on these distributions, the relative impact of each volatility factor on overall volatility was derived.

Volatility factors dealing with the customer's expectations and defined need for the system were rated as the most likely causes of requirements volatility followed closely by instability in the system's scope. Budget stability problems, customer turnover, and technological advances within the systems domain rounded out the group of factors the survey respondents listed as the most likely causes. Based on survey responses, one of the original volatility factors dealing with the effects of an unstable Work Breakdown Structure (WBS) was removed from the model. Additionally, several factor names were changed to reflect the often-repeated responses in part 2 of the survey.

The next step in developing prior information for the BN was to conduct a second survey in which respondents were asked to assess the impact of each volatility factor on a series of five fictional scenarios. Each scenario depicted a system development project with good and bad aspects. By assessing each factor in relation to the scenario and the overall level of requirements volatility the respondent felt the system would exhibit, a matrix of responses was developed. This matrix was then used to “learn” the probability distribution of each node from the input data. Information on probabilistic learning can be found in (Laskey, 2003) and (Robert, 2001).

### **Using the Model to Support Decision Making**

To implement the model to predict the overall level of requirements volatility, the program manager would sit down with his management team during the early stages of the program’s development and record their collective beliefs as to the extent each volatility factor is present in the current program. These beliefs are entered into the model. Based on the beliefs entered for each factor (nodes within the network), the model will produce an overall measure of the requirements volatility that should be expected during the system’s development.

Each volatility factor is ranked from 1 to 4. A 1 represents a factor with minimal expected impact on the system. A 4 represents a factor with a major impact, with a score of 2 or 3 being somewhere between these two extremes. As this rating scheme is inherently qualitative, the program manager must establish some ground rules for determining how these ratings should be applied to maintaining consistent results. Also, while the program manager could

assign one person to make volatility judgments and complete the model, the resulting information would not be as useful as having several people with different perspectives on the program do so. By having several people work on the model, biases tend to cancel out, and a true measure of potential volatility emerges.

After each volatility factor is ranked, the data is entered into the BN tool. For the examples in this paper, the Netica tool was used. Assuming prior information was entered into the BN tool correctly; the tool will return a probability distribution that can be used to predict the level of requirements volatility the program may experience over its lifecycle. For the model in this paper, the Netica tool returns a probability distribution ranked between 1 and 10. A 1 indicates a program with a very low level of potential volatility, while a 10 represents a program with an extreme amount of potential volatility.

The prediction provided by the model is just that – a prediction. It should not be used as the sole basis for justifying risk mitigation strategies, especially expensive ones, to counteract the effects of volatile requirements. Rather, the results of the model combined with the experience of the program team work together to draw a reasonable inference and serve as a tool for mitigating potential requirements development risks. One way a program manager could use the model would be to identify the top 3-4 volatility risks and concentrate risk reduction efforts on those items. This procedure would work extremely well with a spiral development method. During spiral development, requirements will constantly shift, especially early in the lifecycle. Using the BN to predict the most likely sources of volatility and linkages between volatility factors at the

beginning of each spiral will give the program manager the information needed to begin risk reduction activities.

### **Conclusion:**

This paper outlines one method for predicting the level of requirements volatility a system may experience during the development phase of the SE lifecycle. This prediction, with its statistical bases, provides system stakeholders with greater visibility concerning the root causes of volatility in a given program and some clue as to what portions of the system's development lifecycle are most likely to suffer from volatile requirements.

The outlined method is designed to be generic enough to be applied to many different development domains, and data captured about the impact of specific volatility factors can be reused by future programs to provide progressively better predictions of overall requirements volatility. This method also provides a high degree of flexibility to its user. By identifying additional volatility factors or inferring relationships between factors, a program manager can easily customize the BN to reflect the unique issues and other circumstances for his system.

The information used to develop the priors for the model was good enough to show the concept of using a BN for volatility prediction is sound. The next step for the method is to apply the model in a systems development environment and to judge how well the model predictions are useful to the systems stakeholders and reflect the actual level of requirements volatility the system experienced. As more information is added to the model, the model will become a better indicator of potential volatility.

### **References:**

- Armour, F., Catherwood, B., Beyers C., "A Framework for Managing Requirements Volatility using Function Points as Currency." *International Function Point Users Group Annual Conference*, San Diego, CA, September 2000.
- Boehm, B., "Spiral Development, Experience, Principles, and Refinement." *Briefing to the Ballistic Missile Defense Organization*, The Pentagon, February 2000.
- Brooks, F., "No Silver Bullet: Essence and Accidents of Software Engineering." *IEEE Computer*, April 1987, pp 10-19.
- CMMI for Systems Engineering, Software Engineering, and Integrated Product and Process Development v1.1*. Software Engineering Institute, Carnegie Mellon University, Pittsburgh, PA. 2001.
- Christel, M., Kang, K., "Issues in Requirements Elicitation." *Technical Report CMU/SEI-92-TR-12*, SEI Carnegie Mellon University, Pittsburgh, PA. 1992.
- Jenson, F., *An introduction to Bayesian Networks*. Springer-Verlag, NY, 1997.
- Jones, C., *Assessment and Control of Software Risks*. Prentice-Hall International, Englewood Cliffs, NJ, 1994.
- Koller, D., Pfeffer, A., "Object-Oriented Bayesian Networks." *Proceedings of the 13<sup>th</sup> Annual Conference on Uncertainty in Artificial Intelligence (UAI-97)* Providence RI, August 1997, pp 302-313.
- Laskey, G., "Combat Identification with Bayesian Networks." *Student Paper*, SEOR Department, George Mason University, 2002.
- Laskey, K., *Bayesian Inference and Decision Theory, SYST 644 Class Notes*. Dept. of Systems Engineering, George Mason University, <http://ite.gmu.edu/~klaskey/SYST664/SYST664.html> Spring 2003.

- Laskey, K., Barry, P., Brouse P., "Development of Bayesian Networks from UML Artifacts." Dept. of Systems Engineering, George Mason University, 2002
- Lee, P., *Bayesian Statistics, An Introduction*. 2<sup>nd</sup> Ed.; Oxford University Press, New York, 1989.
- Mahoney S., Laskey, K., "Network engineering for complex belief networks." *Proceedings of the 12<sup>th</sup> Annual conference on Uncertainty in Artificial Intelligence. UAI-96*, 1996, pp 389-396.
- Myers, M, "A Knowledge Based System for managing Software Requirements Volatility." *Doctorial Thesis, George Mason University, Fairfax, VA.*, 1988
- Robert, C., *The Bayesian Choice, 2<sup>nd</sup> Ed.* Springer Texts in Statistics, New York, 2001.
- Solomon, P., "Managing Software Projects with Earned Value." *Northrop Grumman Corporation Internal Paper, [solompa@mail.northgrum.com](mailto:solompa@mail.northgrum.com)*, 2000.
- Sommerville, I., *Software Engineering, Fourth Edition*. Addison-Wesley Publishing, Wokingham, England, 1992
- George Mason University, Fairfax, VA. 2001.
- Stark, G., et al., "An examination of the Effects of Requirements Changes on Software Maintenance Releases." *To be published in the Journal of Software Maintenance*, available at <http://members.aol.com/geshome/iblieve/jsmregres.pdf> 2002.
- Stokes, D., "Requirements Analysis." *Computer Weekly Software Engineer's Reference Book*, 1991, pp 3-21.
- von Winterfeldt, D., Edwards, W., *Decision Analysis and Behavioral Research*, Cambridge University Press, Cambridge MA, 1986
- Wheatcraft, L., "Delivering Quality Products that Meet Customer Expectations." *Cross Talk, VOL 16, No. 1*, January 2003, pp 11-14.
- Wilson, W., "Writing Effective Requirements Specifications." *Proceedings, Software Technology Conference*, April 1997.
- York, D. "An early indicator to Predict Requirements Volatility" *Doctorial Thesis, Ge*



# Virtual Diagnostic Interface: Aerospace Experimentation In The Synthetic Environment

**Richard J. Schwartz, Senior Research Engineer; Andrew C. McCrea, Research Engineer**  
ATK Space, supporting the Advanced Sensing and Optical Measurement Branch  
NASA Langley Research Center, Hampton, VA 23681 USA  
[Richard.J.Schwartz@NASA.gov](mailto:Richard.J.Schwartz@NASA.gov), [Andrew.C.McCrea@NASA.gov](mailto:Andrew.C.McCrea@NASA.gov)

**Abstract.** The Virtual Diagnostics Interface (ViDI) methodology combines two-dimensional image processing and three-dimensional computer modeling to provide comprehensive *in-situ* visualizations commonly utilized for in-depth planning of wind tunnel and flight testing, real time data visualization of experimental data, and unique merging of experimental and computational data sets in both real-time and post-test analysis. The preparation of such visualizations encompasses the realm of interactive three-dimensional environments, traditional and state of the art image processing techniques, database management and development of toolsets with user friendly graphical user interfaces. ViDI has been under development at the NASA Langley Research Center for over 15 years, and has a long track record of providing unique and insightful solutions to a wide variety of experimental testing techniques and validation of computational simulations. This report will address the various aspects of ViDI and how it has been applied to test programs as varied as NASCAR race car testing in NASA wind tunnels to real-time operations concerning Space Shuttle aerodynamic flight testing. In addition, future trends and applications will be outlined in the paper.

## INTRODUCTION

The advent of affordable very high powered desktop computer processing has provided a level of access to advanced three-dimensional computer graphics that have never before been available outside of a limited, computer science orientated environment. Today's personal computer (PC) based systems with workstation class graphics cards are capable of displaying and manipulating highly complex three-dimensional geometries with detailed texture maps under stunning simulated lighting conditions. The origins of these capabilities began to emerge in the early 1990's, and were immediately put to use in support of advanced wind tunnel and flight test instrumentation systems being developed at the NASA Langley Research Center (LaRC). This paper will review the applications of three-dimensional modeling and simulation work developed in the Advanced Sensing and Optical Measurement Branch (ASOMB) over the last seventeen years, concentrating on the current suite of applications and our plans for future development.

### What is ViDI?

The Virtual Diagnostics Interface, or ViDI, is a methodology of applying two-dimensional image processing, three-dimensional computer graphics, physics-based modeling, and the handling of large data sets to use in solving complex aerospace testing and data visualization problems. To date, most of the two-dimensional image processing has been developed within NASA, while the three-dimensional visualization capabilities have been derived from Commercial-Off-the-Shelf (COTS) software packages. However, these COTS programs were chosen for

their ability to be programmed and work seamlessly with custom user interfaces and libraries of physics based simulation software. There are three main areas in which ViDI is utilized: (a) pre-test planning, which involved the simulation of an experiment and the planned instrumentation system in a three-dimensional virtual world as shown in Figure 1, (b) real-time data visualization in an interactive virtual environment, and (c) post-test data unification, where disparate forms of data are brought together *in-situ* in the virtual environment to help obtain a more global perspective on the causes and relationships of experimental parameters and the resulting physical phenomena reported by the data [1].

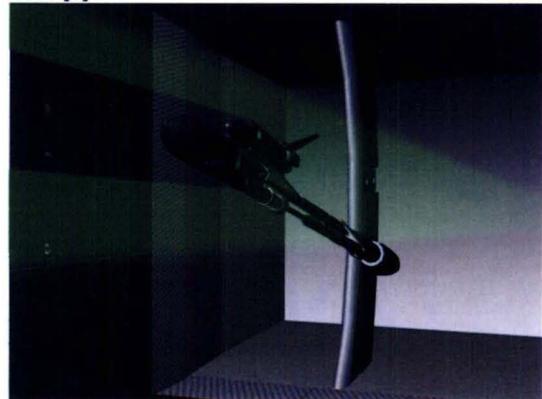


Figure 1: ViDI Visualization of laser light sheet for aerodynamic flow investigation on Space Shuttle wind tunnel model

### History of ViDI

In 1990 NASA embarked upon the development of an instrumentation system to be placed aboard a flight test aircraft which would use lasers and

cameras to obtain quantitative images of the velocity of the airflow of the aircraft. This instrumentation would record megabytes of data in a few seconds of operation, which at the time was problematic on several fronts. One of the first issues was how to display the time variant data images in a meaningful way. Fortunately, the first three-dimensional computer visualization systems were coming on the market, and a DOS based commercial program, running on an Intel 486 class computer was utilized to experiment with mapping the data images into the virtual environment, Figure 2.

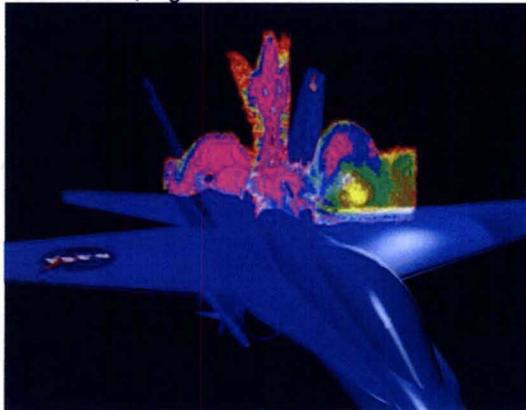


Figure 2: First data mapping of wind tunnel data representing airflow velocity over F/A-18 aircraft in preparation for flight tests of this configuration.

Almost immediately it became apparent that the camera simulation capabilities of the virtual environment would be ideal for use simulating the experiment as a whole for planning purposes. Ultimately, the instrumentation flight test series was cancelled by NASA due to funding issues, and the continued instrument development was focused on wind tunnel applications. The visualization work, then called Virtual Facilities, continued to grow, supporting a wide variety of ground based aerospace testing techniques, as well as flight test applications. Reincarnated as the Virtual Diagnostics Interface (ViDI) in the late 1990's, the scope of applications has expanded to include real-time data visualization and comparison in virtual environments, new ways of merging experimental and computational data, and support for hypersonic aerodynamic flight testing on the Space Shuttle.

## CURRENT APPLICATIONS

### Pre-test Planning

The cost of experimental aerospace testing is a key driver determining the design cycle of a new configuration. Modern computational techniques

have relieved some of the requirements for experimental testing, but there is still a strong need to validate computational results and run tests on conditions where computational methods are not yet fully developed. ViDI has played a key role in optimizing the design of wind tunnel tests to minimize the test set-up time and ensure the desired data can be acquired.

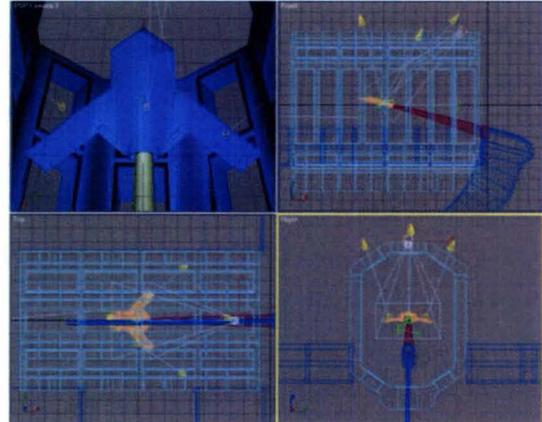


Figure 3. Screen capture of a typical wind tunnel test setup, shown here for PSP. Yellow cones represent lights; the upper left image is the simulated camera image.

ViDI utilizes the virtual world as a stage upon which an experiment can be designed. At the core of the visualization is a carefully scaled model of the test facility, usually a wind tunnel in this case (Figure 3). The model is crafted to represent both the inside and the exterior of the facility. Sufficient detail is required to provide the researcher with enough information to determine where to place items such as sensors or cameras and lights. Additionally, the researcher has to have an accurate three-dimensional computer model of the article being tested, such as an aircraft or rocket. Most often, these geometry files are provided by the company that creates the actual wind tunnel model, and are received in common Computer Aided Design (CAD) file formats. If necessary, these files can be translated into formats readable by the visualization software using commercial software products. The test configuration model files are merged with and scaled to the geometry of the experimental facility. Lastly, the mounting device, known as a sting, is added. This unites the test configuration to the facility model, and often requires dynamic modeling to provide the proper motion of the test configuration as it is pitched, rolled, yawed and translated through the test section.

Given a virtual scene that realistically mimics a planned test; the research may concentrate on how the test will be instrumented. Many advanced wind tunnel techniques utilize cameras

as their sensors. Examples include Pressure and Temperature Sensitive Paints (PSP and TSP), which require specific camera views and lighting conditions on the test configuration. Additionally, Particle Image Velocimetry (PIV) measures air flow velocities using high resolution images of smoke particles in a flow, and Projection Moiré Interferometry (PMI) uses cameras and special lighting techniques to measure surface deflections. The common element is the use of cameras and lighting as integral components of the instrumentation. Using the virtual environment the researcher may experiment with camera placement, required field of view and depth of field. Multiple lighting conditions can be analyzed to eliminate unwanted shadowing, and most importantly the optical access to key regions of the flowfield can be assessed under all possible model positions and orientations. This *in-situ* investigation of experimental setups has proven to save significant time in both the test setup and the running of the experiment by eliminating surprises and providing a clear line of communications to the test team.

The examples below depict test setups for applications as varied as PSP testing in the Langley 30 x 60 Foot Full Scale tunnel on an actual NASCAR racing car to a Laser Velocimetry experiment on the Space Shuttle in the USAF Arnold Engineering Design Center (AEDC) to a NASA Ares rocket in the NASA LaRC Unitary Plan Wind Tunnel (Figure 4).

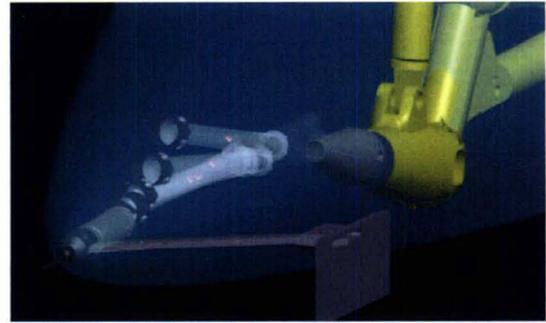
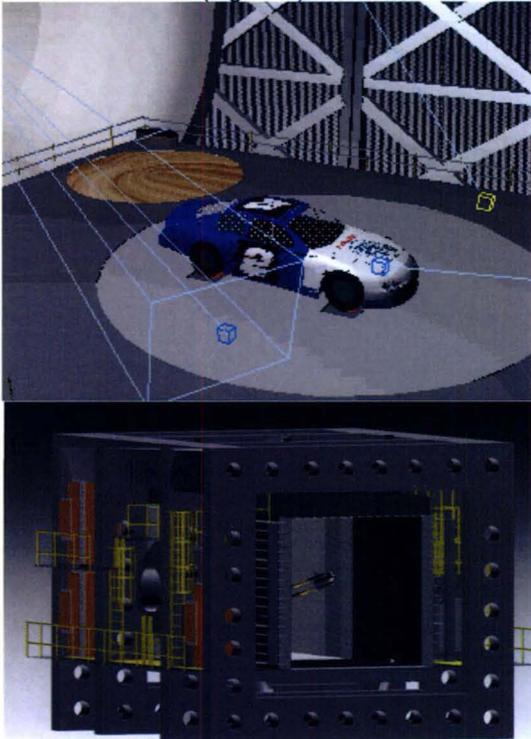


Figure 4. Test setup visualizations used for NASCAR testing, Space Shuttle flow visualization and Ares rocket stage separation testing.

### Real Time Data Visualization

The virtual environment developed in the pre-test phase described above can also be used as the foundation for displaying data in real time in an interactive three-dimensional visualization. To date, three forms of data have been incorporated into the visualization; two dimensional imagery, vector forces, and scalar point information, such as pressure and temperature [2].

At the heart of the real-time data visualization is a custom program developed to feed information into the virtual environment. The first version of this software was designed to interface with cameras to provide real-time streaming video that was embedded into the virtual environment. This was especially useful for techniques such as laser light sheet flow visualization (Figure 5) or Schlieren, which provided a view of the flow that could be rapidly mapped to a plane in the virtual environment.

Following the successful deployment of the real time imagery in the wind tunnel, the system was expanded to interact with the wind tunnel Data Acquisition System (DAS). The DAS is a computer that can process hundreds of scalar parameters defining the wind tunnel environment and test conditions at a given moment in time. This provided a source of information for the three-dimensional virtual environment for pressure, temperature, tunnel velocity, forces and more. Using this information, a comprehensive visualization depicting the actual state of the experimental test article in real time was developed. Pressures and temperatures were shown as bars located at the points where the

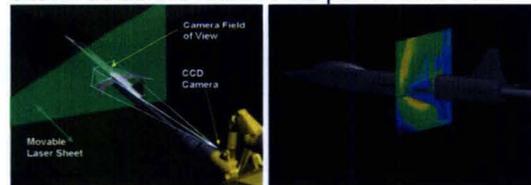


Figure 5. Flow Visualization experiment first used for real-time data visualization in the virtual environment.

sensors were located, arrows of changing magnitude represented forces and moments (Figure 6), and state information such as velocity, angle of attack, or flow temperature was affixed to the display.

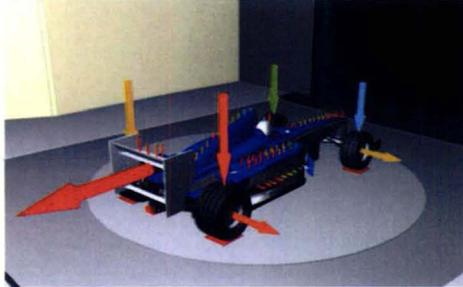


Figure 6. Sample real-time visualization – small bars on model represent measured pressures, arrows show measured forces.

An addition and very important portion of the real-time data visualization capability was the inclusion of pre-computed results from predictive methods such as Computational Fluid Dynamics (CFD). A database of computational results was stored on the ViDI computer in a manner that allows the information to be retrieved based on key test parameters, such as model attitude (roll, pitch and yaw), flow velocity or Mach number, and other pertinent parameters. Then, as the ViDI computer received data from the DAS it automatically retrieved the correct computational solution and displayed it in real time along with the experimental data. It also did a real-time differencing, which rapidly showed the level of agreement between experimental values and prediction (Figure 7). This system was run automatically for hours on end during tests, which allowed the user to concentrate on the data visualization and not the care and feeding of the software.

The real-time software has been developed with wind tunnel testing as the primary application. However, the technology is clearly not limited to just wind tunnel experiments and validation of CFD predictions. This capability can easily be expanded to include computational predictions for any form of analysis – structural, thermal, or other, and the experimental data source may originate from any form of experimental apparatus. The ViDI capability is designed to rapidly allow the user to investigate the fidelity of both the computational and experimental results, and provide a validation capability in real-time that will allow the user to identify issues during a test, while there is still time to affect the way the experiment is being conducted.

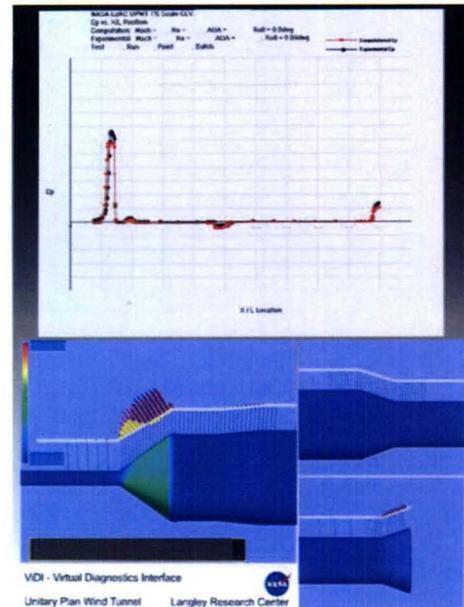


Figure 7. Display from real time test – black bars are experimental pressures, red bars are CFD pressures, surface coloration is CFD pressure distribution, yellow bars show difference from experiment and CFD. X-Y plot is also created in real time to augment visualizations.

### Post-test Data Unification

Over the many decades of wind tunnel testing, data visualization has usually been confined to two-dimensional data plots. With the emergence of CFD data visualization, visualizing data sets of flow features and the physical conditions on an aerodynamic surface (pressure, temperature, shear stress, etc) became integrated with the three-dimensional representations of the test geometries. ViDI has expanded upon this to include disparate forms of experimental data unified into a single visualization, often combined with computational predictions as well. This provides two very important capabilities; the ability to compare very large quantities of experimental and computational results quickly and intuitively, and the ability to obtain a global awareness of the cause and effect relationships between the physical features and trends occurring in the datasets. This level of integration has led to better understanding of the fundamental physics, as well as the often overlooked limitations of either experimental or computational methods. Ultimately, it allows the researcher to have a far superior situational awareness of the experiment that has been conducted than is possible only with a series of traditional X-Y plots (Figure 8).

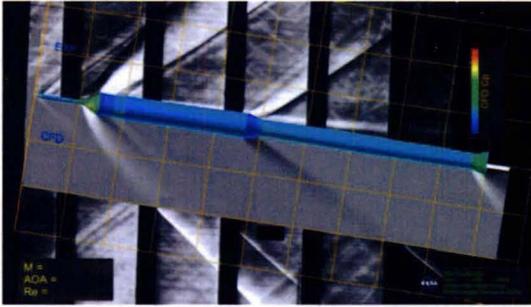


Figure 8. Combination of experiential Schlieren photograph with computational prediction of flow density as well as surface pressure distribution on Ares rocket.

Finally, ViDI has been used as a forensic tool. In instances where experimental results or computational data seem inconsistent, ViDI has allowed researchers to re-create plausible scenarios and experiment with different hypothesis to see if a scenario is physically possible and if the discrete data sources support the hypothesis.

#### HYTHIRM

One expansion of ViDI applications to flight testing involves the HYTHIRM (Hypersonic Thermodynamic Infrared Measurements) project. The HYTHIRM project is tasked with obtaining high resolution infrared imagery of hypersonic vehicles (flying greater than Mach 5) in flight to determine the heating on the vehicle. This is especially critical for reentry spacecraft, whose properly designed heat shields are essential to ensure adequate vehicle performance while ensuring the craft will not burn up due to inadequate protection. To date, the largest and most complex hypersonic vehicle is the Space Shuttle. After more than twenty years and one-hundred twenty flights there are still a number of important engineering questions concerning the fundamental flow physics involved in the aerodynamic behavior of the Space Shuttle during reentry [3].

The HYTHIRM project relies upon aircraft and ground based systems to locate the Space Shuttle during reentry. The vehicle is flying at velocities over Mach 18 (roughly 14,000 miles per hour) many hundreds of miles away. These imaging assets have to track the vehicle optically from close to horizon break to beyond the point of closest approach, which is typically about 30 nautical miles from the deployed infrared camera. For such a mission pre-planning is critical. Reentry trajectories are obtained from the Flight Dynamics Office (FDO) at the Johnson Space Center (JSC). These trajectories are placed in a custom ViDI program tied in with the COTS graphical software to plot the trajectory on a

virtual three-dimensional Earth (Figure 9). In addition, a Space Shuttle model is animated along the trajectory, and the program allows the user to specify an imaging assets (such as a particular telescope mount on an aircraft) and quickly determine the view of the shuttle from the telescope, based on the aircraft position and the point on the Shuttle trajectory being observed (Figure 10).

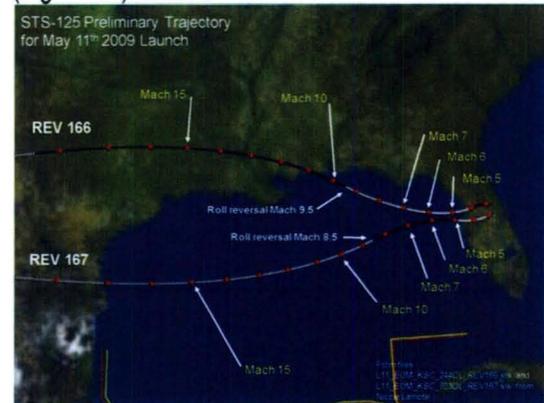


Figure 9. Typical trajectory plot for Space Shuttle reentry, shown here for the STS-125 mission.

The Space Shuttle reentry trajectories have a multitude of variables that can allow the vehicle to approach the Kennedy Space Center (KSC) by flying approaches ranging from the east of Cuba to the center of Mexico. An advanced understanding of how to position the imaging assets is a complex and essential task to ensure mission success. Additionally, it may be only hours before reentry that the actual path is known, and less than an hour before touchdown before a highly accurate track is computed. Real time updates are processed using the ViDI tools in the Mission Control Center in Houston and radioed up to the flight crews or land based telescope operators. To date, ViDI support has been provided to the two HYTHIRM missions, STS-119 and STS-125, both of which had complete success in acquiring and tracking the Space Shuttle from an airborne platform, and obtaining high resolution thermal imagery of the critical underside heat shield of the Space Shuttle during descent (Figure 11).

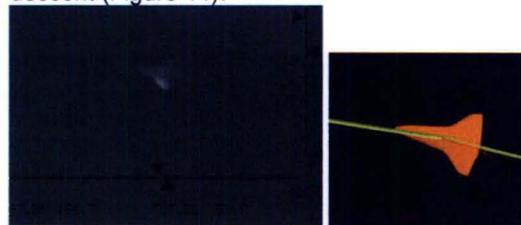


Figure 10. Comparison of flight data image from STS-119 (left) and ViDI virtual prediction prior to reentry (left).

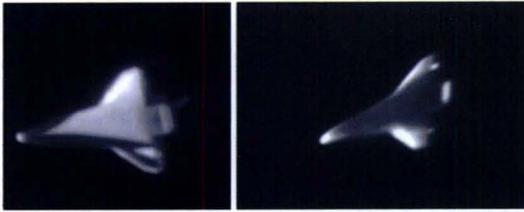


Figure 11. Unprocessed infrared images of STS-119 (left) and STS-125 (right).

Following the data capture, ViDI has been used for mapping the two-dimensional thermal images of the surface of the vehicle back to a three-dimensional geometry. This texture mapping technique relies on a unique application of spatially calibrating the data using a virtual reference. In order to properly texture map the data with scientific rigor, the data must be 'de-warped' to remove optical aberrations, perspective distortion, and foreshortening of the data on the vehicle due to the angle the vehicle makes with the camera. Traditionally, photogrammetric techniques are used that required knowledge of the exact position and orientation of the camera relative to the target object, or the Space Shuttle, in this instance. However, with this virtual calibration technique, this knowledge is not required. A reference pattern of equi-spaced dots is applied to the three-dimensional geometry of the Orbiter, and then a rendering of the Orbiter is made that matched to orientation of the actual Orbiter in the data image as closely as possible (most easily done by making the three-dimensional orbiter transparent and overlaying it on the actual data image, Figure 12.) This virtual calibration rendering is then processed by an image de-warping program (custom written at LaRC) to remap the two-dimensional data to remove all distortions and create a transformed image that appears as if it was taken from a camera directly perpendicular to the underside of the vehicle, with no perspective distortion. This image can then be mapped to the three-dimensional virtual model for data visualization.

#### FUTURE DEVELOPMENTS

With new measurement and visualization technologies emerging and maturing, ViDI will grow and adapt to interface with the new technologies. Despite using commercially available software, the current cost is still high enough to make the sharing and distribution of ViDI results challenging. The solution lies in platform-independent stand-alone applications that the user can open and run without additional software. These applications would use existing methods utilized by three dimensional game rendering engines or web browsers to create applications that allow the user to view and

manipulate the virtual environment. The next milestone for ViDI is three dimensional displays. With three dimensional presentation methods rapidly growing in use, the future of ViDI lies in three dimensional presentations. The ability to view and move through a virtual environment with the use of active or passive three dimensional techniques would only heighten the information and understanding gained by the user.

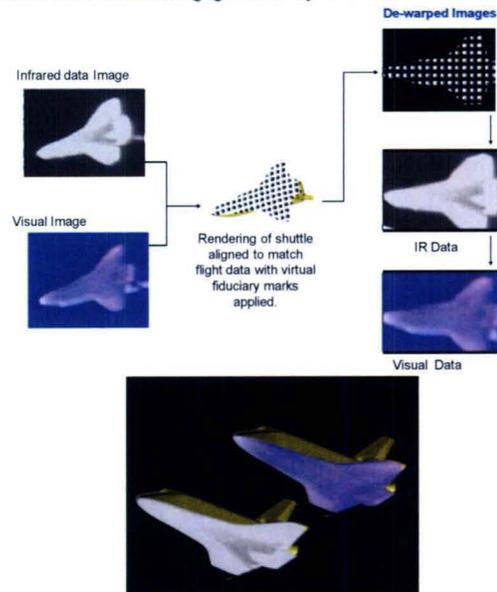


Figure 12. Data de-warping method for texture mapping thermal imagery data to Space Shuttle three-dimensional computer model.

#### CONCLUSION

The Virtual Diagnostic Interface software has become a powerful tool for a wide range of aerospace testing applications. The ability to rapidly combine experimental and computational data sets with three-dimensional geometry into one interactive environment gives the user a greater situational understanding throughout a test. ViDI will continue to grow its visualization capabilities in support of ground and flight test applications. As personal computing power continues to expand these techniques can still be improved upon and increased in scope. Every new project presents its own unique set of challenges to overcome and expands the capabilities of the ViDI software.

#### REFERENCES

1. Schwartz, R.J., "ViDI: Virtual Diagnostics Interface Volume 1-The Future of Wind Tunnel Testing" Contractor Report NASA/CR-2003-212667, December 2003
2. Schwartz, R.J., Fleming, G.A., "LiveView3D: Real Time Data Visualization for the Aerospace Testing Environment", AIAA-2006-1388, 44<sup>th</sup> AIAA Aerospace Sciences Meeting and Exhibit, Reno, Nevada, Jan. 9-12, 2006.
3. Berry, S., Horvath, T., Schwartz, R., Ross, M., Campbell, C., Anderson, B., "IR Imaging of Boundary Layer Transition Flight Experiments," AIAA-2008-4026, June 2008.

# Using Agent Base Models to Optimize Large Scale Network for a Large System Inventories

**Ramez Ahmed Shamseldin, Senior Design Engineer, Verizon Business and Shannon R. Bowling, Assistant Professor, Old Dominion University**  
[shamsrdin@hotmail.com](mailto:shamsrdin@hotmail.com) and [sbowling@odu.edu](mailto:sbowling@odu.edu)

**Abstract.** The aim of this paper is to use Agent Base Models (ABM) to optimize large scale network handling capabilities for large system inventories and to implement strategies for the purpose of reducing capital expenses. The models used in this paper either use computational algorithms or procedure implementations developed by Matlab to simulate agent based models in a principal programming language and mathematical theory using clusters, these clusters work as a high performance computational performance to run the program in parallel computational. In both cases, a model is defined as compilation of a set of structures and processes assumed to underlie the behavior of a network system.

## Introduction

The nature of digital networks, as described in [1] is comprised of non-variable bandwidth channels that transfer data. Furthermore, the growth of demand for transmitted visual data has been abruptly increased to satisfy customer needs, which resulted in the development of multi-video compression standards such as MPEG-2 [2], H.263 [3] and MPEG-4 [4].

In today's world, there are two kinds of video transmissions that have been established, one of them consists of full transmission of stored packets of video from a server to the customer's premises before playback begins; the other is a concurrent transmission which is under a certain restriction of quality of service (QoS) and serves as a real-time application.

The nodes in a network represent a video between customers who requested the service to watch certain movies. The selected video file is downloaded to the customer's computer site according to the system requested. It is also added to an inventory which can be allocated to several other sites in the future. System redundancy has been taken into account with regard to system needs for any overly excessive demands.

Agent-based models (ABM) are used in simulating social life, not only to understand environmental change and human roles, but to be attractive to many practitioners from a variety of subject areas. Human changes can happen through space and on different time scales. Many vital opinions of ABM and simulations are that numerous phenomena, even though system is complexity, dynamically or both combined dynamically complexity, can be described as autonomous agents that are relatively simple and follow certain rules for interaction.

Computer models are used for interesting research practices and testing theories within certain discipline structures. The progression fundamentals of a real-world structure are difficult to be observed and collecting data as well as controlling it under certain conditions is impossible. Assumptions based on theories for these structures can be implemented in a computer model that can perform and compare to this practical data.

Mainly, models used in this conference either use computational algorithms or procedure implementations developed by agent based models in any principal programming language or mathematical theory to underlie the behavior of a network system.

## Literature Review

Storage systems occur in a variety of contexts, including manufacturing, warehousing, and the service sector. Most storage systems do not deal with dynamic complexity because they are static and are usually in the form of physical warehouses. Storage systems dealing with materials can be either continuous or discrete storage. There are three major factors affecting storage systems: depending on size of storage, storage methods and layout of the storage system.

Storage throughput has been used as a measurement to describe the number of storage that can be retrieved per time period storage/retrieval (S/R). From there the size of the storage system is powered by throughput and cost parameters of transferring materials. Storage system mission controls input/output (I/O) functionality that can be determined by storage requirements is distributed centrally over time.

The storage method contains specification of unit load, S/R and storage equipments; these methods can be handled by machine or by humans and can be automatically launched by automatic guided vehicles.

The storage system layout, by using three dimensions - height, length and width - can identify the location of storage items. In this paper, the system layout control other storage parameters throughput and storage method will be discussed and five different storage system layouts will help to understand what the differences are between these physically traditional storage types and virtual layouts for our case scenario.

## Dedicated Storage Location

Every SKU (Stock Keeping Unit) is related to items in a warehouse and has a unique storage dedicated to it is location. Dedicated storage is characterized by the

assignment of fixed storage locations for the items stored in the warehouse [5]. For items to be allocated as measured unites used in the warehouse, they are assigned the cube per order index (COI).

In such cases, the more popular items have to be near the I/O point in the warehouse for reduction in travel time and travel distance according to the S/R. As an example, active items have to be placed in the most convenient and accessible place; this minimizes cost effectiveness and gets item to I/O points.

### **Randomized Storage Location**

The items in the storage warehouse are stored randomly in any available storage location. For an example, when the inbound load arrives for drop off, the item in the closest available slot is designated. This is known as first-in and first-out.

This is common in the case of randomized storage results when less storage space occurs. Having small sized parts stored in a space designed for large size parts wastes storage space, and for the same scenario, storing large parts to fit randomly can be impossible, so adjustable shelves may need to be used.

In randomized storage it is assumed each item of a certain product is equally likely to be recovered when multiple storage locations exist for the product, and the recovery operation is achieved. In the case where the warehouse is pretty full, the travel distances are significantly of the same "equal likelihood" [6].

### **Class-Based Dedicated Storage**

Class based storage is defined as a grouped of SKUs in one class. These classes are assigned to a dedicated storage spot, at the same time; these SKUs within an individual class are stored randomly and in a logical sequence.

The products are distributed according to their demand rates, among the number of classes and have a reserved a region within the storage area for each class. Accordingly, an incoming load is stored at an arbitrary available location for the same class.

We must look to the randomized storage location as a single class case of class-based storage policy where dedicated storage is counted as one class for each item.

In addition, the dedicated storage policy attempts to reduce travel times for S/V (storage/retrieval) by sorting the highest demand to the I/O point as well for class-based storage and calculate the product demand by COI [7].

### **Continuous Warehouse Storage**

Increasing demand for continuous recording of hundreds of millions data daily, a necessary storage media should have the capability to handle data volumes and data flow rates.

These types of data could be called detailed records (CDRs) - which are commonly used by the telecommunication industry - at an individual basis for

each customer. Software applications have been used to pose several challenges related to data volumes and data flow rates to data warehouses and to online analytical processing (OLAP).

### **Shared Storage**

Shared storage is widely used within the computer networking industry and addresses the needs of corporate computing environments for storage systems that propose scalability, availability and flexibility.

Storage systems are known as storage computer systems (hosts) and are connected to multiple individual hosts while using the shared storage by these hosts and are managed independently and historically viewed (host-attached storage).

Shared storage systems enabled by networking technology can provide high bandwidth. In turn, it offers several benefits for today's businesses, for example by improving quality of service (QoS) and increasing operational efficiency.

Moreover, as growing needs for shares (files, data, etc.) become necessary, it is necessary to prevent buying mainframe computer complexes and computer clusters where a modest number of cooperating computer systems share a common set of storage devices.

As computing environments have grown in industry, computer storage systems have grown in storage size and in number as the cost of equipment becomes more reasonable in order to increase the computing environments.

The main disadvantage is that the known computer storage systems processors have failed and replacement parts can be required to get the system back for full operation which wastes time and is followed by a typically propagation delay of the restoration of the data.

### **Virtual Warehousing**

As a physical location is not necessary to locate specific data content, data can be located within many virtual storage hosts. If a customer is looking for specific data to download, random locations can be used without specification and taking into account how many locations have been used.

The storage locations mentioned above, such as dedicated, randomized and class-based storage, can be used to benefit virtual storage warehouses with priority, size and rates of transferring data. On other hand, desired locations for data can be easily tracked and assigned to scale from the highest to the least high activities according to their demand.

At the same time, randomized storage results in a reduction in space and will be significant with regard to data travel time much less so than those traveling from a dedicated storage area.

Also, using other large-scale (shared) storage techniques, such as Internet, without specifying certain hosts is not problematic because data is already restored

within different hosts. Finally continuous warehouse storage techniques use network capability and add more data to different new hosts entering to networks as well using existing hosts.

### Network and Complexity

The internet-wide system is viewed as a large scale structure with an underlying physical connectivity that deploys real experimental studies to evaluate system architectures, however this is not possible. Instead, a randomly generated network connectivity structure is used and has been accepted at the beginning as a node degree distribution technique. A generator - also known as a software based solution - is used to generate network nodes which represent network autonomous systems (AS), original power laws and connectivity to the Internet.

The studies were then used to randomly generate networked topologies and provide precise analyses that show network modeling include [8]:

- Regular topology, such as liner, rings, trees, and stars;
- Well recognized topology, such as ARPANET or the NSFNET backbone;
- Arbitrarily generated topologies.

According to [9] when any two nodes have a relation, one link will be added with a probability depending on the distance between them given by:

$$p(u, v) = \beta \exp \frac{-d(u, v)}{L\alpha}$$

Where  $d(u, v)$  is a distance from  $u$  to  $v$ ;  $L$  is the maximum distance between two nodes,  $\alpha > 0$  and  $\beta \leq 1$ . However, this method does not obligate a large scale structure.

Albert et al [10] describe a system's components for a network as a complex system because of its functionality and attribute is largely to redundancy node connections. A large scale network consists of a complex communication network (CCN) along with groups of telecommunication carriers and ISPs (Internet Service Providers). It is almost impossible to analyze the infrastructure but this can be done within the limited boundaries of individual networks [11].

The redundancy of network connectivity, in other words scale-free network connections, represents an unpredicted degree of robustness for each kind of system, such as the internet, social networks or cellular (metabolic) networks. Network nodes break when faced with an extremely broken down communication rate.

### Agent Based Models (ABM)

Agent based model methodology has been applied to several studies, for example, social dynamics and communication and cooperation under ecological risk [12]; complexity in artificial life applications [13]; common dilemmas for ecological economics [14];

language evaluation [15]; armed forces contradictions [16]; and human social interaction interpolating with regeneration management [17].

Huigen [18] anticipates an ABM structure, called MameLuke, which will study human environment interaction. For like structures agents are categorized according to user definitions and determinations from the objective's study, meaning that individual agent sets can fit into multiple non-divergence categories. Potential option paths (POPs) are rule-based implementation through decision making, which depends on the agent's category.

ABM was significantly used in a spatial interest group within computational mathematical organization theory (CMOT). Today, on the other hand, simulations using ABM have expanded further than the original boundaries of use and have linked up with groups of people and cover work in a variety of different disciplines such as economics, biology, sociology, artificial intelligence, physics, computer science, archaeology and anthropology.

In the last few years, growth of ABM has been significant realized especially after releasing more helpful software toolkits. This was enough to attract many practitioners from different fields to simulate numerous subject areas. Some of the better known toolkits are Swarm, Repast, AnyLogic, MASON, Ascape and NetLogo.

Gilbert et al. [19] express an example of using ABM in the artificial intelligent field for developing cellular automata. At the time, Swarm, introduced in 1996, was the only agent based modeling simulation tool available [20].

The primary characteristic of an agent is the potential to make decisions on individual bases. On the other hand, agents, in a true case, are discrete events handled individually with a set of attributes and policies that influence its actions and decision making capability. In addition, an agent may have supplementary policies that modify its policies or attributes. An agent can be purposely independent in its atmosphere and in its interactions with other agents as well itself if not over an imperfect scope of posts. An agent has objectives to accomplish (not optimize) as goal bound within its actions. Furthermore, an agent is flexible and has the ability to learn and adapt its performance over time based on ongoing skills, in other word, some form of memory.

### Network Description and Functionality

The network that will be handled in this dissertation has a total of 250 nodes which represent the total number of customers carried by this network. These nodes are virtually connected by the internet and each address is recorded and knows the location of each customer. Each node in the network follows these assumptions:

- Each node is connected to the network and works online all year long with no bad connections.

- All nodes share the same bandwidth speed (uploading or downloading), and uploading bandwidth is half the speed of downloading bandwidth.
- The bandwidth speeds that will be used are limited to 128kb, 512kb, 1000kb, 2000kb and 5000kb per second.
- Uploading bandwidth and downloading bandwidth are two different streams and separated at each node.
- All nodes are spread all over the internet and connect to a separate network that can be located physically anywhere with no adverse affects on location or distance.
- All nodes can download simultaneously from the server with no affect on delay or connectivity.
- Each node can be used as virtual storage and upload any necessary file needed by another node upon request and can only to do this one node at a time.
- Each node can be downloaded from the server or from another node according to these guidelines:
  - Each node can download, at the maximum, from two locations and can be the server, the server and a single node or two nodes simultaneously.
  - Only one file can be downloaded at a time.
  - If the file exists in two virtual locations in network, the server will be exempt.
- All 250 nodes will be divided into five categories. Each category includes 50 nodes selected randomly. These categories are Actions, Crime, Comedy, Drama and Romance.
- Each node has an internal storage device and is selected randomly from a set of sizes: 50, 100, 150, 200, 250, 300, 350, 400, 450 or 500GB.

Each node will be studied throughout the year and is equivalent to 8760 hours download time and is evaluated for how many files has been selected and downloaded as these files are selected according to each node's preferences.

Selected files will be chosen randomly and according to each node's preferences. These files have the following characteristics and assumptions:

- The network will handle files of different sizes having different time durations, and is limited to 10,000 files and all files can be downloaded from the server.
- The 10,000 files will be divided to five categories. Each category includes 2,000 files

ranked from the highest priority to the lowest according to the power law degree distribution  $P(k) \sim k^{-y}$  with an exponent  $y$  range between 2 and 3. These categories are Actions, Crime, Comedy, Drama and Romance.

- The files can be downloaded from server, two nodes or a node and server at the same time by splitting the file's size to two batches - each batch contains half of the file.
- The file's batches will be downloaded either simultaneously by dividing the downstream bandwidth in half or downloading individually as the second batch will not start till the first one is completely downloaded.
- These files do not have expiration time but rather are replaceable inside the network's virtual storage which is located at the nodes. If the node's storage device reaches 75%, the files will be deleted according to the file's priority from low to high with the exception of the server. In this case, it will be remain stored as a reference for future requests.

All nodes will be able to download any files from a server at any time with no delay. All files are ranked according to its priority and stored in the server in the five different categories. Any node can search for any file across the entire network and download it in another in order to overcome network load and reach an optimum for the network.

#### **Networks and Their Dynamic Complexity Purpose**

Networks act as huge virtual storage warehouses that are dynamically changed over a period of time. The address of the nodes will be constant, but a file's location will be changed from node to other node with time and determined priority.

The duration of this study is equivalent to 8760 hours over an entire year. This study will follow several procedures to highlight and identify the purposes of this research. In addition, it will also simulate the generated data not only to show the output results but to also understand how the network works with layers of dynamic changes as the files flow across the network.

A network's complexity is represented by nodes and a server that are interactively and laterally ordering files from the server, other neighbor nodes or both at the same time. Also, the simulation of each of these nodes requires further study. The criteria and procedure will follow:

- Gathering data of inter-arrival time which was observed for each node during the 8760 hours.
- Gathering information about what type of file category customers were interested in as well as how many files per node were accessed.
- Gathering information from where the files were downloaded by each node.
- Calculating the arrival time by each node.

- Calculating the inter-arrival download time for each file by each node.
- Calculating the arrival download time for each node.
- Setting up groups of files preferences for each customer at the time of ordering and calculating the watched time at an individual basis for each file with download time to calculate the penalty time that occurred, each group will include random numbers of files between one and six files sets at a time which represent a set ordering one time group.

A network is a set of nodes connected virtually by an intranet at all times. These nodes share their contents at the same time and evaluate best practices to reach an optimal scenario as an ideal network which can change periodically over space and time.

Optimizing a network has a set of fundamentals that are required to help to simulate this kind of network and evaluate the results. In order to do that, the next section describes a set of identifying metrics that have to be fully gathered and understood in order to direct these types of networks to the second stage, which is ready to be integrated as an optimal large scale dynamic complexity network.

### Conclusions and Results

As noted before this data will be analyzed in different measures that are suitable for the type of experimental run and described in this paper. There is numerous data that will be impracticable to present here, but important ones will be presented as a key to show the differences and similarities and as the aim of this paper to show how networks can be dynamic and change complexity in time and space which then can be optimized based on specified performance measures.

Time Series was conducted on a Base Line, as shown on figures below, for different bandwidth loaded on the server with no nodes sharing the load with the server: the hard drive was not a study factor in this experiment and each node using the same value of bandwidth download speed from the server at the same time of analyzes.

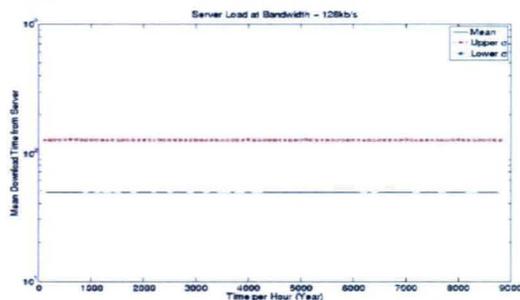


Figure 1: Server load for Base Line case scenario at download bandwidth of 128kb/s.

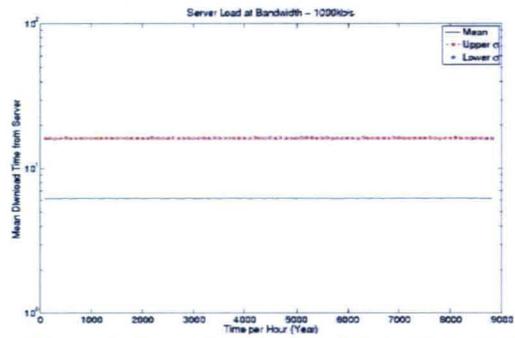


Figure 2: Server load for Base Line case scenario at download bandwidth of 1000kb/s.

Time Series was conducted on a H1, as shown on figures below, for different hard drive sizes and different download bandwidth speed loaded on the server with other nodes sharing the load with the server; the hard drive had different effects on the study in this experiment and each node using the same value of bandwidth download speed from the server and same value of hard drive size at the same time of analysis.

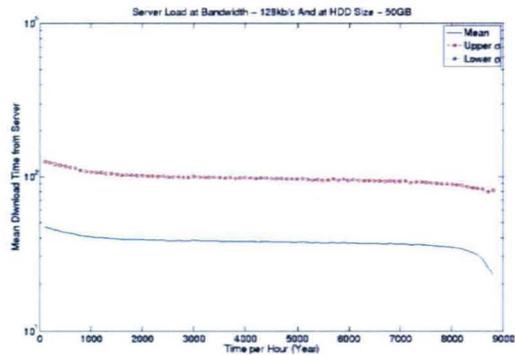


Figure 3: Server Load for H1 case scenario at hard drive size of 50GB and download bandwidth of 128kb/s.

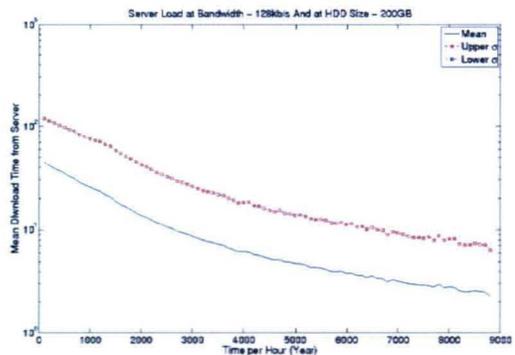
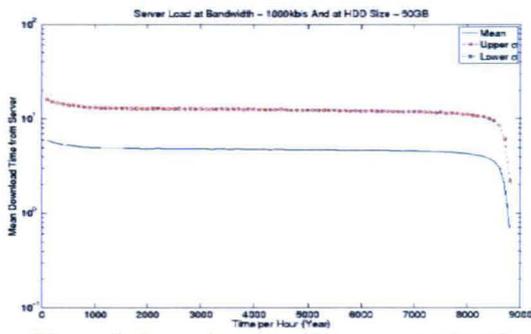
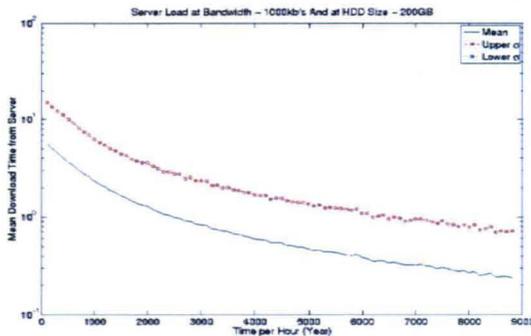


Figure 4: Server Load for H1 case scenario at hard drive size of 50GB and download bandwidth of 1000kb/s.



**Figure 5:** Server Load for HI case scenario at hard drive size of 50GB and download bandwidth of 1000kb/s.



**Figure 6:** Server Load for HI case scenario at hard drive size of 200GB and download bandwidth of 1000kb/s.

## References

- Militzer, M. and Suchomski, M. and Meyer-Wegener, K. (2009) "Improved p-domain rate control and perceived quality optimizations for MPEG-4 real-time video applications," *Proceedings of the eleventh ACM international conference on Multimedia, ACM Press, New York, NY, USA*, pp. 402-411.
- Boncelet, C. (2000) "Handbook of Image and Video Processing," Image noise models, Academic Press: San Diego, CA, USA, Alan C. Bovik.
- Rijkse, K. (1996) "H.263: video coding for low-bit-rate communication," *Communications Magazine, IEEE*, vol. 34, pp. 42-45.
- Militzer, M. and Suchomski, M. and Meyer-Wegener, K. (2003) "Improved p-domain rate control and perceived quality optimizations for MPEG-4 real-time video applications," *Proceedings of the eleventh ACM international conference on Multimedia, ACM Press, New York, NY, USA*, pp. 402-411.
- Malmborg, C.J. and Krishnakumar, B. (1989) "Optimal Storage Assignment Policies for Multiaddress Warehousing," *IEEE Transactions of Systems, MAN, AND CYBERNETICS*, vol. 19, no. 1.
- Francis, R.L. and McGinnis, L.F. and White, J. A. (1992) "Facility Layout and Location: An Analytical Approach," Englewood Cliffs: NJ.
- Hesket, T.J.L. (1963) "Cube-per-Order Index-D a Key to Warehouse Stock Location. Transportation and Distribution Management," vol. 3, pp. 27-31.
- Zegura, E.W. and Calvert, K.L. and Donahoo, M.J. (1997) "A Quantitative Comparison of Graph-Based Models for {Internet} Topology," *IEEE/ACM Transactions on Networking*, vol. 5, no. 6, pp. 770-782.
- Bernard M. Waxman (1991) "Routing of multipoint connections," *Broadband switching: architectures, protocols, design, and analysis, IEEE Computer Society Press*, ISBN = "0-8186-8926-9", Los Alamitos, CA, USA, pp. 347-352.
- Albert, R. and Jeong, H. and Barabasi, A.L. (2000) "Error and attack tolerance of complex networks," *Nature*, vol. 409, pp. 371-384.
- Claffy, K. and Monk, T.E. and McROBB, D. (1999) "Internet Tomography," *Nature*: <http://www.nature.com/nature/webmatters/tomog/tomog.html>.
- Andras, P. and Roberts, G. and Lazarus, J. (2003) "Environmental Risk, Cooperation, and Communication Complexity," *Adaptive Agents and Multi-Agent Systems*, vol. 2636, pp. 49-65.
- Menczer, F. and Belew, R. K. (1996) "From Complex Environments to Complex Behaviors," *Adaptive Behavior*, vol. 4, pp. 3-4.
- Jager, W. and Janssen, M. and Vries, H. D. and Greef, J. D. and Vlek, C. (2000) "Behaviour in Commons Dilemmas: Homo Economicus and Homo Psychologicus in an Ecological-economic Model," *Ecological Economics*, vol. 35, pp. 357-379.
- Bartlett, M. and Kazakov, D. (2004) "The Role of Environmental Structure in Multiagent Simulations of Language Evolution," *In Proceedings of the Fourth Symposium on Adaptive Agents and Multi-Agent Systems (AAMAS-4)*: AISB convention.
- Cioffi-Revilla, C. and Gotts, N. M. (2003) "Comparative Analysis of Agent-Based Social Simulations: Geosim and FEARLUS models," *Journal of Artificial Societies and Social Simulation*, vol. 6, no. 4.
- Deadman, P. and Gimblett, R. H. (1994) "The Role of Goal-Oriented Autonomous Agents in Modeling People-Environment Interactions in Forest Recreation," *Mathematical and Computer Modeling*, vol. 20, no. 8.
- Huigen, M. G. A. (2004) "First Principles of the MameLuke Multi-Actor Modelling Framework for Land-Use Change, Illustrated with a Philippine Case Study," *Journal of Environmental Management*, vol. 72, pp. 5-12.
- Gilbert, N. and Troitzsch, K. (1999) "Simulation for the Social Scientist," *Open University Press*: second edition, 2005.
- Nelson Minar and Rogert Burkhart and Chris Langton and Manor Askenazi (1996) "The Swarm Simulation System: A Toolkit for Building Multi-Agent Simulations," Santa Fe Institute: Working Papers, no. 96-06-042, <http://ideas.repec.org/p/wop/safiw/96-06-042.html>.

# The Application of Modeling and Simulation to the Behavioral Deficit of Autism

**John J. Anton**

*Principal Investigator, IVIR Inc.*

[janton@ivirinc.com](mailto:janton@ivirinc.com)

**Abstract.** This abstract describes a research effort to apply technological advances in virtual reality simulation and computer-based games to create behavioral modification programs for individuals with Autism Spectrum Disorder (ASD). The research investigates virtual social skills training within a 3D game environment to diminish the impact of ASD social impairments and to increase learning capacity for optimal intellectual capability. Individuals with autism will encounter prototypical social contexts via computer interface and will interact with 3D avatars with predefined roles within a game-like environment. Incremental learning objectives will combine to form a collaborative social environment. A secondary goal of the effort is to begin the research and development of virtual reality exercises aimed at triggering the release of neurotransmitters to promote critical aspects of synaptic maturation at an early age to change the course of the disease.

## 1.0 INTRODUCTION

Autism is one of a group of disorders known as autism spectrum disorders (ASD's). It is characterized by developmental disabilities that cause substantial impairments in social interaction and communication and the presence of unusual behaviors and interests. It begins before age three and lasts throughout a person's life. Autism occurs in all racial, ethnic, and socio-economic groups. It is also on the rise. New research indicates a possible mitigation strategy for autism. Deletions of genes or regions potentially involved in regulation of gene expression, suggests that defects in activity-dependent gene expression may be a cause of cognitive deficits in patients with autism. Therefore, disruption of activity-related synaptic development may be one mechanism common to at least a subset of seemingly heterogeneous autism-associated mutations. If the above hypothesis is true, then controlled environmental experiences coupled with calculated experiential exposure might be able to allow treatment through behavioral modification to facilitate learning in normal environments. But given the characteristics of autistic individuals, controlled environmental experiences are difficult to conduct. Reports from teachers, therapists, researchers and parents indicate

that many children with ASD show an affinity for computers. Current research indicates that computer-based tasks can motivate people with autism and encourage learning. Efforts to incorporate the ability to interact with and control virtual characters (avatars) within a computer generated environment are increasing. While there is good evidence that virtual environments are well accepted by individuals with ASD and of potential benefit to them, the use of the technology remains relatively unexplored. There is great potential to re-purpose technology and simulation content developed for the U.S. military that combines 3D video game technology with the constructivist principles of coaching, scaffolding and deliberate practice to help teach cultural awareness and non-verbal communications skills.

This paper describes a research effort designed to leverage this military technology, re-purpose game assets and adapt learning strategies to support virtual social skills training within a computer game environment in an effort to diminish the impact of social impairments on the lives of people diagnosed with ASD. If successful, the potential return on investment is enormous, both in actual cost savings, and in reduction of family suffering. Researching this area is an ethical imperative.

## 2.0 AUTISM



Autism is one of a group of disorders known as Autism Spectrum Disorders (ASDs). They include Autistic Disorder, Pervasive Developmental Disorder – Not Otherwise Specified (PDD-NOS, including atypical autism), and Asperger Syndrome. These conditions all have some of the same symptoms, but they differ in terms of when the symptoms start, the severity of the symptoms, and the exact nature of the symptoms. The three conditions, along with Rett Syndrome and childhood disintegrative disorder, make up the broad diagnosis category of pervasive developmental disorders. ASD begins before the age of 3 and lasts throughout a person's life. It occurs in all racial, ethnic, and socioeconomic groups and is four times more likely to occur in boys than girls. [1]

## 3.0 THE COST



Dr. Michael Ganz, MS, PhD, Assistant Professor of Society, Human development and Health at the Harvard School of Public Health, and respected expert on the societal costs associated with autism and its related disorders claims that Autism is a very expensive disorder costing upwards of \$35 billion in direct (both medical and

nonmedical) and indirect costs to care for all individuals diagnosed each year over their lifetimes. [2]

In a paper published in the Archives of Pediatric Adolescent Medicine, Dr. Ganz details the substantial costs resulting from lifetime care and lost productivity of individuals with autism, their caretakers and society in general. [3] Direct costs measure the value of goods and services used and indirect costs measure the value of lost productivity due to autism. Physician and other professional services, hospital and emergency department services, drugs, equipment and other supplies, and medically related travel and time costs are typical components of direct medical costs. Special education, transportation, childcare and babysitting, respite care, out-of-home placement, home and vehicle modifications and supported employment services are typical components of direct nonmedical costs. Indirect costs are the value of lost or impaired work time (income), benefits, and household services of individuals with autism and their care givers because of missed time at work, reduced work hours, switching to a lower-paying but more flexible job, or leaving the workforce. Behavioral therapies, which are the largest component of direct medical costs, make up 6.5% of total discounted lifetime costs. Those costs, combined with very limited to non-existent income for their adult children with autism combined with potentially lower levels of savings because of decreased income and benefits while employed, may create a large financial burden affecting not only those families but potentially society in general.

## 4.0 AUTISM IS INCREASING

The U.S. Center for Disease Control and Prevention (CDC) states that it is clear that more children than ever before are being classified as having ASDs, however, it is unclear how much of this increase is due to changes in how we identify and classify ASDs in people, and how much is due to a true increase in prevalence. By current standards, ASDs are the second most common serious developmental disability after mental retardation/intellectual impairment. The impact of having a developmental disability is great for families affected and for the community services that

provide intervention and support for these families. It is important that we treat common developmental disabilities, and especially ASDs, as conditions of urgent public health concern, do all we can to identify children's learning needs, and start intervention as early as possible to give all children the chance to reach their full potential. The CDC also states that ASDs can often be detected as early as 18 months and children in high-risk groups - children with a parent or sibling with an ASD - should be watched particularly closely. Studies have shown that among identical twins, if one child has autism, then the other will be affected about 75% of the time. In non-identical twins, if one child has autism, then the other has it about 3% of the time. Also, parents who have a child with an ASD have a 2% - 8% chance of having a second child who is also affected. [4], [5]

## 5.0 RESEARCH DIRECTIONS



New research conducted by Eric M. Morrow et al., of the Division of Genetics at Children's Hospital Boston and Harvard Medical School, states that the regulation of expression of some autism candidate genes by neuronal membrane depolarization, suggests the hypotheses that neural activity-dependent regulation of synapse development may be a mechanism common to several autism mutations. Early brain development is driven largely by intrinsic patterns of gene expression that do not depend on experience-driven synaptic activity. Postnatal brain development requires input from the environment that triggers the release of neurotransmitters and promotes critical aspects of synaptic maturation. During this process, neural activity alters the expression of hundreds of genes, each with a defined temporal course that may be particularly vulnerable to gene

dosage changes. The connection between experience-dependent neural activity and gene expression in the postnatal period forms the basis of learning and memory, and autism symptoms typically emerge during these later stages of development. This finding that deletions of genes regulated by neuronal activity or regions potentially involved in regulation of gene expression in autism suggests that defects in activity-dependent gene expression may be a cause of cognitive deficits in patients with autism. Therefore, disruption of activity-related synaptic development may be one mechanism common to at least a subset of seemingly heterogeneous autism-associated mutations. [6]

If the above hypothesis is true, then controlled environmental experiences coupled with calculated experiential exposure might be able to allow treatment, behavioral modification, and learning to occur in normal environments.

## 6.0 THE USE OF COMPUTERS AND VIRTUAL REALITY



Reports from teachers, therapists, researchers and parents indicate that many children with ASD show an affinity for computers. Previous research has shown that computer-based tasks can motivate people with autism and encourage learning [7]. The social and communication deficits of ASD make it difficult to engage in social interaction, and therefore, access to learning opportunities in these social settings is limited. Computer-based experiences in constructed social environments mitigate this deficiency. New research efforts incorporate the ability to interact with and/or control virtual characters (avatars) within a virtual environment. The Authorable Virtual Peers (AVP) program at Northwestern University uses language-based avatars to enable children diagnosed with ASD to learn about language and social interactions

through collaborative storytelling [8]. The use of avatars has also been found to increase facial recognition, emotion recognition and social interaction skills for children with ASD through repeated practice of multiple different interactions [9]. This affirms the commonly used approach of repeated practice in a natural setting to successfully teach skills to those with ASD.

- Virtual reality uses sight and sound more than touch: auditory and visual stimuli have been found to be most effective in teaching abstract concepts to people with autism.
- In the virtual environment, input stimuli can be modified to a tolerable level.
- The environment can be altered gradually to teach generalization and cross-recognition.
- Virtual reality offers a safe learning environment in which the individual may make mistakes that might be physically or socially hazardous in the real world.

A research study on autistic children conducted at the University of Haifa, focused on the transfer of skills mastered within a virtual environment to the real world and found that the intelligence level or severity of the autism does not affect the ability to understand the system and therefore is an important way to improve their cognitive and social abilities [11]. Six autistic children, ages 7-12 spent one month learning how to cross virtual streets, to wait for the virtual light at the crosswalk to change, and to look left and right for virtual cars using a simulation programmed by Yuval Naveh. The children in the study showed substantial improvement throughout the learning process. At the beginning of the study, the average child was able to use the 2<sup>nd</sup> level of the software, while by the end; they mastered the 9<sup>th</sup> level, which is characterized by more vehicles traveling at a higher speed. A local practice area with a street and crosswalk, complete with traffic signals, was used for validation. The children's ability to cross the street safely was tested in this area, evaluating for example, whether they stopped to wait on the sidewalk or waited for a green light before crossing. The children were brought to the practice area before and after their virtual learning. Here too, the children exhibited an improvement in their skills,

following the training on the virtual street, with three of the children showing considerable improvement. One of the study participants, a 16 year old, had participated in the past in a road safety program in the school, but he was not able to learn how to cross the street safely. Following learning the skill in a virtual environment, he learned how to stop on the sidewalk before stepping into the street, to look at the color of the traffic light, to cross only when the light was green and to cross without waiting too long.

## **7.0 LEVERAGING CURRENT MODELING AND SIMULATION TECHNOLOGY**

While there is good evidence that virtual environments are well accepted by individuals with ASD and of potential benefit to them, the use of this technology remains relatively unexplored. New computer-based game technologies increasingly integrate a social as well as a cognitive component. There is potential to leverage this technology in an innovative new direction to provide a context that can scaffold social interactions and communications skills for children with ASD.

The ability to simulate, test and assess cognitive and social skills within a virtual environment provides professionals with rigorous practice and guidance to increase their chances of success in situations that may not be safe or cost effective to perform in a live training environment.

Simulations and games can supplement traditional training methods by providing challenges and experiences that closely approximate a complex situation in the real world where students must think in real-time and the course of events will be determined by their decisions. Students with ASD may be able to develop a deeper understanding of the knowledge presented and retain that information better when it is learned through the process of repeatedly solving problems in realistic situations. This approach places the learner in a "real-world" environment, which allows the student to learn in context and apply what they have learned. It is this contextual experience of knowledge acquisition in an authentic environment that facilitates the learner to create their own constructs that can be applied to new and unfamiliar situations. There is also an opportunity to provide practical, hands-on

experience in situations that cannot easily be practiced using real scenarios.

There is also great potential to re-purpose technology and simulation content developed for the U.S. military that combines 3D video game technology with the constructivist principles of coaching, scaffolding and deliberate practice to help teach cultural awareness and non-verbal communications skills. The emerging importance of cultural identity and its inherent frictions make it imperative for soldiers and leaders to understand societal and cultural norms of the populaces in which they operate and function. Much of this communication occurs through nonverbal channels, especially when language skills are minimal or absent. It is totally feasible to leverage this military technology, re-purpose game assets, and adapt learning strategies to support virtual social skills training within a 3D video game in an effort to diminish the impact of social impairments on the lives of people diagnosed with ASD.

## 8.0 SPECIFIC APPLICATION STRATEGIES



The symptoms and characteristics of autism can present themselves in a wide variety of combinations. The uniqueness of each individual with autism and the context of their lives provide interesting design challenges for the successful creation and adoption of technologies for this domain. The first goal is to enable children with ASD to not only interact with a virtual environment, but also to build social skills. A socially relevant scenario can be designed which encourages human interaction with artificially intelligent avatars. Inside the virtual world, which includes settings commonly encountered in everyday life such as restaurants, shops, offices, parks and

other social places, autistic individuals will be able to interact with other real people's avatars as practice. The user interface and scripts will be extremely clear and simple, and since previous ASD research has shown benefits of storytelling, each game scenario may consist of a short vignette design to elicit response from the student. Vignettes used for the project may include:

- Teaching the student how to interact through social stories, modeling, role-playing and other activity-based learning.
- Conflict resolution and managing disagreement with compromise and recognizing the opinions of others. Learning not to respond with aggression or immature mechanisms.
- Turn taking and other socially acceptable mannerisms such as verbal interactions, changing conversational topics, introductions to new people and others.

Individuals with autism will encounter prototypical social contexts via a computer interface and will have to interact with 3D avatars within the game that have predefined roles, tasks, and visible body language. Within the context of the game's scenario, the student will trigger events, which equate to learning objectives that they must successfully interact with to advance further into the game. Incremental learning objectives will eventually combine to form a fully collaborative social environment. The initial sequence of tasks will most likely follow a linear model progressing from simple tasks to more complex ones. The effects of the student's actions, will impact the behaviors of the other avatars within the scenario in a realistic fashion. Further, as the scenario reacts to the student's input, it will track performance and provide feedback concerning the consequences of particular actions and or omissions. In terms of people with ASD learning social behaviors, errors do need to be made to support learning. Therefore, a balance must be made between allowing the user to make errors and clearly showing what options are available at any given time within the game.

A secondary goal is to begin the research and development of VR exercises aimed at triggering the release of neurotransmitters to promote critical aspects of synaptic maturation at an early age, to change the course of the disease. Rigorous scientific

evaluation is necessary to estimate the likely benefits of this approach and its application to the individual. Research should also attempt to evaluate the contribution of this technology to any observed gains through comparison with traditional teaching approaches.

Web-based instruction and testing via "intelligent" computer simulations of typical social environments will prove an efficacious means for people with ASD to acquire social skills. Computer-based media allows people with autism continuous access to the curriculum, while concurrently allowing researchers to track the frequency of exposure and/or duration of exposure to a given skill (e.g. time logged onto vignettes and testing). It is expected that participants may engage in virtual skills training for longer durations when compared to traditional lecture-based curriculum.

For the parents, teachers and families of autistic children, understanding and active participation can be critical to their development and eventual independence. Adolescents often play games. These games, whether played in isolation or in a group setting, may be an effective reinforcement for skills that can be practiced in the home with family members and peers. Combining skill instruction with the gaming experience offers the gamer a chance to repeatedly practice skills. Although the skills learned may be constrained to the capabilities of each individual, such strategies may prove to be the gateway to increased socialization and acceptance by peers.

The vision is to use online, PC-based games and immersive 3D environments that leverage existing DOD research and development in modeling, simulation, serious gaming, performance assessment, and after action review technologies. The goal is to establish a seamless management and delivery capability to provide a distributed virtual environment where skills can be practiced and honed as a student interacts within each prescribed scenario. Virtual reality and gaming applications for social skills may prove to be less resource intensive than traditional in-vivo and "Video Self Modeling" training models. Further, this technology may provide a more engaging, and socially controlled environment in which

autistic individuals can practice social skills without excessive distractions. Attention must be paid to new collaborative technologies such as massive multi-player environments that allow interactive experiences for groups as well as individuals. Automated support tools should be investigated to help teachers perform in depth assessments of student performance and to identify and mitigate critical behaviors by providing essential feedback. In this way, distributed and collaborative virtual environments can be incorporated into the continuum of ASD treatment to work in tandem with the full spectrum of other case management interventions [11].

Based on positive outcomes from this effort, additional studies could also begin to look at which social skills acquired and practiced via virtual models will generalize to school, home, and community environments. Scenario authoring capabilities should also be investigated to provide the ability to modify and insert new resources into the virtual environment as required by the various ASD treatment interventions. The ability to tailor virtual scenarios to specific student needs is intended to help the student contextualize social situations and events. The application of this augmented virtual reality technology contained within the science of modeling and simulation could produce tremendous synergy in mitigating the treatment and educational interventions to reduce the rising cost in resources, as well as in pain and suffering. Researching this area is an ethical imperative.

## 9.0 END NOTES

[1] U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, <http://www.cdc.gov/ncbddd/autism/overview.htm>, accessed October 21, 2008

[2] Ganz ML. The Costs of Autism. In: Moldin SO, Rubenstein JLR, eds. *Understanding Autism: From Basic Neuroscience to Treatment*. Boca Raton, Fla: Taylor and Francis Group; 2006.

- [3] Ganz ML. The Lifetime Distribution of the Incremental Societal Costs of Autism. *Archives of Pediatric Adolescent Medicine*, Apr. 2007; 161. Downloaded from www.archpediatrics.com on July 29, 2008
- [4] Boyle C, Van Naarden Braun K, Yeargin-Allsopp M. The Prevalence and the Genetic Epidemiology of Developmental Disabilities. In: Genetics of Developmental Disabilities. Merlin butler and John Meany eds. 2004 (Table 3, p. 716-717).
- [5] Muhle R, Trentacoste V, Rapin I. The Genetics of Autism. *Pediatrics* 2004; 113; 472-486
- [6] Morrow, Eric M., et al. Identifying Autism Loci and Genes by Tracing Recent Shared Ancestry. *Science* 321, 218 (2008); DOI: 10.1126/science.1157657
- [7] M. Hart, 'Autism/excel study' ASSETS 2005: *Seventh International ACM SIGACCESS Conference on Computers and Accessibility*, (2005).
- [8] Tartaro, A., & Cassell, J. (2006). Authorable Virtual Peers for Autism Spectrum Disorders. Paper presented at the *Combined Workshop on Language-Enabled Educational Technology and Development and Evaluation for Robust Spoken Dialogue Systems at the 17<sup>th</sup> European conference on Artificial Intelligence (ECA106)*, Riva del Garda, Italy.
- [9] Hopkins I.M. (2007). Demonstration and Evaluation of Avatar Assistant: Encouraging Social Development in Children with Autism Spectrum Disorders. *Dissertation Abstracts International*, 68, 5-B.
- [10] University of Haifa (2008, January 29). Virtual Reality Teaches Autistic Children Street Crossing, Study Suggests. *ScienceDaily*.
- [11] Smith, B. (2007). Virtual Self and Peer Modeling to Build Social Competence in Children who have Autism, (*Unpublished Report by Engineering & Computer Simulations (ECS)*).



# Blended Training for Combat Medics

Jennifer Fowlkes<sup>1</sup>, Sandra Dickinson<sup>2</sup>, and Todd Lazarus<sup>3</sup>

<sup>1</sup>CHI Systems, Inc.

[jfowlkes@chisystems.com](mailto:jfowlkes@chisystems.com)

<sup>2</sup> US Army Research Development & Engineering Command (RDECOM)  
SFC Paul Ray Smith Simulation & Training Technology Center (STTC)

[Sandra.Dickinson@us.army.mil](mailto:Sandra.Dickinson@us.army.mil)

<sup>3</sup> University of Central Florida

Institute for Simulation and Training

[tlazarus@ist.ucf.edu](mailto:tlazarus@ist.ucf.edu)

**Abstract.** Bleeding from extremity wounds is the number one cause of preventable death on the battlefield and current research stresses the importance of training in preparing every Soldier to use tourniquets. HapMed is designed to provide tourniquet application training to combat medics and Soldiers using a blended training solution encompassing information, demonstration, practice, and feedback. The system combines an instrumented manikin arm, PDA, and computer. The manikin arm provides several training options including stand-alone, hands-on skills training in which soldiers can experience the actual torque required to staunch bleeding from an extremity wound and be timed on tourniquet application. This is more realistic than using a block of wood to act as a limb, which is often how training is conducted today. Combining the manikin arm with the PDA allows instructors to provide scenario based training. In a classroom or field setting, an instructor can specify wound variables such as location, casualty size, and whether the wound is a tough bleed. The PDA also allows more detailed feedback to be provided. Finally, combining the manikin arm with game-based technologies, the third component, provides opportunities to build knowledge and to practice battlefield decision making. Not only do soldiers learn *how* to apply a tourniquet, but *when* to apply a tourniquet in combat. The purpose of the paper is to describe the learning science underlying the design of HapMed, illustrate the training system and ways it is being expanded to encompass other critical life-saving tasks, and report on feedback received from instructors and trainees at military training and simulation centers.

## 1.0 INTRODUCTION

Combat medics provide front line trauma care often in the heat of a battle, with limited resources, and under enormous stress. In modern warfare, medics serve as soldiers first, in keeping with the philosophy that "*The best medicine on the battlefield is fire superiority*"[6] but must be able to transition to a medic role quickly and decisively in accordance with the tactical situation. In these situations, medics respond to complex medical trauma. Most injuries on the battlefield are the result of explosions (from landmines and Improvised Explosive Devices) and gun shot wounds, each of which presents challenges. Combat medics must not only understand the nature of war related injuries but also the implications for procedures that will be effective. Moreover, as Mazurek and Burgess [5] point out, decisions

regarding casualty treatment are made based not only on the special knowledge of war-related injuries, but also on the understanding of the tactical environment, current location, resources available, and own capabilities. Clearly, there are significant training needs in regard to combat medics

The purpose of this paper is to describe the HapMed training system which is aimed at fulfilling at least some of the combat medic training needs. HapMed was designed to provide hands-on training to combat medics, addressing the three most common preventable point-of-injury deaths in combat: tourniquet application for hemorrhage control, needle chest decompression to relieve tension pneumothorax (air trapped within the chest cavity resulting in a collapsed or compressed lung), and creating an airway (this is generally a preventive measure to ensure a casualty airway remains open while

medic is treating another soldier). The Army has recently enhanced training of these and related tasks within combat medic training programs, including the Tactical Combat Casualty Care (TC3) courses [1]. TC3 was developed in response to the growing recognition that medical care on the battlefield requires different skills and mindsets compared to civilian approaches.

The design of HapMed was supported by cognitive engineering techniques including focused cognitive task analysis (CTA) to identify critical cues, decisions, and strategies that support task performance; observation of combat medic training; user feedback as prototypes and artifacts were developed; and training research. Our front end analyses revealed several challenges in regard to combat medic training [8]. Specifically, we faced challenges in four areas.

**1.1 Challenge 1: Need for Hands-on Training**

Despite the recent changes in medic training (e.g., TC3), there are few opportunities to receive hands-on practice. A well-practice skill such as tourniquet application may require at least tens, and maybe hundreds, of training trials, but there are few existing opportunities to get extensive hands-on practice. A common approach is to have students apply a combat application (CAT) tourniquet to a 2x4 board. While students may learn procedural knowledge this way, they do not receive tactile feedback on how tight a tourniquet must be to stop blood flow. This is a significant training need for tourniquet application as this quote suggests.

In the case of lower extremity wounds, which give rise to the most severe hemorrhage controlled by tourniquet, it has been my observation, that too few doctors, much less their assistants, have a concept of the constricting pressure required about the thigh to abolish the flow of blood. *(Emergency Treatment and Resuscitation at the Battalion Level by MAJ Meredith Mallory 1954)*

There are other related drawbacks to current training. Students do not gain the knowledge of

- how long it will take to apply a tourniquet;
- the influence of the size of the casualty on how long it will take to apply a tourniquet;

- complications that can arise when applying a tourniquet
- differences among the different types of tourniquets.

Thus, training systems are needed that provide hands-on training and support the acquisition of this type of knowledge.

**1.2 Challenge 2: Tourniquets are Associated with Misinformation**

Until recently, the tourniquets carried by soldiers were ineffective – they did not provide a mechanical advantage that would help apply enough force to stop bleeding from a significant extremity wound and their narrow widths resulted in excessive damage to limbs. As a result, tourniquet use was shunned by civilian medicine and myths abounded in both military and civilian sectors, some of which are summarized in Table 1. Recent tourniquets, including the CAT and the Special Operations Forces Tourniquet greatly minimize or eliminate these problems and recent research suggests they are highly effective [7]. Thus, our analysis indicated that hands-on training should be supplemented by didactic instruction that addressed misconceptions surrounding the use of tourniquets.

**Table 1:** Misinformation associated with tourniquet use

Myth	Current Guidance
It is good to occasionally loosen a tourniquet to allow some blood to get to the damaged limb.	The tourniquet should remain in place with the blood flow completely stopped until it can be removed. Occasionally loosening the tourniquet may result in death of the casualty.
Once a tourniquet is applied the casualty will lose his or her limb.	A tourniquet can remain in place for several hours without causing major damage to the limb.
Tourniquets should only be used as a last resort.	Tourniquets should be applied to extremity wounds as soon as tactically feasible.

**1.3 Challenge 3: Medics serve in low density MOS assignments**

Medics are one of the most common military occupational specialties, yet perhaps 70% of

medics serve in low density assignments, meaning that they are assigned to infantry or armor battalions and their numbers are small compared to other MOSs in their units. One consequence is that training resources are limited. Thus, training systems supporting medics must be low cost, durable, and reliable.

#### **1.4 Challenge 4: Combat medics must learn to perform under intense battlefield stress**

During the CTA, the harsh realities and complex environment in which combat medics work became clear. Medics perform in environments defined by intense time pressure, danger, complex and sometimes horrific injuries, and the many factors that impede or play a part in tactical decision making including tactical, medical, interpersonal, and environmental stressors. The importance of training medics to work effectively under these conditions was reinforced through interviews with subject matter experts and reviews of lessons learned. However, in examining existing curricula, we found few opportunities to systematically practice combat medicine under these “fog of war” conditions. Lane training and field exercises provide immersive, realistic practice, yet there are a very limited number of these opportunities. Moreover, in field exercises with their assigned units, it is often the case that combat medicine is not the primary focus of training and sometimes it is not addressed at all.

### **2.0 HAPMED TRAINING SYSTEM**

Our solution to the challenges we uncovered during our front end analyses was to design and develop a blended training solution. By blended solution we mean incorporating complementary training interventions linked to the challenges or training needs discussed above. The HapMed training system, currently in prototype form, includes an instrumented manikin arm to provide hands-on skills training for tourniquet application, a PDA that wirelessly controls the arm to provide scenario-based training; and game based didactic instruction to further build knowledge and skills required of medics on the battlefield. Finally, we performed training research from which we determined that haptic and audio stimuli can be combined inexpensively to serve as surrogate stressors. These components are described below.

#### **2.1 HAPMED Manikin Arm for Tourniquet Training**

The HapMed manikin arm, pictured in Figure 1, was designed to provide stand-alone, hands-on skills training (as well as interact with the other system components) in which trainees can experience the actual torque required to staunch bleeding from an extremity wound and be timed on tourniquet application [4]. The features designed into the arm were based on the essential cues for tourniquet application identified from the CTA. The hardware components were based on our assessment of low cost technologies that could be sufficiently ruggedized. In the current design of the prototype,

- Bleeding is depicted through LED arrays clustered at four different locations to represent four possible wound areas. Wound sites can be selected by the trainee or instructor.
- Pressure sensors within the arm gauge the amount of pressure being applied via a tourniquet (or through other source of pressure such as squeezing with the hands at pressure points). The sensors were calibrated from data obtained from surgical applications of tourniquets. (e.g., [3])
- As the tourniquet is tightened, LED lights on the arm indicate to the trainee that the bleeding is being slowed by the tourniquet. When the trainee has tightened the tourniquet enough, per amount of torque needed on an actual human, the red lights turn to green indicating that the bleeding has been successfully stopped.
- A timer on the arm shows the number of seconds that it took to stop the bleeding.
- If a tourniquet is loosened once having been applied, the “bleeding” will begin again (and the timer will start again).



**Figure 1:** HapMed manikin arm prototype.

- A pulse can be felt that weakens as the tourniquet is applied.
- The arm will respond realistically to almost any tourniquet. This provides a tool to be used

by medical personnel to try out or compare different tourniquets.

- The arm can be affixed to a full body manikin or to a wall (see Figure 3).

During the course of development a great deal of feedback was obtained from the combat medic community regarding the arm. In addition, we conducted a more formal try out with a National Guard unit during which the arm was incorporated into combat medic and combat lifesaver training. This work has resulted in the design of a system suitable for use by the military.

### 2.2 HapMed PDA

The sensitivity of the pressure sensors on the HapMed arm can be manipulated through wireless connections so that different body sizes can be simulated, a feature that we incorporated into a PDA interface. The HapMed PDA provides a portable and intuitive interface to specify training scenario parameters such as wound location, casualty size, and whether the wound will require more than one tourniquet---these are all distinctions a medic or combat lifesaver needs to know when applying a tourniquet in combat. The PDA also provides performance feedback to trainees. Specifically, the PDA shows the location of the tourniquet(s) applied to the arm and indicates whether they have been placed correctly (2-4 inches above the wound site), the amount pressure being applied relative to the amount of bleeding, time to control bleeding, and status of the casualty. Finally, the PDA provides prompts that can be used by an instructor to facilitate discussion about the variables important in tourniquet application. The PDA can be used by an instructor to control the manikin arm in a classroom or in the field as part of lane training.

Figure 2 shows one of the PDA screens. Figure 3 shows an instructor using the PDA to wirelessly control the arm during combat medic training. Our usability evaluations showed that instructors could easily and immediately use the PDA. Much of the feedback we obtained pertained to additional controls and features that instructor's would like to see incorporated into the PDA.

### 2.3 Game-based Training

A computer-based training framework was designed that would allow trainees to access game-based training modules, combat medic

stories, and demonstrations that would interact wirelessly with the manikin arm. The game-based training modules were the components that we initially developed for the HapMed prototype. These included training modules focused on myths associated with tourniquet applications and identification of injuries suitable for tourniquets (screen for which is shown in Figure 4). In addition, a scenario-based decision making exercise was created. To

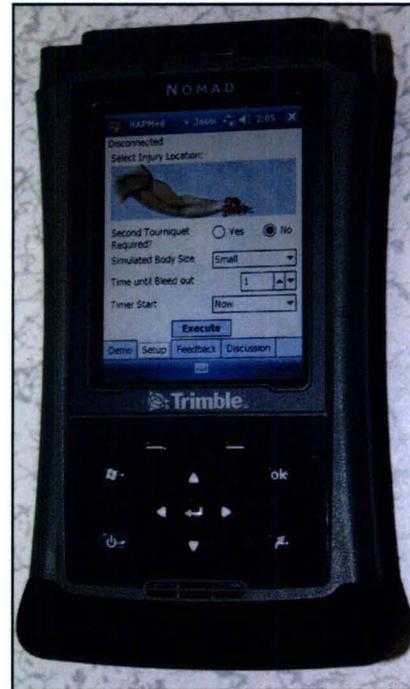


Figure 2: HapMed PDA scenario control screen.



Figure 3: Instructor using the PDA.

enhance the training, each of the training modules sets the stage for the trainee through an advanced organizer, provides performance performance feedback, and concludes with a guided reflection exercise. The game-based instruction has, thus far, received the least attention in terms of validation work. We intend to evaluate the effectiveness of the game-based training format compared to more traditional methods for information presentation in future work.



**Figure 4:** Card sort game-based training module screen.

## 2.4 Stress Training

Training research was conducted examined to how noise and vibrating tactile devices, or factors, can be used as surrogate sources of stress during the performance of a primary tourniquet application task and a secondary task requiring participants to identify pictures of soldiers as friendly or enemy. In combat, soldiers encounter stress from multiple sources including loss of sleep, extremely high levels of physical and psychological discomfort, extended periods of increased vigilance, and intense danger. The research examined the interactions between vibration stress (vibration presented through a haptic vest), audio stress (white noise) and time pressure on task performance for a tourniquet application task. Participants were 33 male and 12 females.

The results showed that the combined stress condition caused worse performance than the single-stress conditions, which caused worse performance than the control condition. Data for

the time to apply a tourniquet for each of the conditions is shown in Figure 5. Also, time pressure caused worse performance than no time pressure. This was true on both the primary task (tourniquet application) and secondary tasks (identifying images as friendly or foe), which indicates that resources were so limited that participants could not maintain performance on the primary task, even if they neglected the secondary task.

The performance decrements obtained in this research are operationally significant. For example, tourniquet application time was increased on the order of 10-15 seconds for the vibration and noise conditions, and on the order of 20-30 seconds for the combined condition. For a task in which casualties can die within four minutes, 30 seconds is a significant loss of time. This decrement, combined with the time distortion that occurred, is startling. If a medic's performance is slowed by 30 seconds, and he thinks he has been applying a tourniquet for one minute instead of two, grave consequences could likely be seen. Thus, the results indicate that vibration, audio, and time pressure, all of which are inexpensive interventions, could be added to the HapMed system to provide surrogates for battlefield stress. By adding system components, trainees could 1) observe stress effects in themselves, and learn strategies for coping with stress.

## 3.0 CONCLUSION AND FUTURE PLANS

Combat medicine lessons learned from recent engagements have led to a heightened awareness within the Department of Combat Medic Training of the need for incorporating warrior tasks during medical skills training and increased immersion in relevant combat medic training scenarios. In response, training interventions are being explored, many of which involve simulation [7]. The HapMed system was based on a needs analysis to design a training system that met some of the current requirements and that complemented other training approaches.

We are implementing HapMed within an overall "build a body" concept, in which different user groups can select the components of the training system that are relevant to their training needs. For example, through ongoing work we are adding instrumented torso and head manikins to support the cricothyroidotomy training in which a medic surgically opens an airway on the battlefield, and we are laying the hardware and

software foundations to support nasopharyngeal and needle chest decompression training. As these components are added, we plan to continue the evaluation research efforts to determine training system effectiveness, identify usability issues, and identify useful training niches such as was accomplished through the stress research. We believe a blended training solution, providing opportunities for information, demonstration, practice and feedback, in an inexpensive training suite, will be one key to the effectiveness of the HapMed system.

#### 4.0 REFERENCES

1. Butler et al. (2008). Tactical combat casualty care for 2007. Section I: Evolving concepts and battlefield experience. *Journal of Special Operations Medicine. Training supplement to Volume 8 (4)*.
2. Fowler, S. Smith, B. & Litteral, D. (2005). A TC3 Gamebased Simulation for Combat Medic Training. *Proceedings of the Interservice/Industry Training, Simulation & Education Conference*. Orlando, FL.
3. Klenerman L. & Hulands G.H. (1979). Tourniquet pressure for the lower limb. *J Bone Joint Surg*, 61B:124.
4. Lazarus, T.; Martin, G.A.; Nayeem, R., Fowlkes, J. & Riddle, D. (2008). *E-MAT: The Extremities Multiple Application Trainer for Haptics-based Medical Training*. Paper presented at the Virtual Reality Conference, 2008.
5. Mazurek, M.T. & Burgess, A.R. (2006) Moderator's summary: Stabilization of long bones. *Journal of the American Academy of Orthopaedic Surgeons*, 14, S113-S117.
6. Parsons, D. (2004). Tourniquets: Lifesavers on the battlefield. *Journal of Special Operations Medicine*, 4(4), 51-53.
7. Richey, S.L. (2007). Tourniquets for the control of traumatic hemorrhage: a review of the literature. [World Journal of Emergency Surgery](http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2151059) 2. Retrieved 8/20/09 from <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2151059>
8. Riddle, D. L., Fowlkes, J., Lazarus, T., Martin, G., Daly, J., & Fowler, S. (2006). Integrating instructional strategies and haptic technologies to enhance the training efficacy of an Army 91W (Combat Medic) medical skills training simulation. *Proceedings of the Human Factors and Ergonomics Society 48<sup>th</sup> Annual Meeting [CDROM]*. San Francisco, California: HFES.

#### ACKNOWLEDGEMENTS

The authors would like to express our thanks to Retired Lieutenant Colonel Donald L. Parsons, PA-C, with the Department of Combat Medic Training, Fort Sam Houston, TX, for so generously sharing his time and expertise, for his efforts to ensure the HapMed team receives quality domain knowledge to support 91W10 instruction.

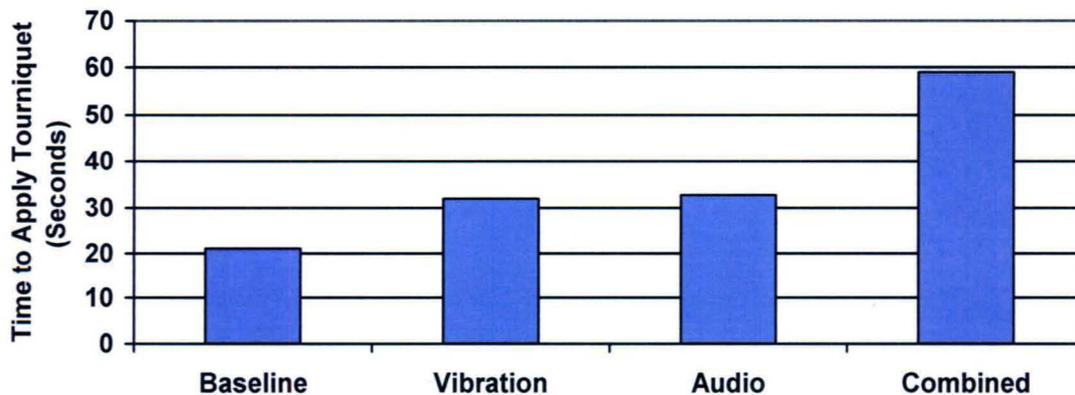


Figure 5: Results from the stress research.

# DigitalHuman (DH): An Integrative Mathematical Model of Human Physiology

Robert L. Hester, Ph.D.; Richard L. Summers, M.D.; Radu Ilescu, M.D. Ph.D.; Joyee Esters; and Thomas G. Coleman, Ph.D.

*Department of Physiology & Biophysics and Department of Emergency Medicine. University of Mississippi Medical Center. 2500 North State Street, Jackson, MS. 39216  
rhester@physiology.umsmed.edu*

**Abstract.** Mathematical models and simulation are important tools in discovering the key causal relationships governing physiological processes and improving medical intervention when physiological complexity is a central issue. We have developed a model of integrative human physiology called DigitalHuman (DH) consisting of ~5000 variables modeling human physiology describing cardiovascular, renal, respiratory, endocrine, neural and metabolic physiology. Users can view time-dependent solutions and interactively introduce perturbations by altering numerical parameters to investigate new hypotheses. The variables, parameters and quantitative relationships as well as all other model details are described in XML text files. All aspects of the model, including the mathematical equations describing the physiological processes are written in XML open source, text-readable files. Model structure is based upon empirical data of physiological responses documented within the peer-reviewed literature. The model can be used to understand proposed physiological mechanisms and physiological interactions that may not be otherwise intuitively evident. Some of the current uses of this model include the analyses of renal control of blood pressure, the central role of the liver in creating and maintaining insulin resistance, and the mechanisms causing orthostatic hypotension in astronauts. Additionally the open source aspect of the modeling environment allows any investigator to add detailed descriptions of human physiology to test new concepts. The model accurately predicts both qualitative and more importantly quantitative changes in clinically and experimentally observed responses. DigitalHuman provides scientists a modeling environment to understand the complex interactions of integrative physiology. This research was supported by NIH HL 51971, NSF EPSCoR, and NASA

## 1.0 INTRODUCTION

Mathematical simulations of physiological processes have become an important tool in understanding normal and pathophysiological processes within the body. Beard and colleagues have presented a very detailed simulation of cardiac metabolism, in particular energy metabolism during cardiac ischemia (2). Secomb and colleagues have simulated microcirculatory hemodynamics in vascular networks (10). Oxygen delivery to tissue has been extensively modeled, (9) along with simulations of VEGF release demonstrating that hypoxia-induced VEGF release is important to direct angiogenesis towards hypoxic tissue (7). There are an extensive number of publications in the literature describing mathematical simulations of individual organ systems, but there are no comprehensive models demonstrating the integration across different organ systems. Olufsen et al. has described integrative models of the baroreflex/sympathetic nerve system interaction demonstrating blood

pressure regulation during orthostasis (8), but these are also somewhat limited in the integration of the entire body responses. In this paper we describe a detailed integrative model of human physiology, designed such that the physiological descriptions can easily be changed by the user.

Starting with the Guyton cardiovascular model in the late 1960's and continuing to the present, Guyton and Coleman demonstrated the use of computer simulations for education purposes and to develop and test hypotheses concerning physiological systems (1; 5; 6). Perhaps one of the best known historical models of integrative physiology is the Coleman HUMAN model which contained a detailed description of circulatory function (4).

There are currently three available integrative models of the human body. All are based on the original model "Human" written by Drs. James Randall and Thomas Coleman between 1981 and 1987. The 1980 version of Human is

available at Skidmore College and is used as a web-based teaching tool. QCP2005.exe (Quantitative Circulatory Physiology) developed at the University of Mississippi Medical Center (UMC) incorporates the cardiovascular, renal, respiratory, endocrine, and nervous systems. A major limitation of this model is that the program is written in C++ and is compiled. Parameter values can be changed using slider bars and other active screen objects, but it is not possible to change or add underlying equations. Therefore we have developed DigitalHuman (DH), a simulation of human physiology where the underlying physiological relationships are written in XML. The model is completely specified in XML (which is both machine and human readable). This paper describes the basics of the current version of DH.

## 2.0 METHODS

The DigitalHuman simulation package is comprised of a series of files that describe the physiology (Structure files), a set of files that describe the display characteristics of the simulations (Display), and the executable DigitalHuman.exe file. To solve and display the XML based physiological descriptions the executable DigitalHuman.exe file is a compiled (C++) code which consists of a fast XML parser, numerical methods used in solving algebraic and differential equations, and the code that generates screen updates. No unusual libraries are used. The code is currently compiled for Windows using one of several Microsoft C++ compilers.

**2.1 XML Structure Files** For DH the XML files describing the physiological responses are located in a folder called "Structure". Within the "Structure" folder are a series of subfolders, each describing a particular organ or responses. Within these sub folders are simple text files having a .DES extension that contain the physiological descriptions. This organization is designed for ease of use in finding particular files. The individual XML files are simple text files and modifications can be made by anyone using a text editor.

The physiological variables and relationships are described using Extensible Markup Language (or XML). The extensible in XML's name identifies one of its major strengths as XML can be customized to meet specific needs. In the case of mathematical model documentation, we have developed an XML schema that is used to represent the details of mathematical models,

including the structure of the model, the control of solutions and the display of results.

The XML data files have several important elements. With respect to the XML schema used by DH, these are:

**<variables>** which is used to declare (and sometimes define) variables. There are a variety of variable types including ordinary, parameters, constants, timer variables and random variables.

**<equations>** Declare and parameterize differential and implicit algebraic equations.

**<functions>** Define curvilinear functions.

**<definitions>** Create the blocks of math that calculate the derivative values and do additional ancillary math.

One element that is important is the XML element **<curve>**. Sometimes, the precise underlying physiological description is not known, but can be described as a curvilinear function. To accomplish this we have developed the element **<curve>** which is described by the following text demonstrating the relationship between PO<sub>2</sub> and erythropoietin section.

```
<curve>
  <name>PO2Effect</name>
  <point><x> 0.0 </x><y> 4.0
  </y><slope> 0 </slope></point>
  <point><x> 35.0 </x><y> 0.0
  </y><slope> -0.14
  </slope></point>
  <point><x> 60.0 </x><y> -1.0
  </y><slope> 0 </slope></point>
</curve>
```

The parser reads the text and fits the data to a cubic spline. Since most data sets are small we have determined that the model builder must specify the slope of the curve at each data point to provide a better fit. That is the approach used in DH and the curve for the above equation is shown in Figure 1. The benefit of this method of describing a curve is that the researcher does not have to do any mathematical analysis.

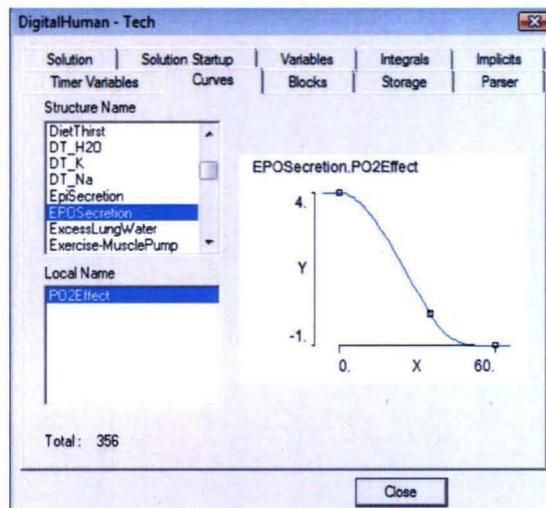


Figure 1: Curve fitting capability of DH

The examples of the code and Figure 1 describing erythropoietin secretion demonstrate the ease of writing code to describe physiological responses. The "curve" function allows the investigator to describe a physiological relationship with an arbitrary number of points (the minimum is two), along with the slope at each point. The numerical methods then fit a cubic spline curve to the data and use this fitted equation in DH. The investigator can view the curve and determine if the curve accurately demonstrates the physiological responses.

## 2.2 Parser and Solver

DH parses mathematical expressions directly and evaluates them in its own math engine. There are some important rules for writing the XML code. A common source of error is the failure to comply with these rules, however the parser will inform the user of any rules that are broken so the code can be corrected. The XML parser makes the following tests:

1. Is the document well-formed? This means does it conform to the general rules of XML. For example, is each opening tag paired further down with a corresponding closing tag?
2. Is the document valid? Does it conform to its XML defined document structure? Are all the element names recognized?
3. Is the data acceptable? Is text interpretable? Is the literal a valid number? Is the number within range?

DH uses a serial access (SAX) parser that completes all three tests in one pass. The parser stops if an error is detected and the exact spot in

the document is logged and reported when the error is detected.

Parsing of DH requires <4 seconds on Dell desktop (2.8 GHz, 3 GB, XP). Subsequent parses require less time since Windows memory maps recently read files. The quick re-parse response is convenient during model development when lots of parses are needed.

## 3.0 RESULTS

### 3.1 Physiological responses

DH allows the user to adjust many characteristics of the patient's physical environment, from global conditions such as altitude to local qualifiers such as temperature, humidity, and barometric pressure. Other external parameters that can be adjusted include partial pressures of the individual gases in inspired air and nutritional composition and amount of ingested food and fluids. Control of the patient's daily routine schedule allows the user to adjust basic functions such as sleeping, working, and feeding on an hour-to-hour basis, whereas the Exercise Panel facilitates studying the effects of differing types of exercise on both a short and long-term scale.

Various panels such as the Organ Details and Basic Physiology button groups allow the user to investigate and adjust physiological parameters on a more in-depth basis. Panel features include graphical data displays, information buttons, and adjustable variables. In some instances, pathophysiological states can be mimicked through the use of radio and slider buttons that allow hormone levels to be clamped or fixed at a given level.

In addition to manipulating physiological parameters, DH allows the user to administer pharmacological agents to treat a simulated patient. Currently, several drugs are available for interventional treatment: chlorothiazide, digoxin, furosemide, midodrine, insulin and epinephrine.. Additional drugs can be added as the quantitatively relationships regarding their effect upon physiological functions are determined. Several additional treatment options are available in the DH model, including placing the patient on a ventilator, administering fluids via an IV drip, and performing a blood transfusion.

The "Chart" panel provides graphical description of blood pressure, body temperature, heart rate, and ventilatory rate. The full drop-down menu provides additional windows with organ and

physiological responses. Panel features include graphical data displays, information buttons, and adjustable variables. Pathophysiological states can be mimicked through the use of radio buttons and slider bars that allow values to be changed or fixed at a given level.

Currently DH has ~5000 variables describing a variety of physiological responses. The following demonstrates the physiological response to exercise, both at a global and at a tissue level. To run a simulation, DH is started and there is approximately a 5-10 second period (depending on the speed of the computer) while the program parses all of the XML code. The simulation is initiated once the user activates the dropdown menu under "Go", and "brings the person to life" for a period of time, from 1 second to 1 month, depending on the desire to observe acute or chronic physiological changes.

Figure 2 demonstrates the cardiovascular responses upon standing, followed by a period of exercise. The simulation is run for 10 minutes while the subject is lying down, followed by another 10 minutes of standing, followed by 20 minutes of exercise. Each 10 minutes of simulation take approximately 2 seconds of computation time. Note that the increase in heart rate (Figure 2 upon standing). For exercise we can make the subject exercise for 20 minutes on an exercise bicycle at an level of 200 watts. Note the rapid increase in blood pressure, heart rate, and respiratory rate. In Figures 3-5 we provide examples of the detail of DH, demonstrating acid-base balance, liver metabolism, and neural activity.

DH saves the values of all variables along with the state of the timer variables. Saving complete solutions is also supported. A solution can subsequently be reloaded for viewing and can even be continued. File sizes may be large, with the file size for the experiment in Figure 2 requiring ~ 10 MB of disk space.

#### Gender Specific DH model

We have added detail and scaling to DH to simulate certain aspects of female physiology. In DH the parsing of the XML document is conditional depending on the criteria set in a simple text file.

Figure 6 shows the initial screen for the DH female version (note the female morphology). Figure 7 shows the simulation results for ovarian secretion of estradiol, under cyclic LH control for

one month. We are working to add additional content to DH to fully simulate female physiology.

#### 4.0 CONCLUSION

We have developed an integrative model of human physiology. The physiological equations, variables, parameters and quantitative relationships as well as all other model details are described in XML text files. The advantage of using XML is that it is a universally used format; it can be rapidly parsed by computers but also be read and edited by humans as a text-based document. Additional detail can be added by simply writing a XML file with the appropriate description and including it into the structure folder. The use of scaling allows DigitalHuman to simulate male and female physiological responses and many other traits. DH provides a rich environment for understanding human physiology and provides researchers with an environment to easily make changes in the model.

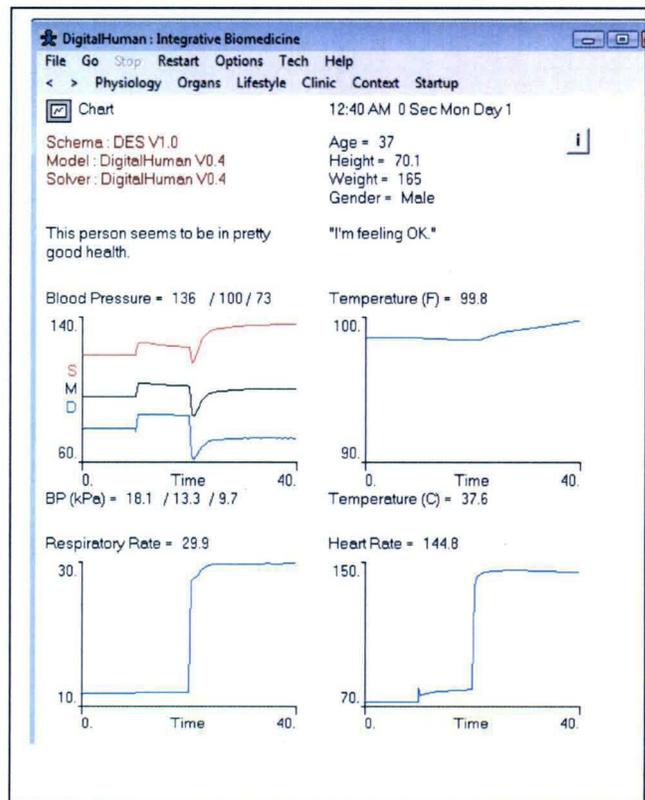


Figure 2: Physiological responses during 20 minutes of exercise

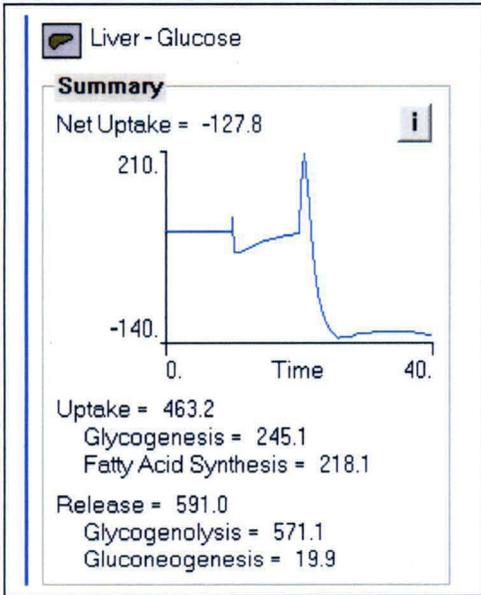


Figure 3: Liver glucose metabolism during exercise

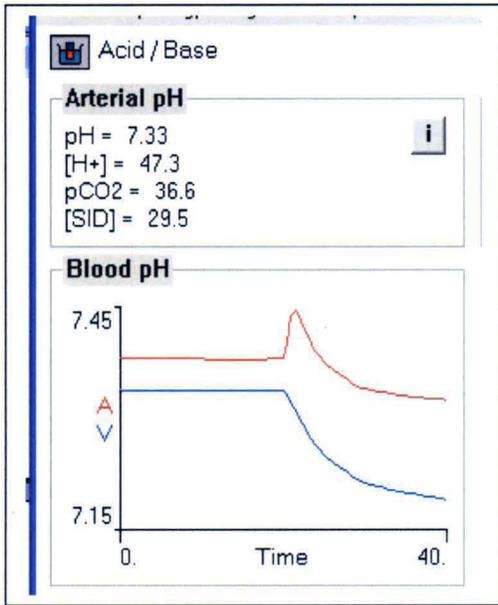


Figure 4: Blood pH changes during exercise

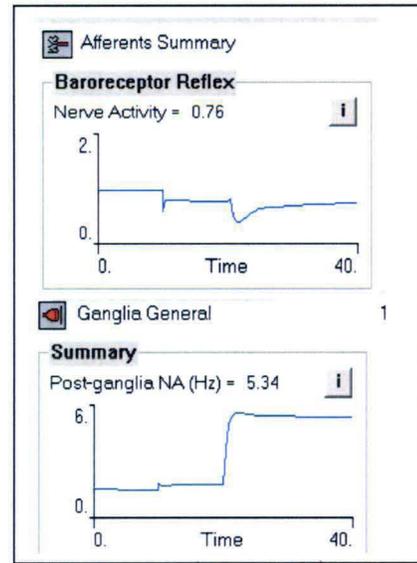


Figure 5: Neural activity during exercise

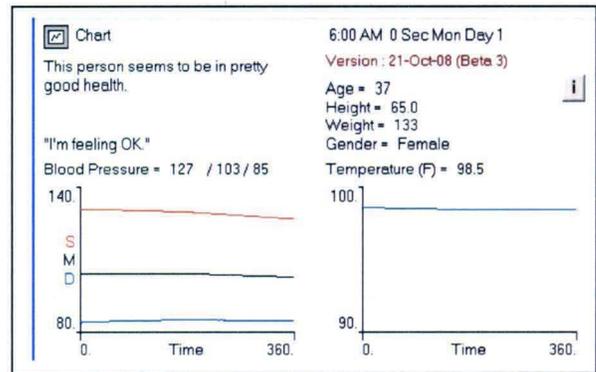


Figure 6: Female version of DH

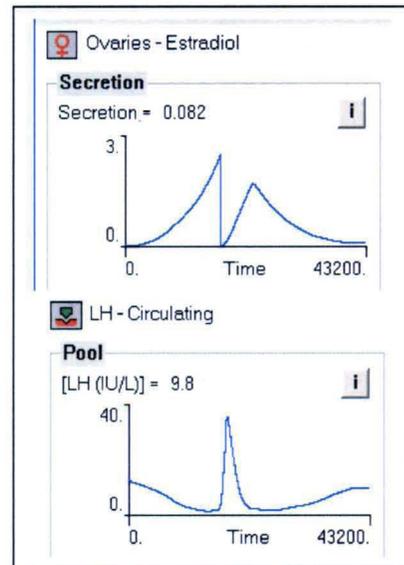


Figure 7: Estradiol responses in the female versions of DH

Free downloads of the software are available at [groups.google.com/group/modelingworkshop](http://groups.google.com/group/modelingworkshop)

Work in this manuscript was supported by National Institutes of Health grant HL-51971, a NSF EPSCOR grant and NASA.

## 5.0 REFERENCES

1. Abram SR; Hodnett BL; Summers, RL; Coleman, TG; & Hester RL. (2007) "Quantitative Circulatory Physiology: an integrative mathematical model of human" physiology for medical education. *Advances in Physiology Education* 31: 202-210,
2. Beard DA. (2006) "Modeling of oxygen transport and cellular energetics explains observations on in vivo cardiac energy metabolism," *PLoS Comput Biol* 2: e107.
3. Coleman TG & Randall JE. (1983) "HUMAN. A comprehensive physiological model," *Physiologist* 26: 15-21.
4. Guyton AC & Coleman TG. (1969) "Quantitative analysis of the pathophysiology of hypertension," *Circulation Research* 24: 1-19.
5. Guyton AC; Coleman TG & Granger HJ. (1972) "Circulation: overall regulation," *Annual Review Physiology* 34: 13-46.
6. Ji JW; Mac GF & Popel AS. (2007) "Skeletal muscle VEGF gradients in peripheral arterial disease: simulations of rest and exercise," *American Journal Physiology* 293: H3740-H3749
7. Olufsen MS; Alston AV; Tran HT & Ottesen JT and Novak V. (2008) "Modeling Heart Rate Regulation-Part I: Sit-to-stand Versus Head-up Tilt," *Cardiovascular Engineering*.8: 73-87
8. Roy, TK & Popel, AS. (1996) "Theoretical predictions of end-capillary PO<sub>2</sub> in muscles of athletic and nonathletic animals at VO<sub>2</sub>max," *American Journal Physiology* 271: H721-H77.
9. Secomb TW; Styp-Rekowska B & Pries AR. (2007) "Two-dimensional simulation of red blood cell deformation and lateral migration in microvessels," *Annals Biomedical Engineering* 35: 755-765.

# Uterine Contraction Modeling and Simulation

Miao Liu, Lee A. Belfore, Yuzhong Shen, and Mark W. Scerbo

*Virginia Modeling, Analysis, and Simulation Center  
Old Dominion University, Norfolk, VA, USA, 23529  
{mliux003, lbelfore, yshen, mscerbo}@odu.edu*

**Abstract.** Building a training system for medical personnel to properly interpret fetal heart rate tracing requires developing accurate models that can relate various signal patterns to certain pathologies. In addition to modeling the fetal heart rate signal itself, the change of uterine pressure that bears strong relation to fetal heart rate and provides indications of maternal and fetal status should also be considered. In this work, we have developed a group of parametric models to simulate uterine contractions during labor and delivery. Through analysis of real patient records, we propose to model uterine contraction signals by three major components: regular contractions, impulsive noise caused by fetal movements, and low amplitude noise invoked by maternal breathing and measuring apparatus. The regular contractions are modeled by an asymmetric generalized Gaussian function and least squares estimation is used to compute the parameter values of the asymmetric generalized Gaussian function based on uterine contractions of real patients. Regular contractions are detected based on thresholding and derivative analysis of uterine contractions. Impulsive noise caused by fetal movements and low amplitude noise by maternal breathing and measuring apparatus are modeled by rational polynomial functions and Perlin noise, respectively. Experiment results show the synthesized uterine contractions can mimic the real uterine contractions realistically, demonstrating the effectiveness of the proposed algorithm.

## 1. Introduction

Uterine contractions are stimulated by uterine muscle cells. Uterine contraction (UC) variations reflect the physiological changes of the uterus during both pregnancy and labor [1]. As a critical component in fetal heart rate (FHR) monitoring during labor and delivery, uterine contractions provide important information regarding maternal and fetal wellbeing. There are three main methods to record uterine contractions [2]: tocography, electrohysterography, and using intrauterine pressure catheter. Tocography measures the strain exerted by uterus on the maternal abdomen via external a tocotransducer. Electrohysterography records the electrical uterine activities from the maternal abdomen. The intrauterine pressure catheter (IUPC) measures intrauterine pressure invasively, and is mostly used during labor. Regardless of their differences, all three methods aim at providing records of the contraction patterns and their relationship to FHR. Proper interpretations of fetal heart rate and uterine contractions require special training, while monitoring of both are only available when pregnant women are hospitalized in parturiency. Building a training system that can simulate fetal heart rate and uterine contractions can help medical personnel learn critical patterns of both signals without putting patient in danger. To gain better understanding of uterine activities, it is necessary to develop mathematical models to quantitatively describe various uterine contraction patterns and this is the problem to be addressed in this paper.

Even though FHR monitoring is now the standard

practice during labor and delivery, surprisingly, there were only a few studies on uterine contraction modeling and simulation. Young used polynomials to model five characteristics of uterine contractions: 1) gradual onset, 2) a linear rising segment, 3) a plateau region, 4) a symmetrical fall, and 5) gradual offset, and fitted the simulated contractions with recorded IUPC data [3]. Their results matched their physical analysis. However, their simulations were not perfect especially in the tail region of the contraction curve. Vauge et al. [4] developed a system of differential equations that describe the dynamics of uterine pressure during human parturition. This method was based on three simplified assumptions: identical contractile properties of all myometrial cells, intrauterine pressure proportional to the number of contracted myometrial cells, and that all cells have three states, namely, contraction, recovery, and resting. Their model was simple and effective especially for normal contractions which begin in the fundus, reach the apex, and then proceed symmetrically downward toward the fundus. But this method did not consider the fact that asymmetry can occur when the uterine cells function independently causing ineffective uterine contractions and minimal dilatation [2]. Recently, Kemal et al. [5] employed two methods to simulate uterine contractions. The first one is based on the same mathematical model proposed by Vauge [4]. The second approach is based on recorded patient data. They first applied Hilbert-Huang Transform (HHT) [6] to identify the contraction locations from real patient data, and then developed spatial-temporal simulations of uterine contractions. All these methods discussed above are capable of illustrating the dynamics of uterine activities. However, these methodologies are

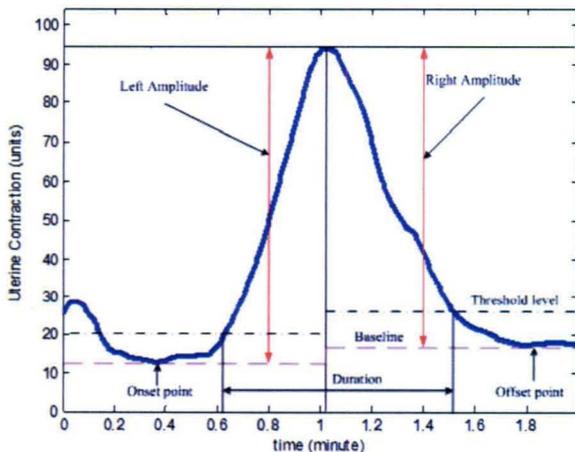
deterministic and no comprehensive parameter estimation for these models was developed. Furthermore, the noise caused by fetal movements and maternal breathing were not considered and modeled. To address the problems in existing uterine modeling and simulation, the paper proposes a novel algorithm that integrates three major components: asymmetric generalized Gaussian function (AGGF) for modeling contractions, Perlin noise for modeling maternal breathing and instrument noise, and impulsive noise for modeling fetal movements. The parameters of the asymmetric generalized Gaussian function are estimated using the least square method based on detected uterine contractions from real patient records.

The remainder of this paper is structured as follows. Section 2 first introduces the proposed asymmetric generalized Gaussian functions for modeling uterine contractions and then estimate the parameters of generalized Gaussian functions based detected contractions. Section 3 describes Perlin noise generation for maternal breathing and low amplitude noise modeling. Section 4 presents impulsive noise generation for simulating disturbances caused by fetal movements. Section 5 summarizes the simulation procedure and compares the simulation results. Section 6 concludes this paper and discusses future research directions.

## 2. Uterine Contraction Modeling

### 2.1 Asymmetric Generalized Gaussian Function

A typical uterine contraction curve of a real patient is depicted as a continuous waveform in Figure 1. This curve is characterized by a basal tone varying from 0 to 20 units, and a deflection of the contraction curve above the baseline, whose amplitude and duration are within a certain range of values [2]. Moreover, one should note this curve is asymmetric.



**Figure 1:** Typical uterine contractions. Also shown are parameters used for contraction detection.

Since the asymmetry of contraction curve matches

real cases of ineffective contractions and comprises the symmetrical case, we propose to use an asymmetric generalized Gaussian function to model uterine contractions as follows.

$$f(t) = A_l \exp\left\{-\frac{(t-t_0)^{\alpha_l}}{\beta_l}\right\} [u(t-t_l) - u(t-t_0)] + A_r \exp\left\{-\frac{(t-t_0)^{\alpha_r}}{\beta_r}\right\} [u(t-t_0) - u(t-t_r)] + b(t) [u(t-t_l) - u(t-t_r)], \quad (1)$$

where the parameters of (1) are shown below.

$A_l, A_r$	Amplitudes for the left and right sides
$\alpha_l, \alpha_r$	Exponents for the left and right sides
$\beta_l, \beta_r$	Variances for the left and right sides
$u(t)$	A unit step function
$t_l, t_r$	Left and right cut off time
$t_0$	The position where $f(t)$ reaches its maximum
$b(t)$	Baseline representing some basal strain exerted by the uterine muscle when contractions do not occur

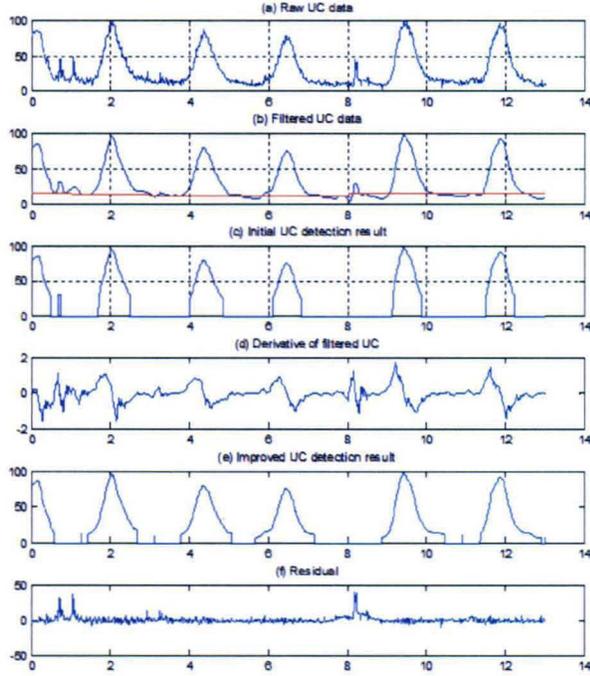
To simulate uterine contractions, we need to determine the range within which the above parameters lie and how they vary with time. So in the next step we will detect uterine contractions from real patient data and estimate the parameters for the asymmetric generalized Gaussian function from the detection results.

### 2.2 Uterine Contraction Detection

Several methods were proposed to detect uterine contractions for different purposes. Radhakrishnan et al. [7] developed a higher-order zero crossing based method and studied the frequency of occurrence of contractions in different pregnancy stages. Novak et al. [8] described two UC detection approaches: amplitude- and derivative-based algorithm. By comparing the results from these two methods, they suggested combining both methods together to achieve better detection results. Aiming at quantitatively analyzing uterine contractions in time domain, Jezewski et al. [9] introduced a statistical method to determine the threshold and also considered duration condition for UC detection. These methods were used to calculate the regular parameters of uterine contractions such as amplitude, duration, frequency of occurrence. Our method is based on a combination of the last two methods, in which thresholding is first performed to detect the presence of uterine contractions, and derivative-based method is applied subsequently to include the samples whose amplitudes are below the threshold, but still belonging to the contractions.

The algorithm proposed by Jezewski et al. [9] starts with low pass filtering with cutoff frequency of 0.04 Hz

to suppress the artifacts caused by fetal movements and maternal breathing. Then the record is analyzed by using a moving window with a length of 4 minutes and 1-minute step. Within each window, the histogram of uterine pressure samples is constructed first and the mode of the histogram is then selected as the baseline value. Finally, the threshold level is set as 10 units above the baseline and the validity of data segment is examined. A valid contraction should remain above the threshold level for a duration longer than 30 seconds and the amplitude of contraction exceeds 20 units.



**Figure 2:** Intermediate results in uterine contraction detection. (a) Raw UC data. (b) Filtered UC. (c) UC detection by thresholding. (d) The derivatives of filtered UC. (e) Improved UC detection results. (f) The residual between raw UC and filtered UC which is detected.

After the uterine contractions are detected, there are still two remaining problems. The first is that the amplitude of detected samples are mostly above the threshold, thus the estimated parameters may not be accurate enough to model the tail areas. We need to include more samples from tail areas. On the other hand, the tail areas are prone to other various sources of noise. Before including them for parameter estimation, we must differentiate contaminated samples and uncontaminated or less-contaminated samples. Since the derivatives in noisy area change drastically, we propose to employ derivative-based detection method to differentiate them. In other words, for less contaminated contractions, the derivative of the left side of the contraction should be positive, while the derivative of right side should be

negative. Starting from the peak of the contraction, we proceed to its left and compute the derivatives of the smoothed uterine pressure signal and then search the position where the derivative changed to negative for left side curve, and denote it as the tail point. The sample points between the peak and the tail point will be utilized for curve parameters estimation. The same principle applies to the right side of contraction with positive derivatives being searched.

The results of uterine contraction detection are shown Figure 2, where (a) is the original uterine pressure signal; (b) is its filtered version, the red solid line represents the baseline; (c) illustrates the sample whose values are above the threshold; (d) is the derivatives of filtered UC data; (e) is the improved uterine contraction detection result; and (f) shows the residual between the detected contractions and original signal (a).

### 2.3 Uterine Contraction Parameter Estimation

After the uterine contractions are detected, we need to compute the parameters of the asymmetric generalized Gaussian function (1) in order to simulate the detected contractions. Here we only consider the left half of the asymmetric generalized Gaussian function (1), that is,

$$f(t) = A_l \exp\left\{-\frac{|(t-t_0)|^{\alpha_l}}{\beta_l}\right\} + b_l [u(t-t_l) - u(t-t_0)] \quad (2)$$

for the parameter estimation, while the right half can be handled similarly and thus is omitted in this paper.

First, the baseline  $b_l$  is estimated as the contraction value at the onset point  $t_l$  which is the lowest point between two filtered contractions. Thus

$$A_l = f(t_0) - b_l.$$

After simple manipulation, the remaining known variables and unknown parameters are separated by taking the logarithm of both sides of (2),

$$\ln \frac{f(t) - b_l}{A_l} = -\frac{|(t-t_0)|^{\alpha_l}}{\beta_l}. \quad (3)$$

Since  $f(t) - b_l < A_l$ , adding minus sign to both sides of (3) and taking logarithm again gives

$$\ln\left(-\ln \frac{f(t) - b_l}{A_l}\right) = \alpha_l \ln|t-t_0| - \ln \beta_l. \quad (4)$$

Now we denote the detected uterine contraction data set as  $\{f(t_i), t_i\}$ ,  $1 \leq i \leq N$ , where  $N$  is the number of samples. Substituting the data set into equation (4) and writing each term in matrix form, we

have

$$\mathbf{Y} = \begin{bmatrix} \ln\left(-\ln\frac{f(t_1)-b_l}{A_l}\right) \\ \ln\left(-\ln\frac{f(t_2)-b_l}{A_l}\right) \\ \vdots \\ \ln\left(-\ln\frac{f(t_N)-b_l}{A_l}\right) \end{bmatrix}, \mathbf{X} = \begin{bmatrix} \ln|t_1-t_0|, & -1 \\ \ln|t_2-t_0|, & -1 \\ \vdots & \vdots \\ \ln|t_N-t_0|, & -1 \end{bmatrix},$$

$$\mathbf{p} = \begin{bmatrix} \alpha_l \\ \ln\beta_l \end{bmatrix}. \quad (5)$$

Then equation (4) can be transformed into a concise linear equation represented by a matrix product,

$$\mathbf{Y} = \mathbf{X}\mathbf{p}, \quad (6)$$

where  $\mathbf{X}$  and  $\mathbf{Y}$  are matrices containing detected sample information,  $\mathbf{p}$  is the parameter vector to be estimated. The parameter estimation problem can be expressed as the following minimization problem,

$$\mathbf{p}_o = \arg \min_{\mathbf{p} \in \mathbb{R}^2} \|\mathbf{Y} - \mathbf{X}\mathbf{p}\|_2^2, \quad (7)$$

where  $\mathbf{p}_o$  is the solution. This is a least square estimation problem whose solution is

$$\mathbf{p}_o = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}. \quad (8)$$

Thus we have obtained the parameter estimator for the proposed asymmetric generalized Gaussian function for modeling uterine contractions.

### 3. Perlin Noise

The low amplitude noise caused by maternal breathing and measuring apparatus are random yet exhibiting both low and high frequency characteristics as shown in Figures 2(a) and (f). Common random number generators cannot be used directly to generate such noise, since they are too random to exhibit the natural outlook of continuity and self-similarity of the noise. To address this problem, we propose to use Perlin noise generator [10] to simulate the low amplitude noise.

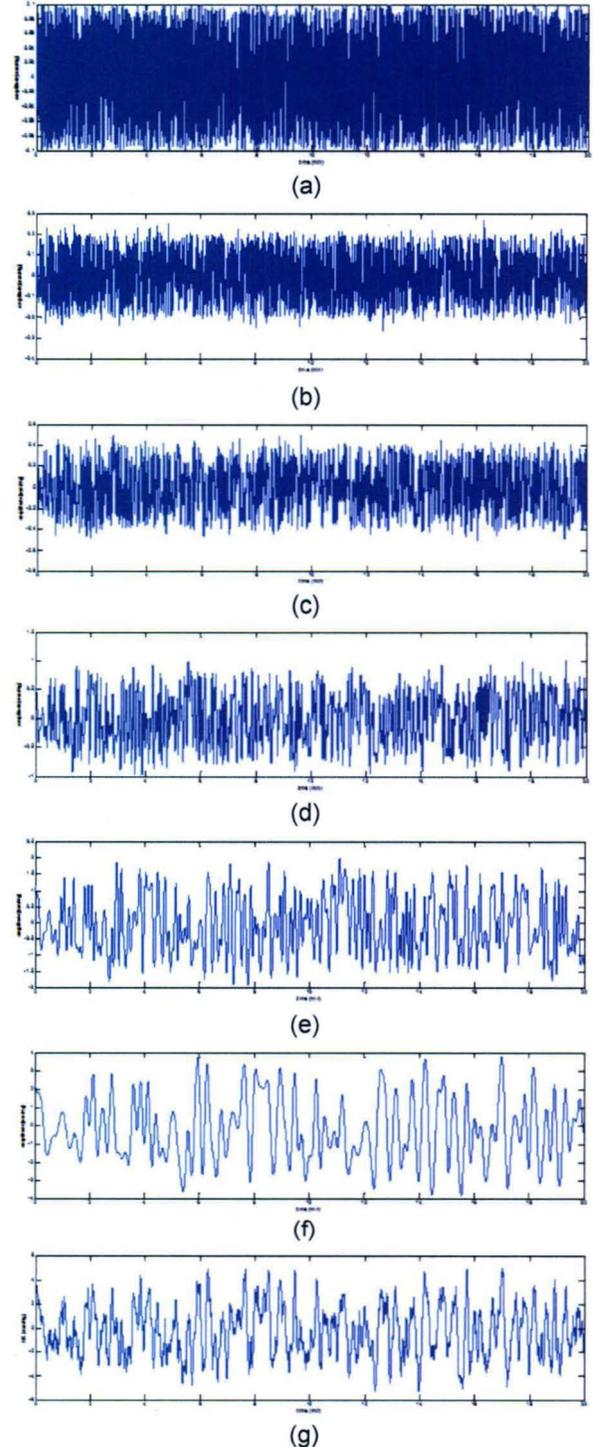
A Perlin noise generator is composed of two components: a noise function and an interpolation function. The basic procedure of Perlin noise generation is

- (1) Generate a series of random numbers  $n_l$  of length  $N$  from uniform distribution  $U[-0.1, 0.1]$ ;
- (2) Decimate the series to size  $N/2$ ;
- (3) Upsample the decimated series to size  $N$  by B-spline interpolation and increase the amplitude of the new series by a factor of  $p$ ;
- (4) Repeat (2) and (3) until reaching the specified level  $s$ , and we obtain a set of series  $\{n_1, n_2, \dots, n_s\}$ .

By adding the above generated series together, we have the Perlin noise

$$f_p(t) = \sum_{i=0}^s p^i n_i(t), \quad (9)$$

where  $t \in [1, N]$ , and  $p$  is called persistence factor which controls the amplitude.



**Figure 3:** (a) - (f) are waveforms corresponding to 6 levels noise series  $\{n_1, n_2, \dots, n_6\}$ . (g) is the Perlin noise.

One fact worth of explaining is the mechanism for how the continuous noise is produced. Since  $n_1$  is generated from i.i.d. (independent and identically-distributed) uniform distribution, it is actually white noise. Through repeated downsampling and upsampling, the new noise series become the low-pass filtered version of the previous noise series. In other words, decimation by a factor of 2 reduces one half the Nyquist frequency of previous noise, then B-spline interpolation restores the sample number of noise without incurring new frequency contents. Thus the noise  $n_s$  generated in the last step, occupies the lowest frequency band. So the waveforms of the noise series from  $n_1$  to  $n_s$ , become increasingly smoother. Figure 3 illustrates the components of the Perlin noise generated in this work, in which (a) is the noise generated from uniform distribution  $U[-0.1, 0.1]$ , (b) – (f) are low-pass filtered version of noise generated from one level before with  $p = 2$ . Note that the sub-figures in Figure 3 have different vertical scales and low-frequency components have much larger amplitudes than high-frequency components. The final synthesized Perlin noise is shown in Figure 3(g).

#### 4. Impulsive Noise Modeling

The last component to simulate is the spikes or large magnitude impulsive noise in uterine contractions, which suggest possible fetal movements [2]. After examining the shape of those spikes, we propose to use the following rational polynomial function to model them

$$f(t) = \frac{a}{1 + \left(\frac{t - t_0}{b}\right)^2}, \quad (8)$$

where  $a$  is the amplitude which follows uniform distribution with value between 0 and 50 units,  $b$  is the scale parameter following uniform distribution with value between 0 and 10, and  $t_0$  is the spike position. A typical simulation result of the impulsive noise is shown in Figure 4.

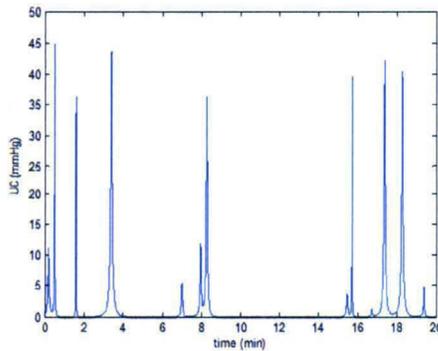


Figure 4: Impulsive noise simulation

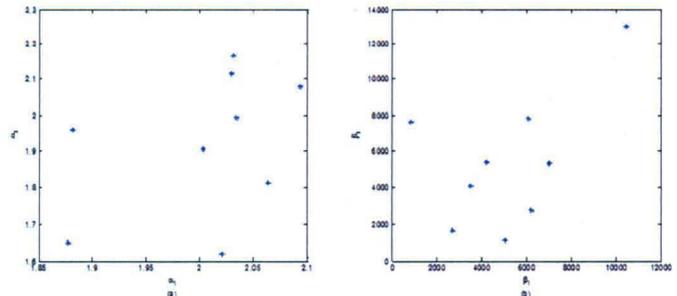
#### 5. Uterine Contraction Synthesis

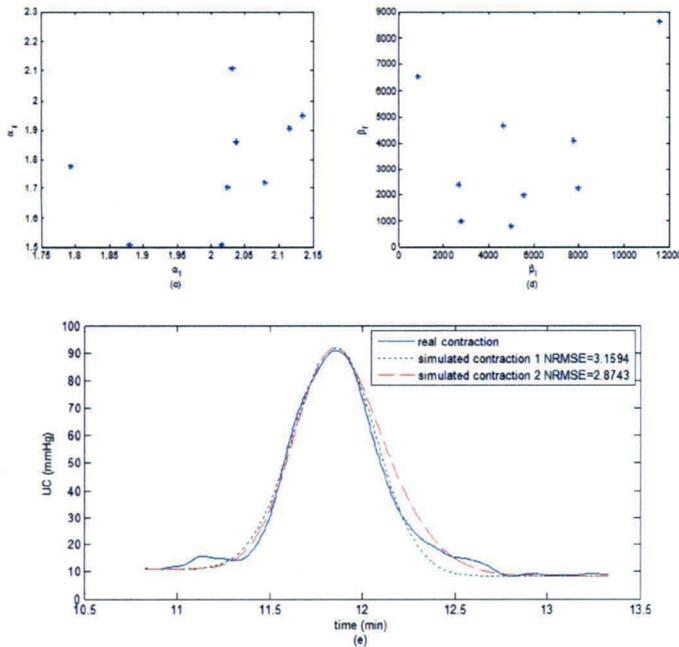
Uterine contraction simulation is finalized by superimposing the components generated by the asymmetric generalized Gaussian function, Perlin noise, and impulsive noise. First, a typical segment of 20 minutes was extracted from real patient record and the parameters for the asymmetric generalized Gaussian function were estimated. Figure 5 compares the parameters estimated from two detected data sets. The first data set contains only samples whose amplitude is above the threshold. The second data set is the expanded version of the first set by incorporating samples whose amplitude are below the threshold but are belong to the contraction. Figures 5 (a) and (b) plot parameters estimated from the first set. Figures 5(c) and (d) are estimated from the second set. It can be seen that the parameters of the left side and right side are different, validating the asymmetry of uterine contractions.

To compare the impact of different parameters on uterine contraction simulation, we select one set of parameters from Figures 5(a) and (b), and another set from Figures 5(c) and (d) to simulate two contractions, namely, contraction 1 (green dotted line) and contraction 2 (red dashed line), as shown in Figure 5(e). It can be seen that the red dash line achieves a better fit to the real contraction at both the peak and tail area of the contraction and has smaller normalized root mean square error (NRMSE), which is defined as

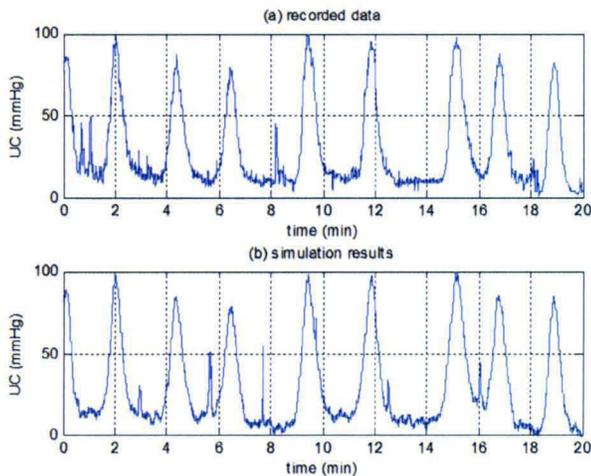
$$\text{NRMSE} = \frac{\sqrt{\frac{\sum_{i=1}^N (f_1(i) - f_0(i))^2}{N}}}{\max_{1 \leq j \leq N} \{f_0(j)\} - \min_{1 \leq k \leq N} \{f_0(k)\}}. \quad (10)$$

This result indicates that the derivative-based analysis could help to recruit more effective data for parameter estimation, and thus increase the estimation accuracy. Finally, by adding together the asymmetric Gaussian function, the Perlin noise, and the impulsive noise, we obtain the final simulation results in Figure 6(b). Comparing the original uterine contraction record in Figure 6(a) and the simulated results in Figure 6(b), it can be easily seen that the proposed algorithm is very effective and produces superb results.





**Figure 5:** Parameters estimation and uterine contraction simulation. (a) is the scatter plot of  $\alpha$ , and (b) is the scatter plot of  $\beta$  for left and right curves. (c) and (d) are  $\alpha$  and  $\beta$  plots of the second data set. (e) shows the simulations of one uterine contraction from two data sets.



**Figure 6:** Simulation results. (a) The original uterine contraction recording. (b) Simulated uterine contraction based on (a). It can be seen that the proposed algorithm is very effective, producing realistic simulations.

## 6. Conclusion

This paper proposed a set of parametric models to simulate uterine contractions. The proposed algorithm contains three major components: AGGF model for contractions, Perlin noise for maternal breathing and instrument noise, and rational polynomial functions for fetal movements.

Parameter estimation for the asymmetric generalized Gaussian function is derived based on real uterine contractions. The proposed algorithm is effective and produces realistic simulation results. Future work includes parameter estimation for the Perlin noise and impulsive noise in order to further improve the quality of uterine contraction simulations.

## Acknowledgments

This work is sponsored by Eastern Virginia Medical School (EVMS). The authors appreciate the data and support provided by EVMS.

## References

1. M Holger, M William, S George, G Robert (2003): The physiology of uterine contractions. *Clinics in Perinatology* 30:665.
2. M Michelle, H Gayle, R Patricia (2007): *Essentials of Fetal Monitoring*, Springer, New York.
3. RC Young (1997): A computer model of uterine contractions based on action potential propagation and intercellular calcium waves. *Obstetrics and Gynecology* 89:604.
4. C Vauge, B Carhonne, E Papiernik, F Ferre (2000): A Mathematical Model of Uterine Dynamics and Its Application to Human Parturition. *Acta Biotheoretica* 48:95.
5. A Kemal, H Tansel, B Coskun, L Kamran (2006): Computer Simulation of Uterine Contraction Dynamics. In: *Proceedings of the 5th WSEAS International Conference on Circuits, Systems, Electronics, Control & Signal Processing*, WSEAS, Dallas, Texas, 1790.
6. NE Huang, Z Shen, SR Long, MC Wu, HH Shin, Q Zheng, N-C Yen, CC Tung, HH Liu (1998): The Empirical Mode Decomposition and Hilbert Spectrum for Nonlinear and Non-stationary Time Series Analysis. In: *The Royal Society*, 903.
7. N Radhakrishnan, JD Wilson, C Lowery, H Eswaran, P Murphy (2000): A Noninvasive Method Utilizing Higher-Order Zero Crossing for Signal Analysis. *IEEE Eng. Med. Biol* 19:89.
8. D Novak, A Macek-Lebar, D Rudel, T Jarm (2007): Detection of contractions during labour using the uterine electromyogram. In: *IFMBE*, 148.
9. J Jezewski, K Horoba, A Matonia, J Wrobel (2005): Quantitative analysis of contraction patterns in electrical activity signal of pregnant uterus as alternative to mechanical approach. *Physiological Measurement* 26:753.
10. Perlin Noise, [http://freespace.virgin.net/hugo.elias/models/m\\_perlin.htm](http://freespace.virgin.net/hugo.elias/models/m_perlin.htm)

# Proposition: High-Fidelity Simulation for Remediation

Katherine Quartuccio, Patricia Lisk, and Karen Mittura  
Germanna Community College  
Locust Grove, Virginia

**Abstract:** Every school of nursing encounters students who are in need of remediation regarding the ability to develop critical thinking skills or develop competent fundamental nursing skills. In order to remediate identified students, schools may select criteria to target at-risk students that may include low course scores and/or clinical skills deficits. Nursing students who have these identified deficits are encouraged to seek remediation as early intervention that is integral to their continuing success. Learning opportunities that allow at-risk students to evaluate their personal learning styles, to assimilate nursing knowledge acquired in the didactic setting, and to accommodate nursing theory in various patient presentations have proven beneficial. Courses designed to develop essential nursing skills integrate a variety of teaching and learning strategies that foster the development of critical thinking. Learning experiences reinforce principles of patient safety, provide opportunities for student integration of nursing knowledge, and supply nursing students with constructive feedback allowing them to improve and manage their learning. Integrating high-fidelity simulation into remediation courses is a strategy that provides a safe controlled learning environment that simulates the clinical setting. In addition, high-fidelity simulation incorporates the use of multiple learning styles while allowing students to integrate didactic theory into practice. High-fidelity remediation scenarios provide at-risk students with the opportunity to practice critical thinking as a supplement to traditional teaching and learning strategies.

## 1.0 Definition

Remediation is defined as an act or process of remedying (Merriam-Webster, 2009). Inherent in all nursing programs regardless of the educational level is the need to identify and meet specific competency standards. The process of remediation facilitates the developmental needs of nursing students and provides a mechanism by which all students are provided the opportunity to meet the required competency standards. The goal of the competency standards is to provide safe and effective patient care. Remediation can be accomplished in a self-directed format, program format, or a combination of both. A program format using high-fidelity simulation is the focus of this paper. The outcomes of remediation can benefit the individual student, the nursing program, and the healthcare community. Establishing this process of remedying deficits early in the

development of the nursing role as a student will hopefully be reiterated in future roles as a new graduate, novice nurse and as a life-long learner developing expertise in the practice of nursing.

According to Beyea & Kobokovich (2004), nursing students as novices experience significant performance anxiety, are rule-governed, and are challenged to multi-task in a complex environment. Therefore, the use of simulation in the remediation process protects patients from errors and facilitates the bridge from textbook theories to real-life application. No matter how experienced a person, learning from errors is a powerful aspect of improving expertise.

## 2.0 The Method

Once nursing students with skills deficits have been identified and recommended for remediation, the facilitator must carefully

orchestrate the remediation session in order to address student needs, create session objectives, choose appropriate high-fidelity simulation scenarios, and outline student demonstration expectations of performance.

### **3.0 Student Identification and Needs Assessment**

To understand the nursing students' need for remediation, at risk students must be identified and a needs assessment conducted to provide direction for remediation development. Faculty members use several methods to identify at risk students in nursing programs. Both academic and non-academic factors may lead to unsuccessful performance. Non-academic factors are addressed on an individual student basis using resources available at the college and within the community. The academic factors are addressed by differing remediation activities available at the college with one using directed high-fidelity simulation.

Early identification of students who may be at-risk academically allows faculty more opportunities to strengthen students' knowledge and skills. Prior to admission, nursing students meet selected academic criteria. These criteria include meeting prescribed benchmarks on a standardized pre-admission test, meeting selected GPA standards, and successfully completing nursing pre-requisite coursework. The standardized pre-admission examination tests the students' abilities in four core knowledge areas including: English, Science, Mathematics, and Reading. Students admitted to the program score above the 45<sup>th</sup> percentile in these areas. Students who have completed all general education courses and prerequisite coursework with a "C" or better are considered for admission. Overall student GPA is considered with students acquiring the highest GPA being priority candidates in the admission process. These measures are taken to select students who are most

qualified to meet the academic challenges of nursing during the admission process.

Despite the diligence and effort expended during the admission process to select candidates who are most prepared, students may encounter difficulty either in the academic environment or within the clinical setting at any point within the nursing curriculum. These difficulties may manifest themselves through academic failure on tests or through unsatisfactory performance in the clinical or laboratory settings. Difficulties in these areas lead the students' faculty advisors to refer them for academic or clinical remediation. When a student has failed two theoretical tests in a nursing course (attaining a score of less than 80% or less), students are referred to appropriate remediation strategies available at the college. Some of these strategies may include individualized tutoring, assistance with test taking, and study skills, and help with the management of test anxiety. Other strategies may include referring students to additional laboratory sessions and experiences to include those integrating high-fidelity simulations.

Another academic indicator used by faculty to identify students at risk for academic or possible National Council Licensure Examination (NCLEX) difficulty is a standardized mid-curricular examination. At present, the test in use is the Health Education Systems, Incorporated (HESI) mid-curricular examination. This test is given to students prior to their final year of study in the program. Students who score less than 850 on this examination complete prescriptive remediation activities. Although open to all students, students scoring below 700 on this test are required to complete successfully a remediation course during the summer session to address identified weaknesses. This course uses a case study methodology augmented with high-fidelity simulation as described later in this paper.

In all cases, whether remediating a beginning or a more experienced student, the facilitator in charge of the high-fidelity learning experience must collaborate with the referring faculty member and discuss skills deemed substandard in the nursing student performance. These deficit areas may include collegial communication, nurse/patient communication, psychomotor skills, or cognitive skills such as plan of care development and critical thinking.

#### **4.0 Objectives**

According to Willford and Doyle (2006), "Simulation allows multiple learning objectives to be taught in a realistic clinical environment ..." (p. 604). Outlining specific objectives provides the high-fidelity simulation facilitator with directions for teaching to specific learning domains. These domains may include professional roles, cognitive application of nursing interventions, or ethical legal dilemma solutions.

#### **4.1 Scenarios/Teaching Strategies**

Effective remediation experiences require the facilitator to understand that the target population of students may be at various levels of learning. These different learning levels may include: knowledge acquisition level, comprehension level, accommodation level, or analysis level of learning. In addition, the facilitator must possess an understanding of experiential learning and adult learning theories that improve student knowledge acquisition and accommodation. Teaching strategies that use these principles of andragogy must focus on the principles of "student empowerment" (Billings and Halstead, 2009, p.207). Strategies for nursing students who are adult learners must foster a safe environment for learning, improve learner's confidence level, respond to learner's life transitions, and create positive attitudes toward learning.

Once the student's needs assessment is completed and objectives are outlined, choosing high-fidelity scenarios that focus on the identified weaknesses will include student role-playing. Selected scenarios will encourage critical thinking that is needed to improve the identified deficiency. As an example, high-fidelity remediation sessions that focus on communication weaknesses may be manipulated to include the nursing student as an emergency room nurse who must provide the telemetry unit nurse with a patient transfer report. In addition to the nursing student role-playing the emergency room nurse, the nursing student who is role-playing the telemetry nurse must be aware of the needed information to properly care for their patient and consider methods that supports full disclosure for collegial communication.

#### **4.2 Student Demonstrations**

According to Hawkins, Todd and Manz (2008), "Simulation is an excellent venue for students to learn experientially and provides opportunities for students to practice problem solving and psychomotor skills in a safe, controlled environment" (p.524).

Depending upon identified student deficits, high-fidelity scenarios are worked and manipulated to solicit specific nursing student competencies that include professional roles, psychomotor skills, cognitive processes, or communication skills. In addition to student demonstration of skills during the high-fidelity simulation experience, requiring written feedback after the scenario provides the nursing student with the opportunity to reflect upon the experience and self evaluate behaviors.

#### **5.0 Conclusions**

Identifying and assisting high-risk students in nursing programs remains a challenge for nursing faculty. Developing a systematic method for identification of at-risk students followed with strategies for remediation can improve student performance. Although

remediation of these students can occur in a variety of ways, one way to stimulate critical thinking and help students integrate theoretical knowledge into practice is with high-fidelity simulation. Practicing in a safe environment, students put into practice theoretical concepts, reinforce principles of patient safety, and enhance clinical reasoning and critical thinking skills. Implementing selected case studies with high-fidelity simulation as described is one way to prepare students for ongoing success in nursing and eventual NCLEX passage.

## **6.0 References**

Beyea, S.C., & Kobokovich, L.J. (2004). Human Patient Simulation: a teaching strategy. *AORN Journal*, (3)2, 153-159.

Billings, D.M., & Halstead, J.A. (2009). *Teaching in Nursing: A Guide for Faculty* (3rd Ed.). Philadelphia, PA: W.B. Saunders.

Hawkins, K., Todd, M., & Manz, J. (2008). A unique simulation teaching method. 47(11). Retrieved June 15, 2009, from Proquest database.

Merriam-Webster Online Dictionary. Retrieved July 29, 2009, from <http://www.merriam-webster.com/dictionary/remedy>.

Willford, A. & Doyle, T. (2006). Integrating simulation training into the nursing curriculum. *British Journal of Nursing*, 15(7). Retrieved June 13, 2009 from EBSCOhost database.

# Modeling Pilot Behavior For Assessing Integrated Alerting And Notification Systems On Flight Decks

Mathew Cover, Graduate Student; Thomas Schnell, Associate Professor

Operator Performance Laboratory, University of Iowa  
mcover@engineering.uiowa.edu; tschnell@engineering.uiowa.edu

**Abstract.** Numerous new flight deck configurations for caution, warning, and alerts can be conceived; yet testing them with human-in-the-loop experiments to evaluate each one would not be practical. New sensors, instruments, and displays are being put into cockpits every day and this is particularly true as we enter the dawn of the Next Generation Air Transportation System (NextGen). By modeling pilot behavior in a computer simulation, an unlimited number of unique caution, warning, and alert configurations can be evaluated 24/7 by a computer. These computer simulations can then identify the most promising candidate formats to further evaluate in higher fidelity, but more costly, Human-in-the-loop (HITL) simulations. Evaluations using batch simulations with human performance models saves time, money, and enables a broader consideration of possible caution, warning, and alerting configurations for future flight decks.

## 1. INTRODUCTION

The aviation safety (AvSafe) program at NASA, is tasked with assuring that safety of current and future aircraft participating in the National Airspace System is always being evaluated and improved upon [7]. The Integrated, Intelligent Flight Deck (IIFD) program within Aviation Safety, has sponsored a research project at the University of Iowa's Operator Performance Laboratory to mitigate high crew workload and increase situational awareness in the operational NextGen environment. Specifically, the research project seeks to resolve conflicts in caution, warning, and alerts (CWAs) that may be presented to pilots. The mechanism in which this is done will be via a software solution called the integrated alerting and notification (IAN) function. This work is conducted in conjunction with Ohio University and is supported with efforts from Boeing, Rockwell Collins and Delft University of Technology (TUD).

## 2. ARCHITECTURE OVERVIEW

The architecture of our model is best introduced as a closed-loop control system where the aircraft state and IAN function are fed into a human model (Figure 1). The human model then analyzes the stimuli provided by the aircraft displays and sensors and responds accordingly, outputting feedback into the flight model.

The model needs to be run hundreds to thousands of times to test variations of the IAN function and displays on the human model. Using batch Monte Carlo simulations with a human model permits us to test out a wider variety of simulated avionics conditions and operational scenarios than could ever be feasible with HITL testing.

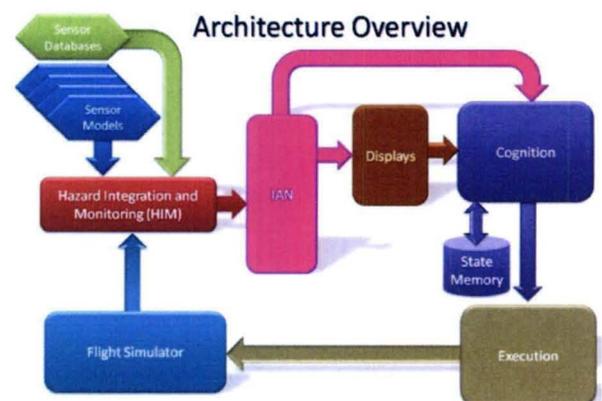


Figure 1: This figure outlines the top-level view of how the closed-loop model is constructed

### A. Simulink

For the project model, Simulink provides a structure that many pieces of the model can plug into to complete the closed-loop architecture. Using some TCP/IP functions, programs that run outside of Simulink are able to communicate and participate in the model such as a flight simulator,

the human model, and display visualizations (Figure 2).

### B. Flight Simulation Interface

The IAN project requires a realistic aircraft flight model for the human model to fly. Microsoft Flight Simulator X (FSX) has been chosen to run the flight model for this project. FSX provides *SimConnect*, a built-in interface that provides a standardized interface for add-on executable programs to communicate with and allows asynchronous communication over a network connection. These features make FSX ideal for multi-threaded applications and allow it to run out-of-process with the rest of the closed-loop model [6].

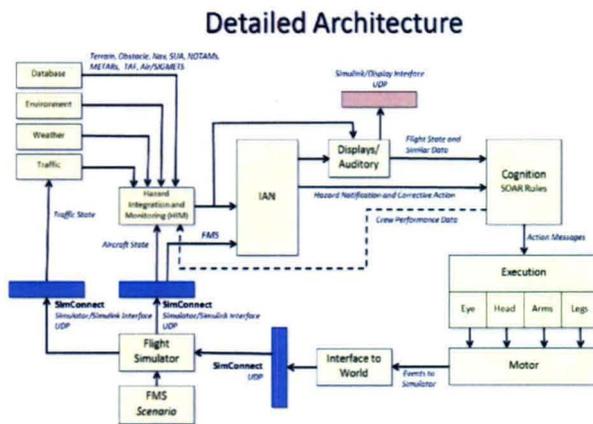


Figure 2: A more detailed view of the software architecture for the IAN / human model

### C. Aircraft Sensors

As part of the aircraft model, additional sensors that make the aircraft NextGen-worthy are being added to the FSX flight model. Among the sensors that are being modeled and incorporated into the architecture are GNSS, ADS-B, TCAS, TAWS/EGPWS, WxR. Future work by Ohio University will include computer models of sensors such as FLIR, millimeter wave radar (MMWR), 3D imager, and Interferometer. All of these sensors are coordinated and filtered by the Hazard Integration and Monitoring (HIM) module. Relevant data is then passed along to the IAN function.

### 3. INTEGRATED ALERT AND NOTIFICATION (IAN)

The flight simulator is only a part of the closed-loop model. The human model will connect to a representation of a NextGen cockpit. This simulation models an important characteristic of the NextGen aircraft – an integrated alerting and notification system. One aspect of this functionality is the ability to sort through alerts and notification and resolve conflicting information prior to presenting it to the pilot. This system is named the integrated alert and notification (IAN) function, as mentioned previously.

This is critical as conflicts in the cockpit consume pilot's time and spare mental capacity to resolve and respond in what are usually challenging situations. For instance, an enhanced ground proximity warning system (EGPWS) may suggest increasing altitude to avoid a mountain while the traffic collision avoidance system (TCAS) may tell a pilot to lower altitude to avoid another aircraft nearby.

Work with Boeing and Rockwell Collins will be able to help provide a set of rules in the IAN function so that visual, auditory, and tactile cues do not conflict and convey information in a useful manner to the pilot. They will also be constructed so as not to fall out of line with standards and common practices followed today and envisioned for NextGen operations.

### 4. COGNITION

The following requirements were considered in the selection of a base modeling approach to incorporate into the IAN system modeling function for this research.

1) We wanted to make sure that the cognitive model could interact with an external environment in a software-feasible way. Having the best cognitive model in the world does us no favors if it is unable to communicate with the flight model and IAN function. This closed-loop architecture has been done with cognitive models previously with prime examples seen by ACT-R [1], Soar [5].

2) The cognitive model needs to demonstrate at least face validity; that is, it should accurately approximate how a pilot would behave in different

circumstances. It is felt that a rule-based system would best be suited for a model of an aviator.

Starting with a system that makes some simple assumptions about generic human cognitive performance allows for more time and effort to focus on tailoring the model to flying-related tasks.

3) In order to mimic a human interacting with an aircraft as much as possible, the interface needs to closely emulate the human body perceiving through the eyes and ears and manipulating with hands and feet. A software module/client, named the execution block, will be written that communicates with the cognitive model and creates a representation of arms, hands, legs, feet, head, and eyes. This helps unload the burden on the cognitive model of keeping track of the details regarding the sources of input and output.

4) We are concerned with efficiency. To this end, we aimed to select an approach that minimized development time, and leveraged existing work. In addition, the human model component is designed to run out-of-process, or asynchronously with the aircraft simulator to facilitate parallel development. Once a common interface is described, the human model should be able to fly many types of simulated aircraft that meet that software specification. This also permits development to not be dependent on the Simulink portions of the model in order to test and run it.

5) We are on a schedule and need to make sure that the cognitive model we select doesn't take an unnecessary amount of time to develop. It is undesirable to spend time re-inventing something that someone else has already done. Any existing models and architectures that exist out in the world should be considered as a potential baseline for our model and taken advantage of.

The following sections highlight some of the candidate cognitive models considered and some of the pros and cons of each with regard to their applicability toward our IAN model.

#### **A. ACT-R/PM**

ACT-R Perceptual-Motor (ACT-R/PM) is a set of extensions to ACT-R which provides perceptual-

motor capabilities for ACT-R. The Perception-Motor layer is made up of modules that handle various aspects of perception and action. Among the modules covered with this extension to ACT-R are vision, motor, speech, and audition [2].

One of the difficulties of working with ACT-R is that it is written in Lisp, a powerful, yet older high-level programming language. While it is known for powering artificial intelligence research over the last several decades, it is not as prevalent as other high-level programming languages such as C, C++, and Java for general programming activities.

Another difficulty of ACT-R is that it traditionally is run as a stand-alone application where the cognitive model does not talk easily with any external application/devices/computers. That was resolved with release of the ACT-R/PM module which does incorporate the ability to interact with a simulated device easily. However, the simulated device must be a Lisp object which must have certain methods defined for it, which in turn, will be called automatically by ACT-R/PM at the appropriate times [2].

There also exists a version of ACT-R called jACT-R which is a java implementation of ACT-R. While not comprehensive of all features that ACT-R provides, it covers most of what one would expect of ACT-R, but written in Java rather than Lisp. jACT-R also provides some benefits to interact with external environments and control the models remotely [4].

#### **B. Air MIDAS**

Air Man-machine Integration Design and Analysis System (Air MIDAS) is a modeling and simulation tool designed to assess human-system integration in dynamic aviation-related environments. It is currently being used to analyze advanced air traffic management concepts at San Jose State University where it was originally developed by the Human Automation Integration Laboratory [3].

Like ACT-R, Air MIDAS is programmed in Lisp, although it interfaces with external simulators more natively. Air MIDAS also has the benefit that it was developed with an aviation emphasis and not just a generic cognitive model.

### **C. Soar**

Soar is a theory of a cognitive model that is implemented as software architecture [10]. Soar research today tries to realize an approximation of human behavior and thought while minimizing the sets of mechanisms that are required. Soar memory is associative which means that the flow of control in Soar is not determined by a sequential, deterministic control structure that is used in most programming languages. In other words, evaluation of relevant knowledge can be done in parallel [10].

Another feature of Soar is the ability to automatically create sub goals to help resolve impasses in decision junctions. Soar also takes into account past experiences when adapting to unfamiliar situations and making decisions. This allows Soar to learn new conceptual knowledge, procedures, and even correct its knowledge as it gains feedback through experience in its environment [10].

The most recent release of Soar version 9.0.1 now includes a reinforcement learning (RL) module. Beta versions of Soar include episodic memory and semantic memory modules. These new memory and learning modules greatly enhanced Soar's ability to approximate human memory [9].

All of these positives, in terms of programming, come at the expense of under specifying the capabilities that must be built into intelligent agents. Most of the knowledge that a Soar agent has, stems from rules that have been programmed into it. For the agent to realize high-level intelligent behavior, the knowledge must be created. Soar also can make simplifications which leads to unrealistic behavior in the model [5].

Perhaps one of the greatest benefits Soar provides for a closed-loop model is a standardized way to connect to external simulators via a language called Soar Markup Language (SML). The method was debuted in Soar version 8.6, and has been supported since. The SML specification allows external programs to send and receive information from Soar which allows external simulations, such as a flight simulator and/or custom applications to interact with the human model easily [8].

### **D. Model Selection**

Based upon the model considered, we have determined that Soar best meets our requirements to integrate into the IAN / human model. Soar provided the easiest method to integrate into the closed-loop architecture, provided models that already existed to build upon (Air-Soar, TacAir-Soar), and is a reasonable cognitive approximation with notable, but acceptable faults.

### **E. Model Extension for IAN**

Our concept in the NextGen IAN functions will not focus on modeling perception as much as it will on comprehension and cognition. We are aware that perception of stimuli in the closed-loop simulation can be a factor in the evaluation of IAN functions; however, parameters such as font sizes, brightness, and contrast ratios are prescribed by detailed design specifications for flight decks. We assume that these same design specifications will be used in NextGen avionics. This minimizes the need to study perceptual parameters in the closed-loop simulation of the model.

However, we intend to use the closed-loop simulation, as described in this paper, to determine design specifications for the cognitive processing aspects of IAN functions. The number of simultaneous or near-simultaneous caution, warning, and alerts that may be presented to the pilot could be competing for scarce cognitive resources. Multiple stimuli could be subject to the psychological refractory period wherein the pilot may delay reaction to an important stimulus while attending another. It is these types of scenarios that we wish to use the human model to determine the best candidate IAN functions.

## **5. CREW PERFORMANCE DATA**

The OPL has years of experience collecting data from human pilots in both aircraft and flight simulators to evaluate pilot performance and estimate workload. This capability is being developed under a separate NASA project entitled Operator State Sensor Investigations and Operator State Classification and Feedback Algorithms (NNL07AA00A). A significant piece of software has stemmed from this project called the Cognitive Avionics Toolset or CATS. This program

is used to provide real-time data exploration and analysis to support effective operator state feedback.

It is our hope that we may be able to also gain an understanding of what the workload is of the human model in the IAN simulation. The OPL has conducted several studies in airborne platforms and ground-based simulators that involved collecting workload measurements as well as physiological responses and eye tracking. Empirical data collected from such studies can help us fine-tune the human model for similar scenarios such as the standard terminal arrival route (STAR) approach and landing.

It is also of interest to allow the IAN function to be aware of pilot state during all phases of flight. Should, during long stretches of low workload, the pilot allow their attention to fall elsewhere (or nowhere at all, should they fall asleep), IAN will become aware of the pilot's inattention. How IAN presents information to an attentive versus an inattentive pilot should be quite different. For example, sounding audio cues in addition to presenting visual cues may be necessary to draw the pilot's attention to significant information.

## 6. FUTURE WORK

As part of the IAN / human model project, human-in-the-loop experiments will be conducted using the candidate formats down-selected by the IAN / human model. The CATS software developed by the OPL will be used as the primary collector of human performance data during these experiments. This allows us to estimate the workload of pilots in the simulator with the new IAN functions being tested. This provides a quantitative way of comparing the different IAN functions in terms of reducing workload for pilots and indicates desirable function and display formats.

After the IAN closed-loop model is able to successfully start testing candidate IAN functions and display properties, there will be a selection of the top four candidates. These top candidates will then be implemented in a flight simulator and flown as part of a pilot-in-the-loop experiment/study that will verify the properties and characteristics of the IAN system. The OPL

houses and maintains several research aircraft and flight simulators, including a 737-800 fixed-based flight deck that we will use to conduct the human-in-the-loop study (Figure 3).



Figure 3: OPL's 737-800 Flight Simulation Facility

## 7. TEST SCENARIOS

Under consideration for scenarios to be tested with the human model are terminal approaches. In a NextGen aircraft, this would not typically involve much from the human model other than to act as a supervisor for the auto-pilot. In order to put the IAN system to the test, a circumstance that would cause pilot intervention needs to be a part of the scenario. While these scenarios are still being finalized at the time of this writing, one scenario being developed is an aircraft receiving a STAR to O'Hare (ORD). This scenario would involve conflicting air traffic interfering with the intended route to the airport. Presentation of this information to the pilot(s) would progress from simple messages and advisories to full warnings (if action was not taken earlier).

## 8. ACKNOWLEDGEMENTS

We would like to acknowledge the National Aeronautics and Space Administration under grant number NNX08BA01A. We also would like to thank Dr. Maarten Uijt DeHaag and his great team at Ohio University. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of NASA, the U.S. Government, or any other organization.

## 9. REFERENCES

1. Byrne, M.D.; Kirlik, A.; Fleetwood, M.D.; Huss, D.G.; Kosorukoff, A.; Lin, R.; Fick, C.S., 2004. "A closed-loop, ACT-R approach to modeling approach and landing with and without synthetic vision system (SVS) technology." *Proceedings of the Human Factors and Ergonomics Society 48<sup>th</sup> Annual Meeting*.
2. Byrne, Mike. ACT-R/PM Theoretical Notes. Computer-Human Interaction Lab at Rice University. 20 April 2009. [http://chil.rice.edu/projects/RPM/theory\\_notes.html](http://chil.rice.edu/projects/RPM/theory_notes.html)
3. Freund, Louis. San Jose State University. Telephone Interview. Conducted 15 July 2009.
4. Harrison, Anthony M. jACT-R. 20 July 2009. <http://jactr.org>
5. Jones, R.M. Laird, J. E., Nielsen, P.E., Coulter, K. J., Kenny, P., Koss, F. V. (1999). "Automated Intelligent Pilots for Combat Flight Simulation." *AI Magazine*. Spring, 1999.
6. Microsoft. "About SimConnect." 4 March 2009. <http://www.fsinsider.com/developersPages/AboutSimConnect.aspx> .
7. NASA, Aviation Safety Program. Aeronautics Research Mission Directorate. 20 July 2009. [http://www.aeronautics.nasa.gov/programs\\_aviation\\_safe.htm](http://www.aeronautics.nasa.gov/programs_aviation_safe.htm) .
8. Pearson, Douglas. "XML Interface to Soar (SML) Software Specification." August 28, 2008.
9. Soar Group. "Soar-EpMem Manual". 11 May 2009. <http://sitemaker.umich.edu/soar/>
10. Soar Technology. "Soar: A Functional Approach to General Intelligence." Ann Arbor: 2002. <http://www.soartech.com> .

# Modeling the Effects of Stress: An Approach to Training

Taryn Cuper

MYMIC LLC

taryn.cuper@mymic.net

**Abstract.** Stress is an integral element of the operational conditions experienced by combat medics. The effects of stress can compromise the performance of combat medics who must reach and treat their comrades under often threatening circumstances. Examples of these effects include tunnel vision, loss of motor control, and diminished hearing, which can result in an inability to perceive further danger, satisfactorily treat the casualty, and communicate with others. While many training programs strive to recreate this stress to aid in the experiential learning process, stress inducement may not always be feasible or desired. In addition, live simulations are not always a practical, convenient, and repeatable method of training. Instead, presenting situational training on a personal computer is proposed as an effective training platform in which the effects of stress can be addressed in a different way. We explore the cognitive and motor effects of stress, as well as the benefits of training for mitigating these effects in real life. While many training applications focus on inducing stress in order to “condition” the stress response, the author explores the possibilities of *modeling* stress to produce a similar effect. Can presenting modeled effects of stress help prepare or inoculate soldiers for stressful situations in which they must perform at a high level? This paper investigates feasibility of modeling stress and describes the preliminary design considerations of a combat medic training system that utilizes this method of battlefield preparation.

## 1.0 INTRODUCTION

Stress is defined as *a condition typically characterized by symptoms of mental and physical tension and strain that can result from a reaction to a situation in which a person feels threatened, pressured, etc* [1]. The effects of stress manifest both emotionally and physiologically and can be beneficial or detrimental to performance, depending on the level of stress and performance requirements. In the case of a combat medic, stress levels can be expected to be extremely high as the lives of both the medic and his comrades are at risk. In addition, the medic's performance requirements are complex, demanding both gross (e.g. running) and fine (e.g. applying a tourniquet, firing a weapon) motor skills as well as unaffected cognitive functioning. Therefore, it is important to take into account the effects of stress when designing any combat medic training program in order to adequately represent realistic situations and to prepare medics with the tools and strategies to effectively do their job.

This paper describes an investigation into the concept of stress response modeling and its inclusion in the design of a Complex Incident Response Training System (CIRTS) for Combat Medics (CM). The CIRTS-CM design is of a PC game-based system for training combat medics to respond to IED-ambush scenes. While the primary foci of CIRTS-CM are site management and

casualty care, the integration of stress is considered an important element of providing a realistic approach to this type of training.

## 2.0 BACKGROUND

The emotional effects of stress include fear and anxiety while physiologically, stress results in, for example, an increased heart rate. Stress also affects cognitive functioning. When cognitive load surpasses capacity, the result can be decreased perception of the environment and an inability to think through known processes or to problem-solve [2]. Though the domains of the effects of stress are often addressed separately, that is not to assume they are exclusive of one another. Instead, the effects of stress are interrelated mechanisms that are divisible only by how they present in the face of a stress source, or *stressor*.

### 2.1 Physiological effects of stress

Siddle [3] refined reported effects of stress to physiological measurements, linking the deterioration of fine motor skills and cognitive functioning and the enhancement of gross motor skills to increasing heart rate:

- >115 bpm: loss of fine motor skills
- >145 bpm: loss of complex motor skills; visual system decreases the peripheral view and attention

- >175 bpm: gross motor skills only; auditory exclusion; tunnel vision; "freezing"; deterioration of depth perception; deterioration of information processing skills

This relationship between performance and heart rate can be explained by the Inverted U Hypothesis, which states that performance increases with arousal (e.g. heart rate) up to a certain point, at which time performance starts to deteriorate. Levitt's research supports both this hypothesis and Siddle's findings, defining this point of descent at 145bpm, with optimal performance between 115 and 145 bpm; heart rates below 80 and above 175 bpm correspond to poor performance [4].

These increases in arousal can affect almost all aspects of performance, especially vision, hearing, physical movement, and cognitive abilities, either singularly or in combination [5] depending on the length and severity of an engagement [6]. Following are some selected physiological effects of combat-related stress from a collection of first-person reports compiled by Artwohl & Christensen [7]:

- Auditory exclusion: Inability to hear or severely diminished hearing of one's own and other's gunfire, as well as sirens and verbal communications; hearing a "pop" noise and/or the sound of the weapon's slide moving forward and backward instead of the gunshot; sound of one's own gunfire sound very far away. In addition, inability to hear the shot or explosion that is close enough to harm.
- Tunnel vision or perceptual narrowing: seeing only the ring on a shooter's finger while being fired at; focusing so intently on the adversary's pistol, the officer shot it and the trigger finger out of his hand; leveraging this phenomena on an adversary by side-stepping to gain the advantage. In addition, the rare occurrence of complete loss of sight, a rare occurrence.
- "Scared speechless": inability to speak intelligibly into radio; making a "weird, inarticulate garbling sound" instead of speaking; focus on training police officers to give specified verbal instructions

constantly so that this is an automatic response in a real situation.

- Slow-motion time: time slowed so that an officer felt he could consider options and carefully aim in what was actually an extremely quick movement; ability to see bullets in the air as if they were low-velocity projectiles, e.g. paintball rounds.

A similar phenomena to these experiences is hypervigilance [8]. Hypervigilance is characterized by freezing in place or existing in a state of extreme agitation in reaction to perceived threat. In such cases, an individual may persist in a frantic search for escape routes, perform irrational actions, or perform the same actions over and over again though they have proven ineffective at achieving the individual's goal. Each of these reactions can be found within the effects of high performance anxiety, which is characterized by the, "temporary loss of cognitive efficiency and a temporary loss of perceptual acuity, perceptual-motor coordination, and motor skills," [9]. This similarity in the effects of combat related stress and anxiety indicates that hypervigilance is a complex response to both a perceived threat and a lack of confidence in one's ability to control the threat.

## 2.2 Coping with and/or overcoming the effects of stress

Training to expected conditions is a common way to prepare individuals for the challenges they will face in a real-life situation. This can be accomplished via live exercise with varying levels of simulation augmentation [10] wherein trainees have direct exposure to real-life or close to real-life stressors. One of the goals of this approach is to make training the first place these stressors are experienced to hopefully mitigate their debilitating effects on an individual when his life is threatened and/or his performance is a determining factor of others' survival.

Stress Inoculation Training (SIT) extends this approach, utilizing lessons learned from cognitive behavioral therapy. This process consists of three phases: conceptualization, skills acquisition and rehearsal, and application and follow through [11]. During the conceptualization phase, the trainee is educated about the nature and possible impacts of stress and trained to view stressors as problems to be solved; in this way, the trainee's mental model of the stress has been modified. Next is the

skills acquisition and rehearsal phase, in which coping skills are taught and practiced in a "laboratory" setting. Last, during application and follow-through, trainees apply the coping skills in increasingly stressful situations, such as realistic training scenarios. SIT has been successful at enhancing one's skill at coping in extremely stressful situations, such as military combat [12] [13]. A particularly significant aspect of SIT is the conceptualization phase, in which trainees are educated about the effects of the stress they will encounter.

### **2.3 Survival stress management**

Survival stress management can be best characterized by first examining Martens' definition of stress [14], which emphasizes threat perception, perception of response capability, and perception of time required for effective threat management. Survival stress management training addresses these perceptions, the primary goal of which is to increase the trainee's confidence in his own capabilities; remaining training goals include experience, visualization, and breathing [3].

As mentioned in the discussion above on the relationship between the effects of fear and anxiety, confidence is a determining factor of the strength of one's survival stress response. Building this confidence takes a two-pronged approach: confidence in the specific skills to be performed and confidence in the context in which the skills are performed. As such, the trainee builds this confidence through both skills training and effectiveness and experience through dynamic exercises.

Recognition of the symptoms leading to adverse effects of stress is a key component of learning to control these effects. Situational awareness plays a role in this step as trainees learn to recognize potential threats and visualize Plan A and Plan B responses. This objective is enhanced by the last training goal of breath control. Among other physiological effects, controlling the rate of breathing can result in a lower heart rate, thereby minimizing the physiological effects of stress associated with bpm above.

### **3.0 CIRTS-CM PRELIMINARY DESIGN FOR STRESS**

Some training modalities are not amenable to stress inducement; in these cases, direct exposure to stressors may not be practical,

feasible, or desired. For example, consider the training environment of the CIRTS-CM. In this case, it is not feasible to impart fear for one's life or the physical effects of running for cover to the trainee when they are safely located in a training room in front of a computer. While the trainee may become so engrossed in training that an accompanying increase in heart rate is experienced, this physiological response will not meet the levels experienced in combat.

In addition, it is not desirable for this tool to induce stress for the purpose of eventual inoculation as this additional training focus may take away from the stated training goals of the system. However, in recognition of a desire for elevated engagement and realism, the element of stress will be addressed in a manner that can be adequately executed on a PC as well as enhance the training goals of the system.

Teaching the effects of stress, not only the effects on perception but also on the accompanying effects on job performance, is a unique opportunity for game-based training. The proposed approach to this challenge is *stress response modeling*. This concept is similar to the conceptualization stage of SIT, in which trainees are educated about the effects of stress. It also addresses the Visualization training goal of survival stress management, in which trainees learn to recognize the onset of survival stress in order to take steps towards mitigation. In this environment, stress response modeling can occur in situ; the nature of the model presents the effect of stress while the resulting inability of the trainee to perform actions that require that perceptive capability compromises performance in the game, as it would in real life. This representation thereby enhances the realism of training modality.

For example, the CIRTS-CM will impose the effect of tunnel vision on the player by blacking out the periphery of the screen. Another possibility is muting the sound temporarily to replicate auditory exclusion. These models can serve as practical experience to reinforce the education of the medic as to what can happen to him in a real-life, stress inducing scenario.

In addition to modeling the effects of stress to the trainee, CIRTS-COM will address the element of stress with regard to task difficulty. External stimuli, i.e. visual and auditory stimuli, will build from level to level, requiring high performance in the face of diminishing cognitive resources.

#### 4.0 DISCUSSION

While examples of human stress response modeling are not common, two similar approaches were identified and are explained below.

One example leverages technology to educate people on the effects of cognitive states other than stress. In the domain of abnormal psychology, the University of California at Davis has created the Virtual Hallucinations Project, which seeks to educate by modeling an exemplar experience of schizophrenia for visitors in the Second Life virtual world [15]. Virtual Hallucinations presents visitors with auditory and visual "hallucinations" similar to what a schizophrenic might experience. This use of computer-based stimuli to communicate the physiological effects on perception is very similar to the stress response modeling concept presented in this paper.

In addition, flight simulation training provides instances that further support this idea. For example, Microsoft's Combat Flight Simulator 3, a PC-based flight simulation application, incorporates G-force effects such as vision degradation and blackout. It recognizes the significance of these effects by making G-tolerance a criteria for pilot profile selection and skill point allocation [16]. Other simulators incorporate similar blackout simulation capabilities, such as Aero-Elite for PlayStation 2, Falcon 4, and Aces High for PC.

G-force blackout modeling is a controversial aspect of flight simulation. Critics claim the models are not accurate and thereby detract rather than add to realism [17]. Supporters respond by highlighting the unrealistic nature of a pilot who is not affected by any level of G-Force (while in some cases the non-player character opponents are affected by G-force). This school of thought maintains that some representation of G-Force effects is better than none when the end goal is to cause the pilot to discontinue high-G maneuvers [18].

#### 5.0 CONCLUSION

While using physics models to simulate the effects of environmental stimuli are a common way to utilize technology in training, examples of similarly modeling human responses to external stimuli are rare. However, it appears possible that this approach might serve some benefit to combat medics who might experience and be confused by the effects of stress in real life situations by

providing them with the knowledge of what these effects are and how these effects can hinder performance. This promises to be an effective method for addressing the non-training focus element of stress in PC game based training.

#### 6.0 REFERENCES

- [1] Agnes M & Guralnik DB (Eds.). (2007). *Webster's New World College Dictionary*. Wiley Publishing Inc: Cleveland, OH.
- [2] Hockey GRJ. (1997). Compensatory control in the regulation of human performance under stress and high workload: A cognitive-energetical framework. *Biological Psychology*. 45:73-93.
- [3] Siddle BK. (1995). *Sharpening the Warrior's Edge: The Psychology & Science of Training*. PPCT Research Publications: Belleville, IL.
- [4] Levitt S. (1972). The effects of exercise induced activation upon simple, two-choice and five-choice reaction time and movement time. Doctoral Dissertation.
- [5] Grossman D & Christensen LW. (2007). *On Combat: The Psychology of Deadly Conflict in War and in Peace*. PPCT Research Publications.
- [6] Huey BM & Wickens CD. (1993). Workload factors. In *Workload transition: Implications for individual and team performance* (pp. 54-93). Washington, DC: National Academy Press.
- [7] Artwohl A & Christensen LW. (1997). *Deadly Force Encounters*. Paladin Press: Boulder, CO.
- [8] Janis IL & Mann L. (1977). *Decision making: A psychological analysis of conflict, choice, and commitment*. Free Press: New York, NY (Accessed July 2009 at: <http://www.cedu.niu.edu/~fulmer/grouphink.htm>).
- [9] Janus I, Defares P, & Grossman P. (1983). Hypervigilant reactions to threat. In H. Selye (Ed.), *Selye guide to stress research*, Vol. 3. Scientific and Academic Editions: New York.
- [10] Seibel JA. (2009). *Reality-based training: Skill development in Survival Stress Response*. Department of Interdisciplinary Technology, School of Police Staff and Command. Flint, Michigan.
- [11] Meichenbaum D (1996). Stress inoculation training for coping with stressors. *The Clinical Psychologist*. 49:4-7.

[12] Meichenbaum D. (1994). Stress Inoculation Training: A treatment and preventative approach. in Lehrer PM, Woolfolk RL, & Sime WS (Eds.). (2007). *Principles and Practice of Stress Management* 3rd Edition. Guilford Press.

[13] Novaco RW. (1977). Stress inoculation: a cognitive therapy for anger and its application to a case of depression. *Journal of Consulting and Clinical Psychology*. 45(4): 600-608.

[14] Martens R. Arousal and motor performance. *Exercise and Sports Sciences Reviews*. 2(1): 155-188.

[15] Cook J. UC Davis: Virtual Hallucinations Project. In Linden Lab's *Second Life* virtual world. Accessed May 2009 at: <http://slurl.com/secondlife/sedig/27/45/22/>.

[16] Rymaszewski M. (2002). *Combat Flight Simulator 3: Sybex official strategies and secrets*. Sybex, Inc..

[17] "Smoke"CO. (1999). Blackout modeling in sim design. Combatsim.com. Accessed August 2009 at: <http://www.combatsim.com/htm/may99/blackout.htm>.

[18] Crenshaw D "Crash". (2009). To blackout or not to blackout. Aviation and Air Combat. Accessed August 2009 at: [http://www.simhq.com/\\_air/air\\_03\\_6a.html](http://www.simhq.com/_air/air_03_6a.html)



# Development of a Human Motor Model for the Evaluation of an Integrated Alerting and Notification Flight Deck System

Ron Daiker, Graduate Student; Thomas Schnell, Associate Professor

University of Iowa

*daiker@engineering.uiowa.edu; tschnell@engineering.uiowa.edu*

**Abstract.** A human motor model was developed on the basis of performance data that was collected in a flight simulator. The motor model is under consideration as one component of a virtual pilot model for the evaluation of NextGen crew alerting and notification systems in flight decks. This model may be used in a digital Monte Carlo simulation to compare flight deck layout design alternatives. The virtual pilot model is being developed as part of a NASA project to evaluate multiple crews alerting and notification flight deck configurations. Model parameters were derived from empirical distributions of pilot data collected in a flight simulator experiment. The goal of this model is to simulate pilot motor performance in the approach-to-landing task. The unique challenges associated with modeling the complex dynamics of humans interacting with the cockpit environment are discussed, along with the current state and future direction of the model.

## 1. INTRODUCTION

According to the National Transportation Safety Board, most civil aviation accidents occur in close proximity to airports [9]. The majority of these accidents are attributable to human error, especially when pilots are operating in Instrument Meteorological Conditions (IMC), due to a loss of Situational Awareness [9]. Tasks that require physical movements can draw the operator's attention away from other duties, thus reducing a pilot's situational awareness [14]. Pilots often experience task overload during this phase of flight due to the higher workload demands of reconfiguring the aircraft for landing while also interacting with Air Traffic Control (ATC) and crew members to safely navigate congested airspace.

At its most basic level, human limitation can be considered from the perspective of physical movement, or how quickly and accurately the pilot interacts with aircraft controls. If aircraft designers and those who develop in-flight procedures better understood the limitations of the human organism, they could improve cockpit layouts and procedures.

These layouts and procedures can be evaluated with a pilot motor model before they are finalized, identifying potential hazards and improving pilot/aircraft interaction. For example, designers may wonder if the operator will be able to complete all of the required tasks (manipulating controls, for example) within the allotted time, or which flight deck layout would be the best for

responding to an emergency situation. Designers could then use the pilot reach model to answer questions like these in the early phases of development safely and affordably, testing several layouts and procedures with a virtual human model.

Ohio University, the University of Iowa, Boeing, and Rockwell Collins, are working under a grant from NASA to develop an Integrated Alerting and Notification (IAN) solution to aid in flight deck decision making. The IAN project is part of the Integrated Intelligent Flight Deck (IIFD) research group of AvSafe, NASA's aviation safety program.

The University of Iowa has been tasked with the development of a virtual human pilot model for testing multiple alerting and notification types within a simulated flight deck. The motor model described in this paper may be used to evaluate multiple cockpit layouts and configurations through a series of Monte Carlo simulations. This may allow for down-selection of flight deck control layout design alternatives. The motor model may also be paired with a perception and cognition model to result in a more complete virtual pilot model. This article outlines the development of a human reach model which will comprise the link between the virtual human pilot and the simulated flight deck.

## 2. BACKGROUND

Many virtual human anthropometric models have been developed and implemented in the area of human factors and ergonomics research. The majority of these models have focused on calculating reach distances and comfort levels while performing a variety of other tasks. Some examples include virtual human models like Jack [2], HUMOSIM [11], HADRIAN [11], MIDAS [13], and Air MIDAS [10].

Jack [2] and HADRIAN [11] are virtual human avatars that can be used to create visual representations of humans interacting within a virtual environment. Both possess realistic limb and joint functions, including basic information for measuring reaching ability and comfort levels. The Jack avatar is often used by human models like MIDAS and HUMOSIM to visualize the model interacting with a virtual world. The HADRIAN anthropometric model was founded on the "design for all" principal, and claims to model a wider variety of body types [11]. Both virtual avatars contain the reach calculations required for cockpit layout evaluations, but they also contain many additional features that would unnecessarily slow down the Monte Carlo simulation used in this project.

The Man-machine Integration Design and Analysis System (MIDAS) human model has been used for many aviation related tasks, from modeling human/cockpit interaction [5] to air traffic control display evaluation [4]. The MIDAS human model contains a powerful cognitive architecture for modeling human behavior and a highly accurate environment model for creating cockpit interiors.

Air-MIDAS is an adaptation of the MIDAS model that includes additional enhancements for modeling pilot cognition and behaviors. The Air-MIDAS model has been used as a predictive model for the evaluation of flight crew performance when interacting with varying levels of automation [10]. Both MIDAS and Air-MIDAS rely on the JACK virtual avatar for the execution of motor functions, but (as stated above) the JACK virtual human motor model is not well suited for this project due to its higher computational demands.

The HUMOSIM model has been used to evaluate automobile seat comfort [16] and human variability in reaching motions [7], and it also contains highly detailed biomechanics and movement prediction models. MIDAS, Air-

MIDAS, and HUMOSIM are very complex human models, but our objective in this study was to obtain very computationally efficient models for use in multiple Monte Carlo simulations.

All of these models provide useful features, but they are generic and were not created specifically for the purpose of modeling pilots controlling an aircraft. Pilots make up a very small subset of the general population, and they are selected based on specific physical attributes (height, vision, physical fitness, etc.). The models presented here were designed specifically to emulate real pilot reach performance in completing the approach to landing task.

## 3. METHOD

**APPARATUS** An experiment was conducted in the Operator Performance Laboratory's flight deck simulator that is based on the Boeing 737-800 form factor. This fixed base simulator features five outside visual projectors, a semi-spherical screen, and an operational cockpit.

An electromyography device (or EMG) was used to record the initiation of reach movements in the frontal deltoid region of each pilot's right arm. All pilot participants were seated in the left seat and instructed to use their right arm for completing reaching tasks during the scenario. Control inputs made by the left arm and the feet were recorded by the yoke and rudder pedals, respectively.

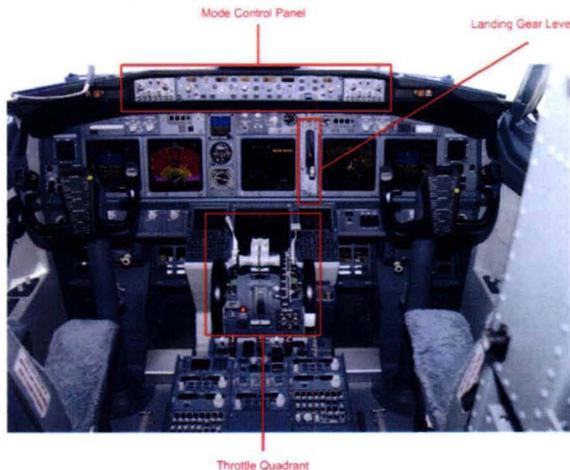
A digital video camera was positioned above and behind the pilot to record the initiation of each movement. The recorded video was later compared to the collected EMG and simulator data to analyze each movement.

**DESIGN** EMG data was collected for nine pilots during the experiment, and each of the participating pilots held at least an IFR rating. Pilots varied in their level of experience with this flight deck layout, and for some participants this study was their first encounter with this cockpit configuration.

The participants were instructed to fly three replications of an approach to landing scenario in IFR conditions. A simulated approach into runway 9R at O'Hare International Airport (KORD) was flown by each pilot under three varying levels of automation; fully coupled autopilot mode, flight director mode with auto-

throttle, and manual approach with only sectional charts.

**MODELING** Three models were developed to measure the time delay required for the completion of pilots' reaching motions in the approach to landing task. Two cockpit layouts were compared, the first being that of a flight deck similar to a Boeing 737-800 in its current configuration (see Figure 1), and the second being a modified cockpit layout which has been designed to reduce reach distances for faster control manipulation. The modified cockpit features controls that have been moved closer to the pilot by approximately half their current reach distances, and was performed mathematically for comparison purposes. This modification illustrates the effect a change in cockpit layout can have on pilot model reach times.



**Figure 1.** Current Boeing 737 Layout

All three models re-create the reaching motions a pilot makes while completing the approach to landing task. The pilot's hands and feet are only modeled as placeholders, with their arrival at a targeted control signifying the completion of a reaching task. The components of the aircraft cockpit that were modeled included the control locations which pilots manipulate during completion of the final approach check list. Each cockpit control has its own "control box", or area in which it can be manipulated by the pilot. Control boxes identify at which point the pilot model's reaching task is completed and control manipulation can begin.

For this experiment, the overall approach to landing task has been divided up into several subtasks consisting of individual reaching actions. The yoke was selected as the point of

origin because pilots are trained to keep their hands on the yoke during the approach to landing task. The target point is the location of the control which the pilot model has been instructed to manipulate. For example, the "Yoke to Gear Lever" task is defined as the time it takes for the pilot to reach from the control yoke to the landing gear lever.

Pilot model reach times were based on two sources; experimental data and calculations derived from a combination Hick's Law [6] and Fitts' Law [3]. Hick's Law was used to calculate reaction time (while considering the number of alternatives) and Fitts' Law was used to calculate the reach time to interact with a control mechanism. The combination of these two methods was used to generate response/reach time values for the Computed Pilot Model. This model was developed purely for the purpose of comparison to the other pilot models currently in development.

$$\text{Fitts' Law: } MT = a + b \log_2(2A/W)$$

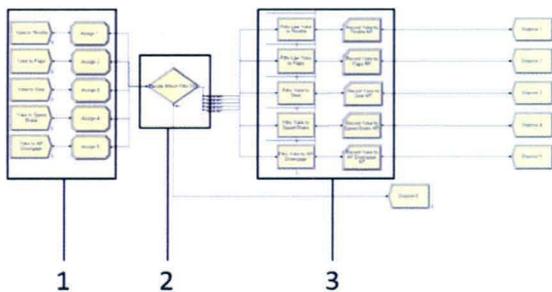
$$\text{Hick's Law: } T = b \log_2(n + 1)$$

The other two pilot models were derived from experimental data collected during the study. Thus far, only the data for two pilots has been analyzed and included in the models. The two pilots varied in their familiarity with the flight deck layout, and the models representing each bear the names "Familiar Pilot Model" and "Unfamiliar Pilot Model" for comparison. The Familiar Pilot was very experienced in locating and manipulating controls, and had participated in at least three experiments in the Operator Performance Laboratory flight deck simulator. This participant holds the most experience of all the pilots who participated in the study. The Unfamiliar Pilot had no experience with the layout prior to the experiment. As more pilot data is analyzed, these two models will continue to grow and change to more accurately represent pilot performance.

The movements of these pilots were recorded and their performance was later analyzed. The initiation of each movement was captured using EMG sensors and reach task completion was recorded in the form of time stamped simulator control inputs. The difference between these two values (accurate to one millisecond) forms the task movement time. Digital video collected during the experiment was also used to identify reach movement initiation and the type of reach being performed.

All three pilot state models (Familiar, Unfamiliar, and Computed) were used to evaluate task completion in both the standard and modified flight deck layouts. The models were developed using Arena, a discrete events simulation software developed by Rockwell Automation. The software can be adapted to fit many systems, from manufacturing processes to liquid flow mapping. A brief overview of a small portion of the model's structure is provided in Figure 2. The model uses multiple "Create" modules to introduce entities into the system at a rate that is controlled by the operator. Each entity represents a reaching task from an origin point to a target destination that is released at a time which corresponds to the final approach checklist.

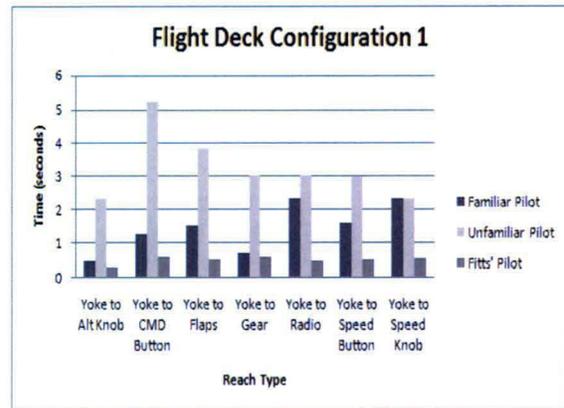
The task entities are then sent to a "Decide" module for sorting based on their origin and target criteria. The Decide module then transfers these task entities to "Delay" modules which apply a delay based on a distribution derived from either experimental data or Hick's Law and Fitts' Law. The Process modules then apply the appropriate delay to the task which is representative of the time it takes for the pilot to complete a reaching task.



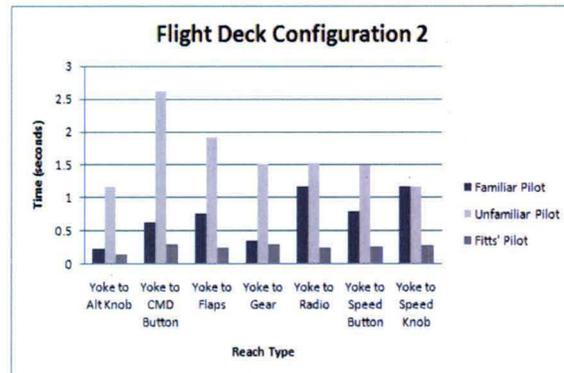
**Figure 2.** A brief overview of the model structure; 1) Create modules, 2) Decide module, 3) Delay modules

#### 4. RESULTS

The reaction and reach time data for each model was fit to a distribution using Arena's Input Analyzer. Each model was then replicated one thousand times in a Monte Carlo simulation to obtain average reach time performance values. The average reach times of all three pilot models in both flight deck configurations is shown in Figures 3 and 4.



**Figure 3.** Average reach times for each model with flight deck configuration one.



**Figure 4.** Average model reach times for flight deck configuration two.

As Figures 3 and 4 illustrate, there is a substantial difference in performance between the three models in completion of the reaching tasks. This may be attributable to experimental data in which pilots initiated a reach but then hesitated to search, not completing the reach until the target control could be located. Even though there are significant differences in performance, the change in cockpit layout resulted in approximately a 100 millisecond improvement in lever reach times across all three pilot models. The relocation of the lever controls to a closer position (approximately half the current reach distance) resulted in an improvement in pilot reach time.

An Analysis of Variance (ANOVA) comparing the performance of each pilot model in both flight deck configurations was performed. A statistically significant effect was found between flight deck layouts 1 and 2 across all three pilot models, with  $F_{1,12}=5.76$ ,  $p=0.0335$  (Familiar

Pilot),  $F_{1,12}=14.65$ ,  $p=0.0024$  (Unfamiliar Pilot), and  $F_{1,12}=34.71$ ,  $p=0.00007$  (Modeled Pilot).

## 5. FUTURE DEVELOPMENT

This article outlines the early stages of development for this motor model. Future work will focus on incorporating all of the collected pilot data into one model and evaluating its performance against other human motor models and more collected pilot data for validation. The completed motor model will provide an accurate reflection of pilot performance in a small, computationally efficient package that will be ideal for the virtual pilot model being developed.

## 6. DISCUSSION

The early results of this study illustrate the usefulness of this modeling tool for measuring the effect of new cockpit layouts on pilot reach time. This model is limited because it only seeks to reduce pilot reach time, so the end results can be summarized as "closer is better". Unfortunately, this is not always the case, especially when considering pilot comfort, frequency of control use, space constraints, and a multitude of other factors important to ergonomic design. In order to serve a practical purpose, this model would need to take into account at least one other factor and balance the two in some meaningful way to come to a conclusion that is more valuable than "closer is better".

The differences in reach completion times between the three models must also be considered. The Computed pilot model (based on Fitts' Law and Hick's Law) fails to capture the time required to search for the correct controls to manipulate during a reaching task. The Familiar and Unfamiliar pilot models more accurately predict real world pilot performance, and should continue to improve in this regard as they are developed further with the incorporation of more pilot reach time data.

The effect these modifications would have on overall crew performance must also be considered. The models developed only consider the pilot acting alone in the cockpit, without a co-pilot who shares tasks and responsibilities. The layout of the controls on a typical transport aircraft flight deck are not optimized for single pilot operations but rather

for use by a crew consisting of a pilot and first officer. Any changes to this layout could have negative effects upon the shared cognition that occurs between the flight crew and the cockpit environment [8].

Also, these models do not take into account reaching tasks being completed by a co-pilot or other crew members. However, this model could be very useful in predicting pilot performance in smaller, single pilot flight decks. The results of this effort have laid the ground work for an interactive operator reach model that (with further development) will be useful for cockpit task analyses. This tool will aid aircraft designers in placing controls in improved locations to reduce pilot movement time during emergency situations. It will also help those who write cockpit procedures for airliners, ensuring that the procedures they define can reduce unnecessary movements. This would be especially useful when outlining tasks to be performed during phases of flight that already require much of the pilot's attention. Safety investigators can also use this tool when reconstructing the events leading up to an aircraft accident. For example, in the case of an equipment malfunction, could the pilot have conducted the necessary actions to avert disaster within the time allowed?

The motor model can also be expanded to evaluate pilot motor function in other phases of flight, or for the evaluation of emergency procedures (landing gear failure, engine fire, etc.). It can even be adapted for the evaluation of other transportation interfaces. With the addition of a repetitive motion damage algorithm, this model can be used to evaluate operator interfaces to improve occupational health and safety. While the narrow focus of this model does not include all factors that contribute to pilot error in aircraft accidents, it provides a useful tool for improving pilot/aircraft interaction.

## 7. ACKNOWLEDGEMENTS

We would like to acknowledge the National Aeronautics and Space Administration under grant number NNX08BA01A. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of NASA, the U.S. Government, or any other organization.

## 8. REFERENCES

1. Azuola, F., Badler, N.I., Ho, P., Kakadiaris, I., Metaxas, D. and Ting, B., 1994 "Building anthropometry based virtual human models", *Proceedings of IMAGE VII Conference.*, Tuscon, AZ, June, 1994.
2. Badler, N., Phillips, C., Webber, B., "Simulating humans: computer graphics animation and control", Oxford University Press, Inc., New York, NY, 1993
3. Fitts, P.M., 1954. "The information capacity of the human motor system in controlling the amplitude of movement," *Journal of Experimental Psychology.*, Vol. 47, pp 381-391, 1954.
4. Gore, B.F., & Jarvis, P.A., 2005, "New integrated modeling capabilities: MIDAS' recent Behavioral Enhancements", *Proceedings of the 2005 Aerospace Congress*, Paper 2005-01-2701.
5. Hart, S.G., Dahn, D., Atencio, A., and Dalal, M.K., 2001, "Evaluation and Application of MIDAS v2.0", Society for Automotive Engineers (SAE paper# 2001-01-2648)
6. Hick, W. E., 1952, "On the rate of gain of information", *Quarterly Journal of Experimental Psychology*, Vol. 4, pp 11-26.
7. Hu, J., Faraway, J., 2001, "Modeling Variability in Reaching Motions", *Proceedings of SAE Digital Human Modeling for Design and Engineering Conference*, Arlington, VA, June 26-28, 2001.
8. Hutchins, E., 1995, "How a cockpit remembers its speeds", *Cognitive Science*, Vol. 19, pp. 265-288.
9. National Transportation Review Board Annual Review of Aircraft Accident Data: U.S. Air Carriers, Calendar Year 2005
10. Pisanich, G., Corker, K., 1995, "A Predictive Model of Flight Crew Performance in Automated Air Traffic Control and Flight Management Operations", *Conference on Aviation Psychology*, Columbus, Ohio, April, 1995.
11. Porter, J.M., Marshall, R., Sims, R.E., Gyi, D.E. and Case, K., 2003, "HADRIAN: a human modeling CAD tool to promote 'design for all'", *Proceedings of INCLUDE 2003: inclusive design for society and business*, Royal College of Art, London, 2003, pp 222-228, ISBN 1 874175 94 2 [CD-ROM].
12. Reed, M.P., Faraway, J., Chaffin, D.B., Martin, B.J., 2006, "The HUMOSIM Ergonomics Framework: A New Approach to Digital Human Simulation for Ergonomic Analysis." *SAE Technical Papers*, 2006-01-2365
13. Tyler, S., Neukom, C., Logan, M., and Shively, J., 1998, "The MIDAS Human Performance Model", *Proceedings of the Human Factors and Ergonomics Society 42<sup>nd</sup> Annual Meeting*. Chicago, Illinois. Pp 320-325.
14. Wickens, C. D., & Hollands, J. G., 2000, "Engineering Psychology and Human Performance (Third ed.)". Upper Saddle River, NJ: Prentice-Hall Inc.
15. Wickens, C.D., Lee, J.D., Liu, Y, Gorden, S.E., 1997, "An Introduction to Human Factors Engineering, 2<sup>nd</sup> Edition". Prentice Hall.
16. Xuedong, Z., Chaffin, D.B., and Thompson, D.D., 1997, "Progress with Human Factors in Automotive Design: Seating Comfort, Visibility, and Safety", *SAE Special Publications; 1243:101-105*, 1997. SAE Transactions 1997.

# Eye Tracking Metrics for Workload Estimation in flight Deck Operation

**Kyle Ellis, Graduate Student; Thomas Schnell, Associate Professor**  
*Operator Performance Laboratory, University of Iowa*  
*kkellis@engineering.uiowa.edu; tschnell@engineering.uiowa.edu*

**Abstract.** Flight decks of the future are being enhanced through improved avionics that adapt to both aircraft and operator state. Eye tracking allows for non-invasive analysis of pilot eye movements, from which a set of metrics can be derived to effectively and reliably characterize workload. This research identifies eye tracking metrics that correlate to aircraft automation conditions, and identifies the correlation of pilot workload to the same automation conditions. Saccade length was used as an indirect index of pilot workload: Pilots in the fully automated condition were observed to have on average, larger saccadic movements in contrast to the guidance and manual flight conditions. The data set itself also provides a general model of human eye movement behavior and so ostensibly visual attention distribution in the cockpit for approach to land tasks with various levels of automation, by means of the same metrics used for workload algorithm development.

## 1. INTRODUCTION

Operators in today's aircraft flight decks find themselves in various situations that change their cognitive workload. Research to improve the interaction between the operator and the aircraft interface is benefited by being able to analyze operator state quantitatively as opposed to the historical standard of subjective feedback. This eliminates the subjective bias across subjects and standardizes feedback to provide more accurate analysis of operator state in different testing scenarios in flight deck operations. The empirical data found within the data set is useful in creating human eye movement simulation models. The eye movement metrics, the experimental procedure, and findings are described in this paper. Together, they are part of ongoing research at the University of Iowa's Operator Performance Laboratory (OPL) initiatives to both model human attention and pilot workload in flight deck environments, funded through NASA research grants.

## 2. BACKGROUND

There are several opportunities to advance flight decks of the future through utilization of real-time pilot workload assessment. Current avionics are not aware of pilot real-time capabilities and limitations resulting from varying workload levels. In flight deck operations there exists the potential for information overload in various phases of flight and various circumstances. Several systems within the flight deck itself, such as the flight management system and autopilot, are very effective at making easy procedures easier and hard procedures harder in situations with dynamic changes; such as unexpected occurrences in flight. If the avionics could be aware of pilot state, they could provide dynamic displays with situationally appropriate information.

The concept of the intelligent flight deck is currently being defined by a NASA project within the Aviation Safety program. The OPL at the University of Iowa is working a project entitled Operator State Sensor Investigations and Operator Feedback

Algorithms. One aspect of this project is to interpret operator workload and overall cognitive state effectively to optimize the flight deck interface.

There are several ways to characterize operator state, including electroencephalogram (EEG), electrocardiogram (ECG, heart rate), galvanic skin response (GSR), respiration rate, flight technical performance, and eye tracking to name a few. Eye tracking is appealing in flight deck operations due to its technical readiness level being higher than that of other sensors or measures of operator state. Since flying on instruments is a visually prescribed activity that is likely to be influenced by over/under-loading the pilot, eye movement behavior provides a rich data set to investigate its ability to characterize operator state.

### 3. EYE TRACKING METRICS

Generally, we speak of two types of eye movements: Fixations and Saccades with respect to attention allocation. Cf. Jacob and Karn [1], define a fixation as a single point of gaze vector within a threshold of two degrees for a minimum duration of 200ms. However, definitions do vary from user to user. Saccadic movement is simply derived based upon the definition of the fixation, by counting a saccade as the movement from one fixation to the next. Saccadic movements are measured by saccadic distance (deg) and velocity (deg/sec). Their Euclidian distance can be derived by determining the plane on which the fixation is occurring and identifying the distance between that specified location and the eye gaze origin.

Fixations are the time in which an individual processing the visual data within the foveal field (<2 deg). Since the foveal field is so limited, saccadic movements are necessary to bring to focus and process information across a person's overall field of view. For example, reading a book consists of several fixations and saccades that trace in spatial segments across the page. A pilot's eye scan behavior is similar, making fixations at specific instruments to

obtain information, performing a saccade to fixate upon the next instrument.

Further metrics may be derived from these two general eye movements. Statistical analysis of each general metric, such as average fixation duration, fixation frequency, fixation duration max and standard deviation of the fixation duration are generally conducted. Scan-path/link analysis is used to quantify saccadic movements and fixation location patterns. These statistics are then used to observe if a correlation exists between the eye tracking metrics and the independent test variables.

Eye tracking entropy is the level of randomness observed in eye movement behavior by evaluating the gaze vector X and Y coordinates. For real-time evaluation, the standard deviation of the X and Y gaze vector components are calculated over a moving time window of 30 seconds [2]. A moving window of 30 seconds was chosen in an attempt to collect enough fixations to evaluate with statistical significance when calculating the standard deviation. With this calculation, changes in entropy values indicate a change from the current scan-path and fixation trends, presumably induced by a change in the flight deck and/or operator.

#### Equation 1. Entropy Equation

$$Entropy = H = \sum p_i \log_2(1/p_i)$$

Each area of interest or fixation point is associated with a state-space probability of subject focus ( $p_i$ ). By assuming when situations are in high entropy, or high levels of randomness, the probability of looking at everything an equal number of times will transition between all areas of interest and stimuli at near equal frequencies. The state-space probability changes over time as scan-path trends change, therefore, changing the entropy value [4]. In theory, as workload increases the observed scan-path becomes less random [3, 4].

Since each display on the flight deck provides specific information utilized by the pilot in

different ways, it is important to characterize the data in two ways: General metric analysis across the flight deck, and specific analysis of the metrics within pre-defined areas of interest (AOIs). Areas of interest used in this experiment were broken down specifically to interpret data on the simulator flight deck interface. They included the Multifunction Display (MFD), Mode Control Panel (MCP), Out the Window (OTW), Computer Display Unit (CDU), and a higher resolution of the Primary Flight Display (PFD) with the standard electronic flight information system (EFIS) display configuration, including the airspeed indicator (ASI), altimeter (ALT), heading indicator (HDG), flight mode annunciator (FMA), and the attitude indicator (AI).

#### 4. APPARATUS

A pilot-in-the-loop study was conducted in the Operator Performance Laboratory's flight deck simulator that is based on the Boeing 737-800 form factor. The simulator is comprised of a flight deck with complete glass cockpit displays, five outside visual projectors, functioning mode control panel (MCP) with autopilot and auto throttle, and standard Boeing 737 flight controls. The head down display (HDD) panel was configured to represent the standard Boeing EFIS display on the PFD. The MFD displayed a moving map depicting the current flight plan and corresponding waypoints, as well as other useful information as would be found standard on a typical 737 glass cockpit.



Figure 1. OPL Flight Deck Simulator

A three camera Smarteye eye tracking system was installed in the simulator cockpit and the camera angles were optimized for the left seat only as shown in Figure 1. OPL Flight Deck Simulator The eye tracking camera configuration was optimized to obtain pilot eye gaze vectors with quality spatial resolution down to one degree and no greater than two degrees of gaze point variation for critical areas of interest as specified earlier. Cameras were mounted directly beneath the glare shield of the flight deck to minimize display obstruction.

#### 5. DESIGN OF EXPERIMENT

The experiment was designed with the intention of yielding a wide range of induced workload across pilots. A total of 12 pilots with at least a private pilot license and IFR rating were asked to fly a simulated instrument approach to runway 9R at KORD (Figure 2) with two design factors, 1). Level of automation (none, flight director, coupled), and 2). Visibility at decision height (DH) (sufficient or insufficient to land)

Level of automation varied among a fully coupled autopilot mode with automated flight controls and auto-throttle, flight director mode with auto-throttle and a flight director overlaid on the PFD, but flight control were manipulated manually by the pilot and manual approach with no automated controls requiring the pilot to control both the yoke and the throttle. Pilots were allowed to utilize only IFR low-altitude en-route charts, approach plates, the localizer glide slope and course deviation indicators as well as a moving map on the MFD. Pilots were also given an approach checklist broken down by waypoint and were required to make standard radio calls and frequency changes.

The "land or go-around" visibility at DH changed between three tenths and a single tenth of a nautical mile. This required pilots to make a decision 200 feet above touchdown height to either land or go-around upon visualization of the runway end identifier lights. This condition only changed the approach

scenario from the last waypoint to decision height.

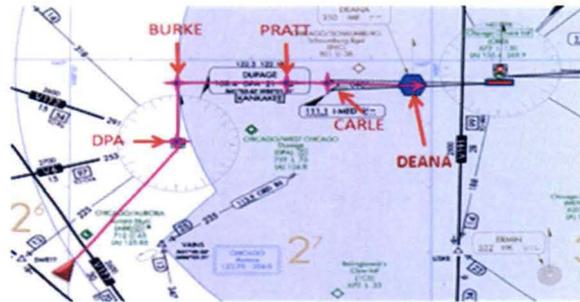


Figure 2. KORD 9R Approach Flight Plan

In flight assessment was administered at each waypoint to increase the resolution of pilot workload using the Bedford workload scale. The Bedford scale is a 1-10 workload rating assessing the current workload perceived by the pilot. Pilots were trained in pre-flight briefing on the scale's decision tree to aid in assessing their workload appropriately according to the scale definition.

The Bedford scale (figure 4) breaks down into subcategories of satisfactory workload (1-3), tolerable workload (4-6), possible to accomplish task workload (7-9), and impossible to accomplish task due to high workload (10).

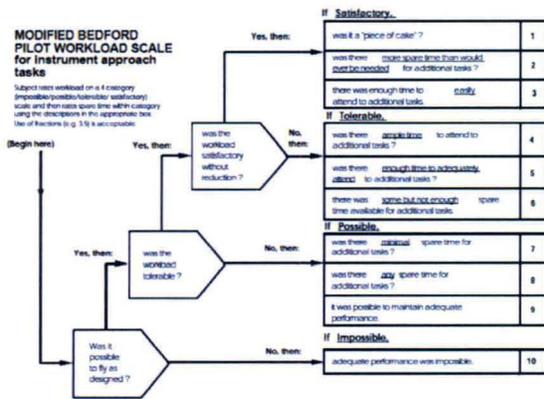


Figure 3. Bedford Workload Scale

## 6. HYPOTHESIS

Two hypotheses were generated in this experimental procedure: 1.) Workload will have a monotonically increasing relationship with increasingly manual flight conditions. 2.)

Saccade length will have a monotonically decreasing relationship with increasingly manual flight conditions. It is reasonable to consider other eye tracking metrics to be in place of saccade length for hypothesis two. This analysis simply addresses saccade length as a simple indicator of fixation dispersion, similar to that of visual entropy or nearest neighbor index analyzed across the flight deck.

The hypotheses are developed by observing typical pilot eye scan behavior in varying levels of flight automation and their respective visual demands. Pilot demand is increased by reducing the level of automation provided. The level of automation provided changes the pilots scan behavior, indicated in this analysis by a reduction in average saccade length with increasing manual flight control. This follows the logic that a pilot is required to more closely monitor the aircraft state gauges on the PFD looking elsewhere less of the time when not flying on autopilot (Full Auto), thereby decreasing the average saccade length.

## 7. DATA SET AND ANALYSIS

Data were processed and analyzed by Minitab version 14. Analysis of the subjective results crossed with the testing conditions indicated that test conditions yielded significant variance in induced workload.

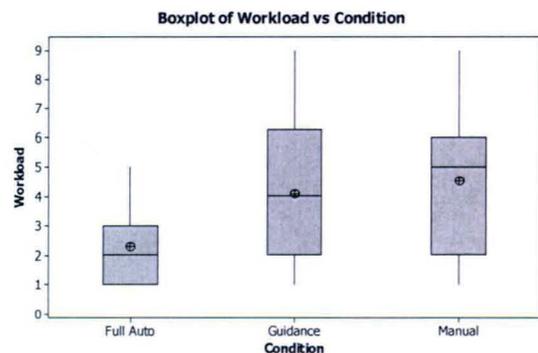


Figure 4. Workload vs. Condition

ANOVA analysis of the effect of condition on workload indicates significant variance among the automation conditions ( $F(2,136)=16.35$ ,

$p < 0.001$ ). A Tukey pair wise comparison test indicated significance between the full automation condition and the guidance condition ( $t = 4.237$ ,  $p < 0.0001$ ), and significant variance between full automation condition and manual condition ( $t = 5.404$ ,  $p < 0.0001$ ). No significant variance was found between the guidance condition and the manual condition ( $t = 1.006$ ,  $p > 0.05$ ).

Analysis of variance of the saccade length observed against automation condition yielded significant results ( $F(2,136) = 11.50$ ,  $p < 0.001$ ). Tukey pair wise comparison tests indicate significance between the full automation condition and the guidance condition ( $t = -4.732$ ,  $p < 0.001$ ) and the full automation condition and the manual condition ( $t = -2.962$ ,  $p < 0.05$ ). There was no significant variance between the Guidance and Manual conditions ( $t = 1.006$ ,  $p > 0.05$ ).

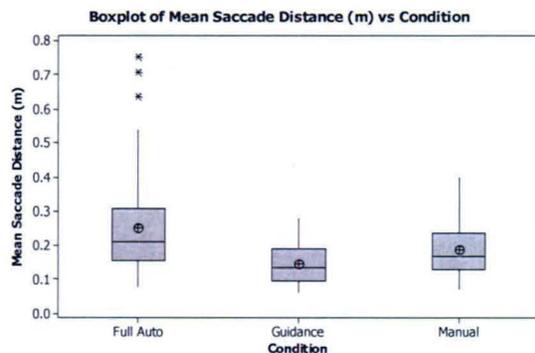


Figure 4. Saccade Length vs. Condition

## 8. DISCUSSION

The lack of variance between the Guidance and Manual conditions is explained by the layout of the standard EFIS display. The source of Guidance information is the flight director displayed on top of the AI on the PFD. A pilot operating in the Manual condition would use the AI combined with other instruments on the PFD to fly the correct flight path. This results in insignificant differences between these two conditions when average saccade length is analyzed across the entire flight deck and not limited to the PFD area of interest.

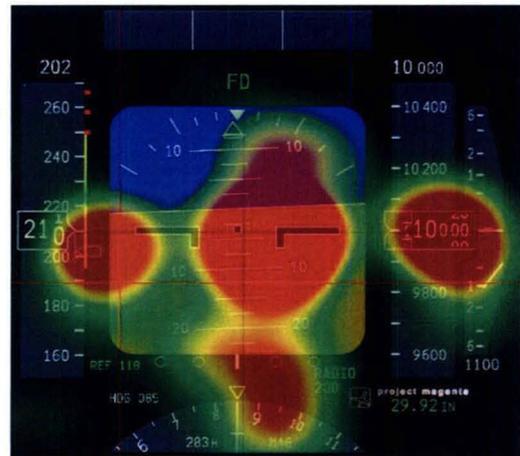


Figure 5. 737 EFIS Manual Condition Heat Map

Figure 5. 737 EFIS Manual Condition Heat Map shows a pilot's fixation heat map with no guidance (flight director) provided.

## 9. FUTURE WORK

The data set itself is a plentiful amount of eye movement behavior of pilots performing an instrument approach. Using this data as an empirical data set, the metrics can be averaged over the pilot population and used to feed into a human computer model for eye movement behavior. Research done at the OPL is performing such research that will utilize this data set. A project done in collaboration with NASA entitled Integrated Alert and Notification (IAN) (Grant: NNX08BA01A) will develop a digital human model capable of interacting in a digital simulation to identify optimal configuration within the flight deck.

This work is presented as a separate paper in this conference (Cover & Schnell [5])

Stemming from this research is the development of a software tool capable of receiving data and characterizing pilot workload in real time and through use of multi-channel inputs, such as EEG, EKG, Respiration rate sensors or any other types of sensors research proves useful in characterizing human workload. Cognitive Avionics Tool Set (CATS) [6] software is currently in development by OPL. Combination analysis tool and real time

classifier is a useful graphical user interface for post processing and analysis.

Future analysis of the raw data will include other metrics such as visual entropy [4], nearest neighbor indexing [3], scan path indexing [7] and fixation mapping [8] to further look into what derivations of saccadic eye movement behavior yield trends that substantially correlate to pilot workload.

Further use of the data set as an empirical data source continues on various projects with the OPL and NASA. One intention is to use the data to drive an eye movement behavior model for future flight deck human performance simulations.

#### 10. ACKNOWLEDGEMENTS

We would like to acknowledge the National Aeronautics and Space Administration under project NNLO7AA00A and their team at NASA Langley for their insight and support. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of NASA, the U.S. Government, or any other organization.

#### 11. REFERENCES

[1] Jacob, R. J., & Karn, K. S. (2003). Eye tracking in human-computer interaction and usability research: Ready to deliver the promises. *The Mind's eye: Cognitive The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research*, 573 - 603.

[2] Ellis, K., & Schnell, T. (2009). *Eye Tracking Metrics For Workload Estimation In Flight Deck Operations*. Iowa City, IA: University of Iowa.

[3] Di Nocera, F., Terenzi, M., & Camilli, M. (2006). Another look at scanpath: distance to nearest neighbor as a measure of mental

workload. In D. de Waard, K. Brookhuis, & A. Toffetti, *Developments in Human Factors in Transportation, Design, and Evaluation* (pp. 1-9). Maastricht, the Netherlands: Shaker Publishing.

[4] Gilland, J. (2008). *Driving, Eye Tracking and Visual Entropy: Exploration of Age and Task Effects*. University of South Dakota.

[5] Cover, M., & Schnell, T. (2009). *Modeling Pilot Behavior for Assessing Integrated Alerting and Notification Systems on Flight Decks*. Iowa City, IA: University of Iowa

[6] Schnell, T., Keller, M., & Macuda, T. (2007). Pilot State Classification and Mitigation in a Fixed and Rotary Wing Platform. *Aerospace Medical Association annual conference* (p. 377). New Orleans, LA: Aviation Space and Environmental Medicine.

[7] Simon, P., Rousseau, F., & Angue, J.-C. (1993). *Quantitative Analysis of Mental Workload Influence on Eye Scanning Movements*. Le Touquet, France: Systems, Man and Cybernetics.

[8] Wooding, D. S. (2002). *Fixation Maps: Quantifying Eye-Movement Traces*. New Orleans: ETRA.

# A Toolset for Supporting Iterative Human – Automation Interaction in Design

Michael Feary, PhD

NASA Ames Research Center

Michael.S.Feary@nasa.gov

## ABSTRACT

The addition of automation has greatly extended humans' capability to accomplish tasks, including those that are difficult, complex and safety critical. The majority of Human - Automation Interaction (HAI) results in more efficient and safe operations, however certain unexpected automation behaviors, or "automation surprises" can be frustrating and, in certain safety critical operations (e.g. transportation, manufacturing control, medicine), may result in injuries or the loss of life. (Mellor, 1994; Leveson, 1995; FAA, 1995; BASI, 1998; Sheridan, 2002). This paper describes the development of a design tool that enables on the rapid development and evaluation of automation prototypes. The ultimate goal of the work is to provide a design platform upon which automation surprise vulnerability analyses can be integrated.

## Introduction

Recent analyses of aircraft accidents (FAA, 1995; BASI, 1998) have shown that aircraft automation is increasing as the major contributing factor to aircraft incidents and accidents. These accidents have shown a disturbing trend in that the automation was performing as designed, and was operated by well-trained operators, but users were surprised with unexpected automation behavior.

These "automation surprise vulnerabilities" are due to a failure in the specification of the behavior of the automation, rather than a failure in the implementation of the automation. The vulnerabilities could be due to a number of possible factors, including: inadequate coverage of the possible situations the automation needs to be able to respond to, or a weakness in the presentation of the automation behavior, such that the human user and the automation do not share a common understanding of the goals, the situation, or the proper behavior to accomplish a goal for a given situation. In either case, the focus needs to be on presenting human operators with predictable automation behavior.

The ultimate goal of the research described in this paper is the development of a viable means of identifying Human-Automation Interaction (HAI) vulnerabilities early in the design process. The focus

for these HAI analyses is on the "cognitive" behavioral aspects of the user and the software or digital hardware in computers. The analyses aim to identify vulnerabilities in the communication of behavioral expectations or intent between the user and the automation.

The Automation Design and Evaluation Prototyping Toolset (ADEPT) was developed to respond to this need, and to focus on the iterative specification of decision logic of the automation being designed. The tool is intended to produce an accurate and complete specification. In addition to the focus on specifying decision logic, the tool was intended to provide a platform for integrating HAI testing and analysis.

The focus of this paper is an examination of the suitability of ADEPT to serve as a platform to upon which to integrate HAI analyses. ADEPT was developed to be usable by a domain expert designer without requiring extensive programming language expertise. This requirement was intended to enable ADEPT to be used early in the design process, by many different design team members (e.g. training, procedure, interface, etc.) The tool should foster communication between design experts from different domains, meaning that the tool should provide a structure that provides specific transition points for design team members to interact with each other.

ADEPT combines a graphical user interface design capability with an automation behavior specification capability and an automatic code generator to enable domain expert designers to create testable software prototypes.

- The User Interface Editor enables the designer to specify the look and feel of the of user interface by placing graphic objects on a canvas. The graphic objects include buttons, knobs, displays, and the ability to import static and dynamic graphical objects created in other software applications. The

properties (i.e. font, size, color, etc) of the graphical objects in can be changed in a property

browser for the User Interface Editor, or can be dragged into the Logic Editor to allow graphic properties to be changed dynamically corresponding to the automation behavior.

- The Logic Editor enables the designer to specify the decision logic and automation behavior of the device, the environment in which the device operates, as well as the behavior of the user interface objects on the user-interface corresponding to the reflect the current state of the device and environment.

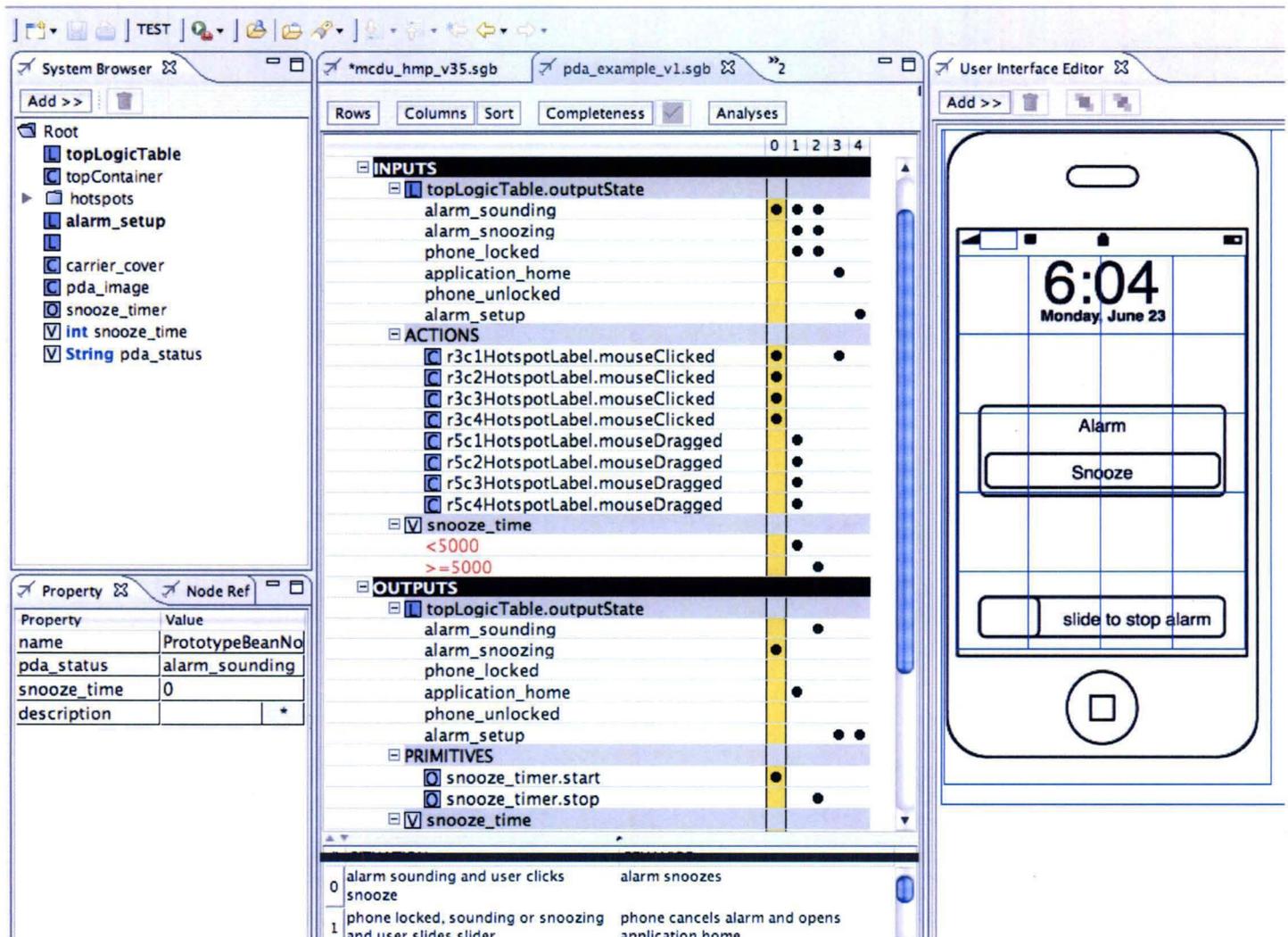


Figure 1. ADEPT in Build mode

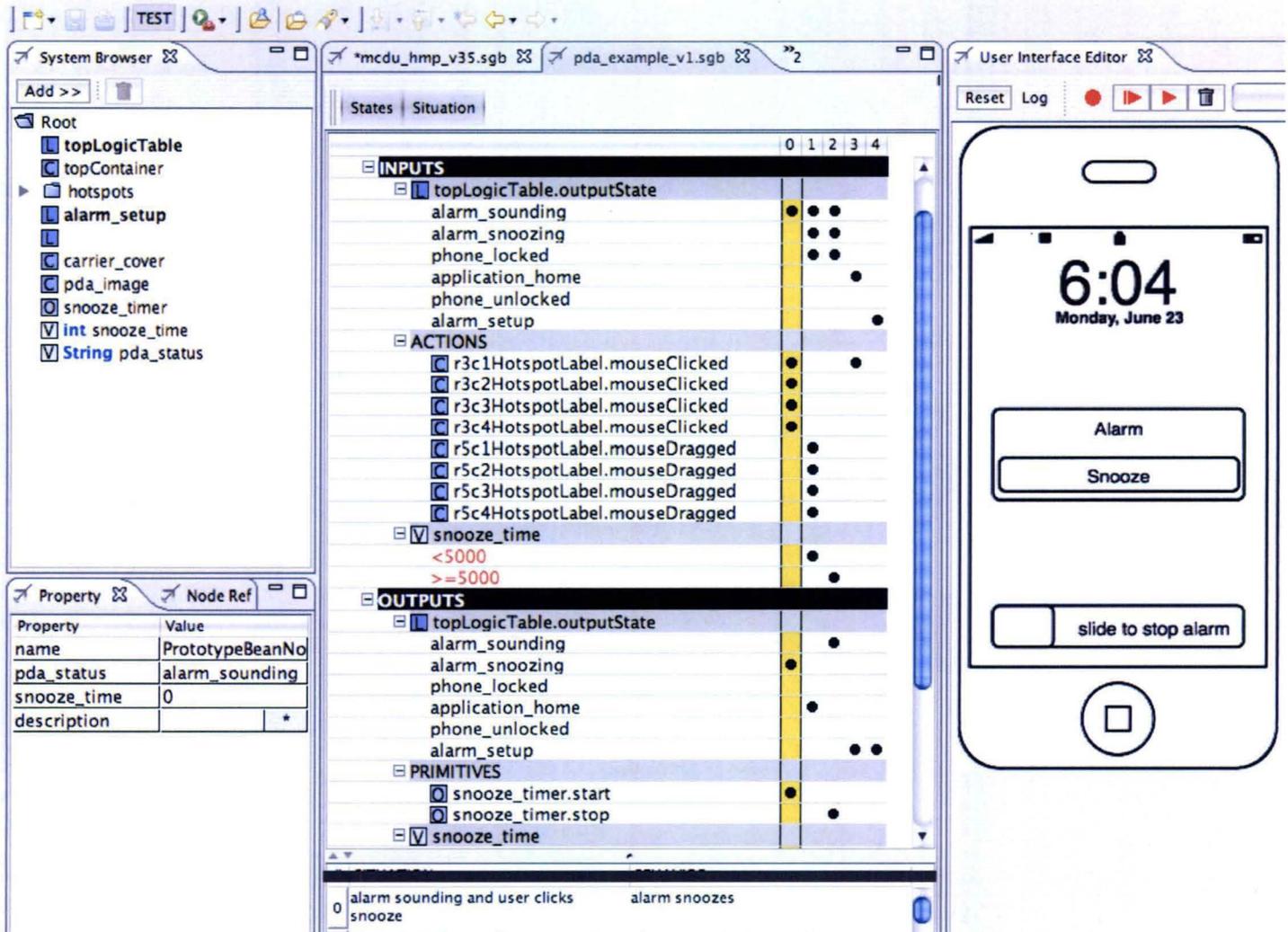


Figure 2. ADEPT in Test mode

### Iterative Build and Test

ADEPT works in two modes, *Build* and *Test*. In *Build* mode the designer creates graphic objects to the User Interface editor, and adds these as well as system objects (e.g. sensor inputs) to the Logic Editor. The designer then uses these objects to construct the logic table. By testing each column as it is added, the designer can start with very simple behavior and iteratively add complexity.

### The User Interface Editor

The User Interface (UI) Editor provides the tools to allow the designer to construct the interface. The User Interface Editor Design Mode Menu is shown at the top right of figure 1. The menu allows UI objects to be added, deleted and arranged.

Figure 1 also shows an example interface constructed in the User Interface Editor with transparent objects on top of the image (shown with blue outlines in figure 1) to create the functionality. The interface could be made to

look even more realistic by importing higher quality images built in other graphical applications.

### The Logic Editor

The Logic Editor is what differentiates ADEPT from a graphics design application. The Logic Editor, derived from the Operational Procedure Table (OPT) method (Sherry, 1996) allows the designer to specify the behavior of the device, and/or the device interface built in the UI Editor.

ADEPT uses a tabular representation of a finite state machine. In contrast to typical state transition tables, the representation used by ADEPT focuses more on presenting information about the situation (input combination) - automation behavior (output combination), and less on presenting information about state transition in a summarized form information. This focus allows a more compact notation, which enables the designer to see more behaviors, making it easier to make a complete specification.

The primary method for building a Logic Table is to select an object in the Object Browser and Drag and Drop it into the table as an input or output. This works for adding inputs and outputs, but it also works for adding variables and other objects as input conditions and output functions.

The table consists of a listing of Inputs and Outputs on the Y-axis, and columns of situation-automation behavior pairs along the X-axis, as shown in figure 2. Figure 2 shows that a black separator bar denotes the Inputs and Output Fields. The Input bar can be translated as an "IF" statement, while the Output bar is read as a "THEN" statement. Between the Inputs and Outputs bars, the thick gray lines between represent "AND" statements, and thin gray lines represent "OR" statements. Note that the outputs only contain "ANDs". The thin gray lines are only used to make the table easier to read.

The pda example shown in figures 1 and 2 can illustrate how the tables are used. In this example, there are two ways to silence the alarm. First, the user can press the snooze button (shown in column 0), or the user can unlock the pda (and go to the alarm page to stop the alarm, which isn't shown in this example).

Examining through the table, column 0 says:

**IF**  
The alarm is sounding  
**AND**  
Any of the row three interface areas (i.e. the snooze button for this page) is clicked  
**THEN**  
The pda status will change to alarm\_snoozing

Similarly, column 1 is read as:

**IF**  
The alarm is sounding  
**OR**  
Snoozing  
**OR**  
Locked  
**AND**  
The row five interface areas (i.e. the lock slider button for this page) is clicked and dragged  
**AND**  
The snooze time is less than 300000 milliseconds (i.e. 5 minutes)  
**THEN**  
Go to the application page

Figure 2 also illustrates how ADEPT can be used to design iteratively. Once the inputs and outputs have been defined, the tabular representation enables the designer to add and test situation-behavior pairs individually using the automatic code generator described in the next section.

## Evaluating Prototypes built in ADEPT

A number of features have been incorporated into ADEPT to aid the designer in evaluating a device and its interface behavior. The automatic code generator creates an executable specification enabling rapid build and test cycles. Figure 2 shows the different functions available in the Test mode of the UI Editor, in contrast to the menu available in design mode, shown in figure 1. The menu shows the buttons for the Reset function, the Log function, and the Scenario Management function, which includes the Record, Reset - Play, Play, and Delete buttons and the Configuration menu.

The Log function is used to begin to record all user actions and all automation behaviors of the device prototype. The Log function generates two files at the moment, one of which is used for traditional usability evaluation and the other is used as a data source for computational human performance models.

The Scenario Management utility consists of the ability to record, playback and delete various configurations that is useful for evaluating the device against different tasks. Pressing the Record button once records all of the user actions and device information. Pressing the Reset-Play button first resets then plays the configuration selected on the configuration menu, while pressing play configures the prototype starting from the existing configuration.

## Method

Three case studies were conducted to test the usability of the ADEPT software. The case studies examined three participants using the ADEPT to design actual prototypes. As the case studies examined the use of the tool across different applications with varying complexity of design, traditional performance metrics (e.g. time, errors, etc.) were not applicable. Therefore descriptive and qualitative measures were used, consisting of complexity metrics and questions about the usability and usefulness of ADEPT.

The four questions consisted of:

- 1: Can designers build testable prototypes in ADEPT?
- 2: Does ADEPT support rapid iteration and modification?
- 3: Does ADEPT focus the design activity on precise and complete specification of the automation behavior?
- 4: Does ADEPT support communication with other design team members?

The three case studies involved the use of ADEPT over a period of between one and six months, and the information gained during these time periods would not have been adequately captured through the use of interview or questionnaire techniques.

## Results

Table 1 provides an illustration of the size and scale of the different projects shown in columns corresponding to each participant (P1, P2, and P3).

**Table 1. Complexity Metrics for the 3 case studies**

Metrics/Questions	P1	P2	P3
Source Lines of Code	1500	20000	4300
Automation behaviors	13	560	13
GUI objects	18	160	100

Table 2 shows the responses to the four questions of the 3 case studies.

**Table 2. Question responses for each case study**

Questions			
1: Construction?	Yes	Yes	Yes
2: Rapid Iteration?	Yes	Yes	Yes
3: Completeness?	Yes	Yes	Yes
4: Communication?	N/A	Yes, needs improvement	Yes, Needs improvement

The results show that tool did enable all of the designers to build prototypes that suited their purposes. This by itself is a notable success for the designers without programming expertise. The designers reported that they felt the tool supported rapid iteration in the design process, a key component of good design. The participants also reported that they felt ADEPT helped them to build more precise and complete specifications of automation behavior, however they felt that some work was needed to make the prototypes they designed understandable to others in their design group.

All of the participants expressed some displeasure with the organization of the hierarchy, and the means of using variables to transfer behavior information in design projects with multiple tables. One of the primary objectives of the tool is the facilitation of communication between design team members.

The evaluation was only intended to validate that domain expert designers can use the proof of concept version of ADEPT to construct testable prototypes, however the case studies served an additional purpose beyond simple validation. Valuable lessons were learned from the length of the case studies and the wide range of expertise of the three participants.

## Discussion

Given the constraints with evaluating new design tools the case studies provide an example of the strengths and weaknesses of ADEPT. Although none of the case studies involved the use of a complete version of ADEPT in a real-world design process, the case studies did test different portions of the tool in real-world design problems. The results of the design exercises and the impressions of the users were positive enough to validate the initial proof-of-concept version of the tool.

This resolves the first development challenge, as domain expert designers can use ADEPT to design testable prototypes without extensive programming expertise or training. The responses from the case study participants indicated that they were able to focus on domain goals and objectives for the devices they were constructing, which was defined as the primary obstacle in software design Curtis et al. (1988).

The case studies have shown that while interpretation of individual tables by novice designers is achievable with the tabular representation, the current organization of multiple tables or representation of multiple tables in a project may obscure the understanding of complete behavior specification of a device.

This has been modified in subsequent versions of ADEPT with the creation of a "Logic Table" object, and the replacement of the action – behavior – feedback table hierarchy with only one "Top Logic Table". In this way individual designers can tailor the organization of multiple tables to suit their needs by connecting the different tables with the Logic Table objects. Additionally, new visualization techniques are being explored.

The case study evaluations also revealed a need for the creation of a library of objects to ease the construction of devices. This is especially true of complex projects where an architecture template can speed the initial construction of the device. This need will be addressed over time as ADEPT gains exposure.

## Conclusions and Future Work

The initial results from the case studies have shown that ADEPT is usable by domain expert designers without requiring extensive programming expertise. While further development work is needed, these results show that ADEPT is suitable as a platform upon which to integrate HAI analyses.

In addition, the case studies showed that an ADEPT-like tool could help fill a niche. By creating a lower fidelity, but still testable prototype in less time with fewer resources, more iteration is possible, which can improve the design process (Gould and Lewis, 1985; Poltrock and Grudin, 1996).

A challenge for the future is the integration of task information. A decision was made to focus on the specification of automation behavior, and leave task information specification for future versions. Rasmussen (1994) Hoffman et al. (2002), and Feltovich et al. (2004) have expressed the need for greater involvement of domain experts in the design process, and the case studies have shown how ADEPT can facilitate greater involvement. A plan to add integrated Human-Automation interaction analyses, should begin to address this need, however the addition of a usable means for integrating or importing the results of other task decomposition or task analysis tools is an idea that deserves future research.

### Acknowledgements

Thank you for technical assistance from Lance Sherry, (George Mason University), Arash Aghevli, and Eugene Turkov, (Singer – Gaffarian Technologies) and Rohit Deshmukh, (San Jose State University).

### REFERENCES

- BASI (1999) Advanced Technology Aircraft Safety Survey Report. *Flight Safety Digest Special Issue*. Flight Safety Foundation June – August, 1999 pages 137 – 216. Available at <http://www.basi.gov.au/sprog/advtek.htm> (9/00)
- Curtis, B., Krasner, H., and Iscoe, N. (1988). A Field Study of the Software Design Process for Large Systems, *Communications of the ACM*, 31(11), ACM, p. 1268-1287.
- Federal Aviation Administration (1996) *The interface between flightcrews and modern flight deck system*. FAA human factors team. Federal Aviation Administration., Washington, DC, Available at [http://www.faa.gov/education\\_research/training/aqp/library/media/interfac.pdf](http://www.faa.gov/education_research/training/aqp/library/media/interfac.pdf)
- Feltovich, P. J., Hoffman, R. R., Woods, D., and Roesler, A. (2004). Keeping It Simple: How the Reductive Tendency Affects Cognitive Engineering, *IEEE Intelligent Systems*, May-June, IEEE Computer Society Publications Office, Los Alamitos, CA, p.90-95.
- Gould, J. D. and Lewis, C. (1985) Designing for Usability: Key Principles and What Designers Think, *Communications of the ACM*, 28, (3), ACM Press, New York, NY, p. 300-311.
- Hoffman, R. R., Klein, G., and Laughery, K. R. (2002) The State of Cognitive Systems Engineering. *IEEE Intelligent Systems*, January-February, IEEE Computer Society Publications Office, Los Alamitos, CA, p.73-75.
- Leveson, N. (1995) *Safeware: System Safety and Computers*. Addison-Wesley, New York, NY, USA.
- Mellor, P. (1994) CAD: Computer-Aided Disaster, *High Integrity Systems*, 1(2), p.101-156.
- Poltrock, S. E. and Grudin, J. (1996). Organizational Obstacles to Interface Design and Development: Two Participant Observer Studies, In *Human Computer Interface Design* Rudisill, M., Lewis, C., Polson, P. G., and McKay, T. D. (Eds.), Morgan Kaufmann Publishers Inc., San Mateo, CA. P. 303-337.
- Rasmussen, J. (1994) *Cognitive Systems Engineering*. John Wiley and Sons, New York, NY.
- Sheridan, T. (2002) *Humans and Automation: System Design and Research Issues*, John Wiley & Sons, Inc., Santa Monica, CA.
- Sherry, L. (1996) Personal communication regarding the use of the Operational Procedure methodology in the design of Honeywell product development.

# Use of Dynamic Models and Operational Architecture to Solve Complex Navy Staffing Challenges

Darby Grande\*, J. Todd Black\*\*, Jared Freeman\*, Tim Sorber\*\*, Daniel Serfaty\*

\* *Aptima, Inc., dgrande@aptima.com, jfreeman@aptima.com, serfaty@aptima.com*

\*\**Klett Consulting Group Inc., todd.black@kcg-inc.net, tim.sorber@kcg-inc.net*

**Abstract.** The United States Navy established 8 Maritime Operations Centers (MOC) to enhance the command and control of forces at the operational level of warfare. Each MOC is a headquarters manned by qualified joint operational-level staffs, and enabled by globally interoperable C4I systems. To assess and refine MOC staffing, equipment, and schedules, a dynamic software model was developed. The model leverages pre-existing operational process architecture, joint military task lists that define activities and their precedence relations, as well as Navy documents that specify manning and roles per activity. The software model serves as a “computational wind-tunnel” in which to test a MOC on a mission, and to refine its structure, staffing, processes, and schedules. More generally, the model supports resource allocation decisions concerning Doctrine, Organization, Training, Material, Leadership, Personnel and Facilities (DOTMLPF) at MOCs around the world. A rapid prototype effort efficiently produced this software in less than five months, using an integrated process team consisting of MOC military and civilian staff, modeling experts, and software developers. The work reported here was conducted for Commander, United States Fleet Forces Command in Norfolk, Virginia, code N5-OLW (Operational Level of War) that facilitates the identification, consolidation, and prioritization of MOC capabilities requirements, and implementation and delivery of MOC solutions.

## 1. INTRODUCTION

The Navy developed the Maritime Operations Center (MOC) concept to enhance its command and control of forces at the operational level of warfare [1]. To oversee the development of the MOC concept, the Navy gave United States Fleet Forces Command (USFFC) the responsibility to standardize MOC staff functions and processes. This standardization will enable interoperability with the joint community and promote commonality across all Fleet and principal headquarters.

USFFC code N5-OLW (Operational Level of Warfare) used the Department of Defense Architecture Framework (DoDAF) to develop Business Process Models (BPM) for MOC processes. These BPMs, called Operational Views (OV-6c) in DoDAF [2], define MOC processes, their sequence, the organizational elements that execute them, and the products of those work activities. The dynamic modeling work reported in this paper transformed the static DoDAF documents into an executable software model called the MOC Performance Assessment Tool (MOC-PAT). The MOC-PAT is designed to support decisions regarding MOC staffing, such as whether a staffing plan is adequate to execute the many MOC processes required to support a specific mission set at a specified operational tempo. The first application of this tool supports planning and execution of Navy exercises to accredit Fleet MOCs.

This paper outlines how a multi-disciplinary team developed an innovative solution for the Navy

leveraging existing architecture products and software modeling approaches. Section 2 of this paper defines the problem. Section 3, describes the technical development of the initial version of the MOC-PAT. This is followed by a discussion of the data used to exercise the model and a presentation of initial results in Section 4. Finally, in Section 5 presents our conclusions and the directions for our future work.

## 2. PROBLEM DEFINITION

The MOC concept is a recent development in the Navy. In order to ensure MOCs meet mission objectives for Fleet and Combatant commanders while implementing necessary interoperability standards, USFFC tasked Commander Second Fleet to establish a MOC Project Team to explore and document MOC doctrine, organization, training, material, leadership, personnel and facilities. As this effort evolved and the MOC Project Team transferred to USFFC as code N5-OLW, it was evident that a means of linking mission tasking to MOC manning and performance was needed to ensure MOCs are staffed and equipped to mission requirements.

USFFC N5-OLW developed BPMs of over 30 MOC processes, documenting hundreds of activities within each process. These BPMs were created using the DoDAF standard OV-6c format. Typically, these diagrams are developed to support acquisition decisions and reside in a central Navy architecture repository, the Syscom Architecture Development and Integration Environment (SADIE). The MOC-PAT leverages

these preexisting BPM documents and uses them to develop accurate models of the operating MOC.

This is accomplished by linking MOC processes back to Joint Mission Essential Tasks Lists (JMETL), first identifying core missions a MOC staff is required to execute and then relating those mission tasks to the associated JMETL tasks. Manning information based on existing MOC manning documents and role data (developed from surveys and onsite observation) is combined with process activity workload observations (i.e., time to complete activities, or work products required to complete activities) to populate the OV-6c BPM documents in the MOC-PAT. These data are then available to support model runs to analyze MOC staff execution and support accreditation events.

### 3. MODEL DEVELOPMENT

The initial version of the MOC-PAT is designed to enable skilled analysts at USFFC N5-OLW to test the impact of MOC manning estimates on MOC performance at the Numbered Fleets executing Normal & Routine (N&R) Missions. In this section we introduce the model, its assumptions, the dynamics, and the output capability for the users.

#### 3.1 Model Introduction

A mission scenario that the user constructs contains a number of *processes*, each of which is made up of a series of *activities*. While the processes are executed, a fixed schedule of *battle rhythm events (BRE)* occurs. It includes special working group meetings and regular briefs to senior staff. The BRE and the activities produce and consume (that is, require) *information products*, and these well-defined products serve as the linkages between different parts of the organization and their many processes and BRE. For example, a planning activity may produce a plan (a document) that is a required input to an assessment activity to communicate which indicators of progress should be monitored. The MOC organization that will accomplish this mission is made up of multiple *organizational units (OU)*. Each OU has several *billets* (individuals) assigned to it, and each billet is assigned a collection of *roles* he may take on, one at a time throughout the mission. These roles currently serve as proxies for more detailed information about billets' associated knowledge and skills, which we hope to incorporate in future versions of the MOC-PAT.

The work discussed in this paper was conducted for an initial proof-of-concept phase, so a number of simplifying assumptions were necessary. As the work continues, we are re-visiting each of these to refine and enhance the model. We assume:

- Billets are available to work 24 hours each day
- The MOC is operating under Normal and Routine conditions
- Each process begins at scheduled times, according to user-specified cycles
- Each activity cannot begin until its preceding activities (within the process) are concluded
- If an activity is prompted to start at time  $t$  (by the schedule or by the conclusion of its preceding activities), then it must conclude at a deadline created by adding the longest required processing time by any of its roles to this earliest triggered start time
- Information products have a user-specified shelf life, after which their level of completion decays. This is to ensure that we capture the fact that an activity which is unable to update or produce an information product on time will affect the ability of an activity which required the information product to execute completely.

#### 3.2 Model Dynamics

The purpose of this effort is to help the Navy determine whether the MOCs as envisioned and instantiated are meeting the mission support and interoperability goals. This specific, evaluation goal led us to implement our model of the MOC in a simulation, rather than pursue optimization of the many variables – staff size, schedule, process step configuration, communication strategies, etc. With this simulation, the MOC-expert user is able to configure the particular mission he would like the simulation to “play,” and the MOC organization is then evaluated against this mission scenario.

More specifically, the model enables analysts to answer several questions about MOC activities:

- **The Activities:** Do activities get the resources they need? Which processes & activities began with incomplete resources: human, information, time? Which activities could not begin at all?
- **The Organization:** Do we have enough staff in the right roles? Which organizational elements & staff were overloaded? Which were under-loaded?
- **The Information Products:** Are the information products complete and current when they are needed? What information was incomplete or missing when it was needed?

Analysts answer these questions in a process that consists of four stages: (1) Populate a database, (2) Configure the data and the model that processes them, (3) Run a mission simulation, and (4) Analyze the results. The analyst then

typically returns to step (2), to refine the configuration and continue analysis iteratively.

Data entry and configuration is conducted using a component, called Adaptive Modeling Environment, that imports data specifying mission activities (tasks), activity information requirements, activity schedules, human resources (number and roles of staff), and organizational structure. The AME provides users with standard lists and graph representations of these data, through which the user can add, delete, or edit most data objects.

The mission simulation, designed collaboratively by the USFFC N5-OLW and the development team, is a discrete-event simulation engine. It drives the assignment and execution of the processes over the course of a mission. This simulation operates as follows.

For the purpose of the model, let

$d_{ij}$  = the amount of time role  $i$  is required to spend on activity  $j$ . (If the role is not required for the activity, then  $d_{ij} = 0$ );

$$d_{\max} = \max_{i=1 \dots R} (d_{ij});$$

$c_m^t$  = the completeness of information product  $m$  at time  $t$ ;

$v_i$  = calculated completeness percentage attainable for the current execution of activity  $i$ ;

$v_i'$  = completeness percentage attained in the most recent execution of activity  $i$ ;

$\alpha$  = activity repair coefficient, that is, the rate at which deficient information products input to an activity are improved by that activity;

$\beta_i$  = minimum completeness threshold for activity  $i$ ;

$\tau_i$  = minimum execution time for activity  $i$ ;

$\sigma$  = completeness decay rate for activities;

$w_1, w_2, w_3$  = the weights used in calculating  $v_i$  to balance the importance of preceding activity completeness, information product input completeness, and fulfillment of roles required,

$$\text{such that } \sum_{k=1}^3 w_k = 1;$$

$N_p$  = the number of activities in process  $p$ ;

$M$  = the total number of information product types.

Additionally, we employ the following variables:

$$r_{ij} = \begin{cases} 1 & \text{if role } i \text{ is required for activity } j \\ 0 & \text{otherwise} \end{cases}$$

$$x_{kj}^t = \begin{cases} 1 & \text{if billet } k \text{ is assigned to activity } j \text{ at time } t \\ 0 & \text{otherwise} \end{cases}$$

$$a_{mj} = \begin{cases} 1 & \text{if info. prod. } m \text{ is an input to activity } j \\ 0 & \text{otherwise} \end{cases}$$

$$l_{ij} = \begin{cases} 1 & \text{if activity } i \text{ directly precedes activity } j \\ 0 & \text{otherwise} \end{cases}$$

Each time an activity within a process is prompted to begin (either by the process schedule, or by the completion of all the preceding activities), the activity's potential completion score is computed to determine whether the activity has available the resources it needs: the required roles among available staff members; recently updated information products required by the activity; and required preceding activities. The score calculated for activity  $i$  at time  $t$  is

$$v_i = \frac{w_1}{\sum_{h=1}^{N_p} l_{hi}} \left( \sum_{h=1}^{N_p} l_{hi} * v_h' \right) + \frac{w_2}{\sum_{m=1}^M a_{mi}} \left( \sum_{m=1}^M a_{mi} * c_m^t \right) + \frac{w_3}{\sum_{n=1}^R r_{ni}} \left( \sum_{b=1}^B x_{bi}^t \right)$$

The activity begins immediately if  $v_i \geq \beta_i$ , that is, if the score is above the minimum completeness threshold. (The user can define this threshold differently for each activity to reflect varying priorities for the resources). If the score is not sufficient (the activity does not have enough of the required resources available), the activity will delay its start. The required completion deadline for the activity remains fixed, so any delay in the activity start time reduces the overall duration of activity execution. As the duration of the activity is reduced, the overall quality of the actions, communications, and products of an activity declines.

At each time interval after the initial time at which the activity was prompted to begin, the activity's score is recalculated: increased with the possible addition of any newly available resources, and decreased by the decay rate due to the shorter time for execution and any resources that have become unavailable during the delay. That is, the new score is computed after a starting delay,  $\delta$ , as  $v_i(\delta) = v_i - ((1 + \sigma)\delta)$ . If  $v_i(\delta) \geq \beta_i$ , then the activity may begin. Otherwise, the delay is continued until (1) the activity is able to begin, or (2) the delay has lasted too long ( $d_{\max} - \delta < \tau$ ) at which time the activity fails.

The overall quality of activity is measured by the activity's completeness score, which conveys to subsequent activities thus propagating the effects of shortages of input resources and time. The staff of subsequent activities can partially repair the deficiencies of prior activities, and this is represented by a multiplier on incomplete input we call the repair rate, whose effect grows with the

actual duration of the task<sup>1</sup>. This repair rate is employed to calculate the concluded activity's completeness:  $v_i = \min(\alpha * v_i, 1)$ . The calculations given for activities throughout this section are used similarly to compute completeness percentages for the BRE.

### 3.3 Model Output

The output of the simulation consists of several measures, which are presented graphically within the software tool to help the analyst rapidly diagnose deficiencies in the staffing plan and mission schedule, and to refine their configuration. These measures are:

- **Activity Completeness:** For each activity that is executed in the mission, we calculate its completeness as the weighted sum of the states of its required inputs at the start of the activity, augmented by a 25% repair rate. The Input Weights are configurable by the user for each activity in order to capture variations in requirements across the three input categories: required information products, required roles, and required completion of prior activities.
- **Manning Employment:** As a mission simulation evolves, organizations dedicate staff (billets) in suitable roles (specific knowledge and skill packages) to activities. Each staff member takes on one of potentially many roles at a time. For each organization element, we return the percent employment (0 – 100%) of its staff over time. For each role, we return over time the percentages of all the billets capable of fulfilling the role that are currently employed in the role. Finally, for each billet, we return the instantaneous and average workload over the course of the mission. Instantaneous workload is currently dichotomous, as the billet is either employed or is idle.
- **Information Product Completeness:** Each information product has a shelf life that is configurable by the user. Each time an information product is updated by an activity or battle rhythm event, its completeness returns to 100% and remains there for the duration of the shelf life. After this time, the completeness of the information product decays as the information becomes increasingly outdated. For each information product, we return its completeness measure

each time that it was required as input by an activity or a battle rhythm event.

The output of the model has thus far proved accurate and useful when compared to actual MOC staff process execution as observed by USFFC N5-OLW Subject Matter Experts, and during an initial application to a MOC accreditation exercise, as discussed below.

## 4. ACCREDITATION DATA AND RESULTS

In 2008, the Chief of Naval Operations mandated that each MOC be accredited to validate its proficiency at MOC core tasks. The MOC-PAT is used to support this process by analyzing the performance of selected MOCs during accreditation. MOC accreditation is accomplished by USFFC via on-site observation of the MOC staff during a “stressing” event such as a major military exercise. These exercises can span weeks and involve hundreds of MOC staff members exercising a complex combination of the processes based on an assigned mission. The accreditation team must place its few observers where and when stress is likely to show its effects, and conduct analyses that help the MOC refine its staffing, schedule, and processes. In the section below, we discuss the types and sources of the data used for the initial MOC-PAT demonstration and evaluation, and present our initial findings based on these data.

### 4.1 MOC Data Types and Sources

Because the emphasis of this work was to develop a model that could be in use by the end of its initial six-month development period, populating the model with operationally-relevant data was of vital importance. The data required to run the model are billet information, which can be imported from existing command manning documents or manually input through the MOC-PAT configuration interface; role information, which specify the jobs or roles needed to accomplish activities (note that multiple roles can be assigned to individual billets); process diagrams imported from the approved OV-6c diagrams; the “Battle Rhythm”, or daily schedule of leadership meetings and roles of attendees; and the information products that each activity in the process model requires and creates. The data used in the MOC-PAT originate from authoritative sources: billet information from Activity Manning Documents; role information from on-site observation, as well as survey results and workshop interviews; process diagrams from the SADIE architecture repository; and the Battle Rhythm from the MOC's schedule. Additionally, an analysis of the assigned mission is conducted, and mission specific tasks from the Universal Joint Task List (UJTL) are identified. The analyst selects these mission tasks in the software

---

<sup>1</sup> By design, none of the algorithms implemented in MOC-PAT are stochastic in nature at this time; that is, none injects variance into the dataset.

configuration editor, and the MOC PAT then automatically identifies the processes to run based on a mapping by USFFC N5-OLW of tasks to processes. After this initial data import and input, the analyst can generate additional configurations easily in the model, and specify the length of a given mission to test the durability and reliability of an organizational configuration.

The software typically runs each modeled mission simulation in less than a minute, allowing users to rapidly assess and reconfigure the organization as required. Each simulation run produces graphs illustrating workload on staff, process execution success, and the availability of information products to subsequent processes during the simulation. Analysts use these outputs to assess effectiveness of an organizational configuration, to diagnose potential failures, and to specify solutions.

#### **4.2 MOC-PAT Initial Outcomes**

The MOC-PAT was tested during a major Fleet exercise in the spring of 2009. Initial testing indicated that the MOC-PAT results are consistent with observed outcomes in the MOC when reliable data are used and processes in the model adjusted to reflect how the MOC staff conducts its mission tasking.

During the spring 2009 exercise, the MOC-PAT identified several areas of interest that were not noted during on-site observation. These findings were discovered during the exercise, because reconfiguring and running the MOC-PAT was so rapid. The findings helped focus the efforts and attention of on-site observers, and allowed identification of how the MOC staff had spontaneously developed workarounds for some issues. These discoveries were documented as "best practices" to share with other MOC staffs. Observers confirmed other problem areas identified in model runs during on-site observation. These discoveries provided confidence that the model was accurately describing how a MOC staff coped with an assigned mission set. The MOC-PAT was also used to explore how process synchronization and staffing issues might evolve over time, by running the model for missions sets that were far longer than those executed in the live exercise. This analysis identified issues for the MOC staff to explore after the exercise was complete.

#### **5. CONCLUSIONS AND FUTURE WORK**

This first iteration of the MOC-PAT proved the value of executing an operational architecture in software to assess complex Navy organizations and their processes. The MOC-PAT accurately modeled an operational staff's performance, and can provide analysts with insights into issues of

staffing and scheduling of complex process flows. The speed of configuration and simulation enabled analysts to rapidly revise the model to diagnose performance failures and test alternative configurations of the organization.

The next iteration of the MOC-PAT will include more advanced analysis tools, including reports that will support analysis and reporting by a MOC assessment team. In addition, the model is being revised to show the impact of role experience and proficiency on process execution speed (e.g., inexperienced personnel in a billet should slow activity execution while experienced staff accelerate activities.) The next version will also model shifts with greater fidelity than the current version.

In the Fall of 2009, the MOC-PAT will be used to support accreditation team observation of a Fleet MOC staff. The model is also intended to support MOC manning levels determination using data from a separate effort to identify MOC staff competencies and activity durations.

The MOC-PAT makes innovative use of an operational architecture (DoDAF OV-6) by providing a configurable, scalable, valid and executable representation of Fleet MOCs. This fusion of authoritative architectural data with simulation technology has proven to be a cost effective way to analyze complex organizational structures and human interactions.

#### **REFERENCES**

1. Department of the Navy (2009), *MOC Fact Sheet (Updated April 30, 2009)*. [https://www.portal.navy.mil/mhq-moc/fltcomm/Shared%20Documents/Fact%20Sheets/MOC%20Fact%20Sheet\\_043009.pdf](https://www.portal.navy.mil/mhq-moc/fltcomm/Shared%20Documents/Fact%20Sheets/MOC%20Fact%20Sheet_043009.pdf).
2. DoD (2007) *DoD Architecture Framework Version 1.5*. 23 April 2007. [http://jitic.fhu.disa.mil/jitic\\_dri/pdfs/dodaf\\_v1v1.pdf](http://jitic.fhu.disa.mil/jitic_dri/pdfs/dodaf_v1v1.pdf)



# NATO Human View Architecture and Human Networks

Dr. Holly A. H. Handley<sup>1</sup>; Dr. Nancy P. Houston<sup>2</sup>

<sup>1</sup>Pacific Science & Engineering Group, [HollyHandley@pacific-science.com](mailto:HollyHandley@pacific-science.com)

<sup>2</sup>NATO Allied Command Transformation, [houston@act.nato.int](mailto:houston@act.nato.int)

**Abstract.** The NATO Human View is a system architectural viewpoint that focuses on the human as part of a system. Its purpose is to capture the human requirements and to inform on how the human impacts the system design. The viewpoint contains seven static models that include different aspects of the human element, such as roles, tasks, constraints, training and metrics. It also includes a Human Dynamics component to perform simulations of the human system under design. One of the static models, termed Human Networks, focuses on the human-to-human communication patterns that occur as a result of ad hoc or deliberate team formation, especially teams distributed across space and time. Parameters of human teams that effect system performance can be captured in this model. Human centered aspects of networks, such as differences in operational tempo (sense of urgency), priorities (common goal), and team history (knowledge of the other team members), can be incorporated. The information captured in the Human Network static model can then be included in the Human Dynamics component so that the impact of distributed teams is represented in the simulation. As the NATO militaries transform to a more networked force, the Human View architecture is an important tool that can be used to make recommendations on the proper mix of technological innovations and human interactions.

## 1. INTRODUCTION

At the Prague Summit in November 2002, NATO recognized that transformation of the military based upon information age principles was essential, and pursued a course of transformation termed NATO Network-Enabled Capabilities (NNEC)<sup>1</sup>. The objective was to initiate a culture of information sharing to induce better situational awareness, faster decision making, and improved collaboration between nations. The potential NNEC benefits would also include improved efficiency, increased interoperability between nations, secure information sharing, improved information quality, and faster speed of command. While NNEC is often perceived as a technical transformation, in practice NNEC emphasizes people first, then processes, and finally technology. The challenge of NNEC is to achieve the proper mix of new human behaviors and competencies, organizational changes, and innovative technologies.

In order to make the transformation to NNEC, methodologies on how to represent the integration of technology and human/social systems are needed. The objective of this paper is to describe the NATO Human View Architecture, and specifically the Human Network product. Human networks can connect different

individuals performing roles in the same or different locations and the same or different organizations. The performance of the process supported by the human network is affected by the assignment of roles, responsibilities, and the existence of needed relationships. Attributes of human networks can also be implemented in a dynamic model to simulate the effect on process performance outcomes. The Human View Architecture can “effectively use information technology to rapidly mesh the individual skills of strangers into interdependent work products” [7].

## 2. THE NATO HUMAN VIEW

The NATO Research and Technology Organization (RTO) Human Factors and Medicine (HFM) Panel 155 convened a Human View Workshop in July 2007 to discuss and propose a cross-national Human View; that is an architectural viewpoint that focuses on the human as part of a system. A Human View is required to explicitly represent the human and to document the unique implications humans bring to the system. The workshop panel evaluated emerging human view concepts, proposed a candidate human view construct, and developed an outline of a NATO-wide Human View. The Human View was designed to be independent of any specific architecture framework and adaptable to different processes. The outcomes of the workshop resulted in the definition of a NATO Human View composed of eight products [5].

---

<sup>1</sup>[http://www.nato.int/cps/en/SID-1F7151AF-2FE364A1/natolive/topics\\_54644.htm](http://www.nato.int/cps/en/SID-1F7151AF-2FE364A1/natolive/topics_54644.htm)

The NATO Human View can be used to capture the human requirements and the way that humans interact with other elements of a system. It can be a design aid to specify future systems or it can be the basis for a methodology to answer questions regarding systems that have already been created. The main focus of the Human View is to capture human data and information about the interactions between humans and between humans and other elements of a system. The set of eight products that compose the NATO Human View are:

- HV-A: Concept - a conceptual, high-level representation of the human component in the enterprise architecture.
- HV-B: Constraints - sets of characteristics that are used to adjust the expected roles and tasks based on the capabilities and limitations of the human in the system.
- HV-C: Tasks - descriptions of the human-specific activities in the system.
- HV-D: Roles - descriptions of the roles that have been defined for the humans interacting with other elements of the system.
- HV-E: Human Network - the human to human communication patterns that occur as a result of ad hoc or deliberate team formation, especially teams distributed across space and time.
- HV-F: Training - a detailed accounting of how training requirements, strategy, and implementation will impact the human.
- HV-G: Metrics - a repository for human-related values, priorities and performance criteria, that maps human factors metrics to any other Human View elements.
- HV-H: Human Dynamics - dynamic aspects of human system components defined in other views.

The objective of the Human Dynamics (HV-H) product is to capture the interaction of the human system components defined in the other products (HV-A to HV-G). The design decisions recorded in the static Human View products can be appraised through a dynamic evaluation of the human system performance using the Human Dynamics. A tool such as the Improved Performance Research Integration Tool (IMPRINT)<sup>2</sup>, a human performance modeling tool developed by the US Army Research Laboratory (ARL), can be used to implement the Human Dynamics product and help system developers

predict the impact of operator attributes on system performance. Trade off analyses can also be conducted to determine the impact of system parameters on human performance metrics.

### 3. HUMAN NETWORK (HV-E)

The Human Network (HV-E) product focuses on the interaction of the human elements of the system: what nodes they reside at, how the human functions are distributed and what technology-based communication network enables collaboration. The HV-E maps frequent or critical types of information exchanges related to human roles as a way of expressing communication-based dependencies; this may include information exchange links to technological systems [1]. Elements of the HV-E may include:

- Role groupings or teams formed, including the physical proximity of the roles and virtual roles included for specific team tasks.
- Type of interaction - i.e., collaborate, coordinate, supervise, etc.
- Team cohesiveness indicators - i.e., trust, sharing, etc.
- Team performance impacts - i.e., synchronization (battle rhythm), level of engagement (command directed).
- Team dependencies - i.e., frequency/degree of interaction between roles.
- Communication/Technology impact to the team network - i.e., distributed cognition, shared awareness, common operational picture, etc.

The HV-E architecture product can be decomposed into several sub views to represent different types of information. An example of this is shown in Figure 1. The HV-E has been subdivided: HV-E1 Role Groupings (Teams), HV-E2 Team Interactions, and HV-E3 Information Requirements. This figure also shows the interconnections between the HV-E and other architectural products, including Operational Views (OV) and System Views (SV).

---

<sup>2</sup> <http://www.arl.army.mil/ARL-Directorates/HRED/imb/imprint/Imprint7.htm>

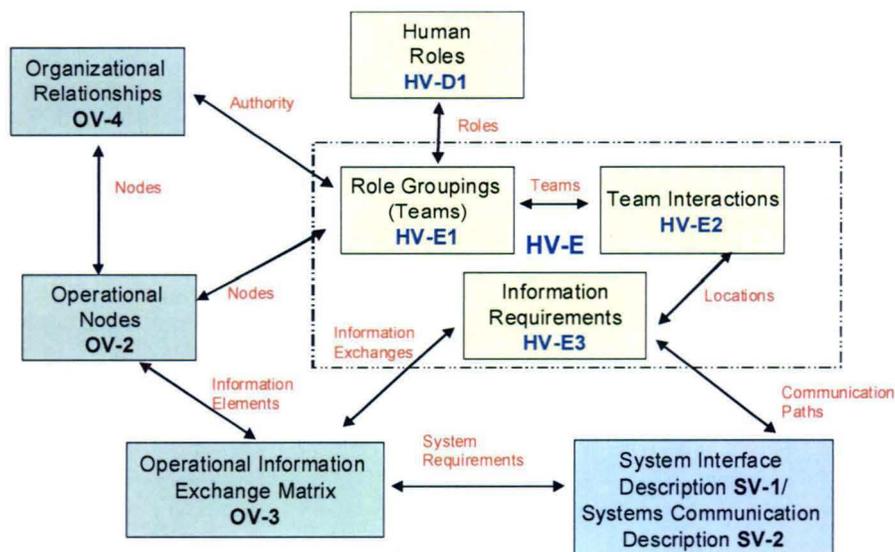


Figure 1: Human Network Relationships

#### 4. HUMAN NETWORK AND DYNAMICS

The need for a technology-supported human information network is often driven by a *reachback* situation. Reachback is “the process of obtaining products, services, and applications, or forces, or equipment, or material from organizations that are not forward deployed<sup>3</sup>.” This term is becoming widely used in the military community to indicate a virtual team. Virtual teams exist when decision-making activities are distributed across a team and the team is also distributed across physical locations. This has implications as to which types of communication media are suitable, how complex data can be shared so they are meaningful, how to organize team members across locations to minimize technology needs, or where to place the authority to ensure effective communication of commands [1].

An example of a HV-E, shown in Figure 2, depicts the collaboration requirements of a distributed military team. The need to conduct an operational activity shared between several roles creates the need for communication independent of where they are located and the need for technology supporting collaborative work. Some of the variables that may be inferred from the diagram include: which roles are at which locations, which activities are performed by what role, what roles need to share information, and

what networks are active between locations. Additionally, the roles (ovals) are shown grouped by their work centers (boxes).

The Human Dynamics captures the interaction of the human system components defined in the other products. The impact on performance of the information captured in the Human Networks product can be assessed through the dynamic model. For example, human networks, representing a collaborative team, can connect different individuals fulfilling roles in the same or different locations and the same or different organizations. The performance of the process supported by the human network is affected by the assignment of roles, the organizations to which they belong, and the connectivity between them. Personnel fulfilling designated roles must have the training and experience required to complete the task and roles in different organizations may have different leadership styles and procedures. Collaborative teams also have differing degrees of situation awareness and team history. Remote nodes experience the greatest impact of system connectivity and interoperability. Table 1 identifies the three human-related entities of role, center (organization) and collaborative team, the relevant attributes of each entity, the baseline requirement in a co-located situation, and the impact on specific performance variables as reachback may impact each attribute.

<sup>3</sup> Dictionary of Military and Associated Terms, US Department of Defense, 2005

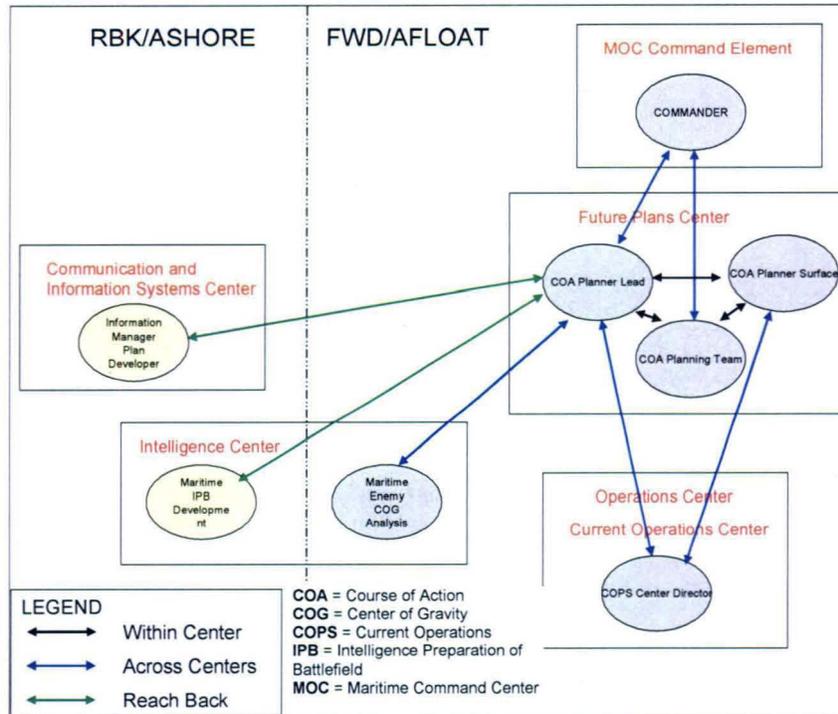


Figure 2: Human Network (HV-E) Example [4].

Table 1: Reachback Impacts

Entity	Attribute	Baseline Requirement	Performance Impact (as deviate from Baseline)
Role	Training & Experience	Fully capable to complete task.	Accuracy
Center (Organization)	Leadership & Objectives	Common goal, sense of urgency and commitment.	Timeliness
Center (Organization)	Guidance & Procedures	Same/similar task, techniques, and procedures.	Accuracy
Collaborative Team	Shared Awareness	Remote roles currently engaged in the operation	Completeness
Collaborative Team	Team History	Prior interaction with the team – leads to understanding of what knowledge, expertise, and resources each member possesses and requires.	Completeness
System	Connectivity/ Reliability/ Accessibility	Reliable and adequate communication; access to necessary information.	Timeliness

## 5. HUMAN DYNAMICS IMPLEMENTATION

The Improved Performance Research Integration Tool (IMPRINT) is a human performance modeling tool to help system developers predict the impact of operator attributes on system performance. IMPRINT can be used to predict the impact of design decisions captured in the HV-E on the performance of the operators of a system; the system can then be optimized by building models representing alternative human and technology allocations [6]. Data are entered through user interfaces and task-network diagrams; underlying human performance algorithms are then employed to perform simulations. IMPRINT incorporates task analysis, workload modeling, performance shaping and degradation functions and stressors, and embedded personnel characteristics data. Performance time and accuracy requirements are collected and workload profiles are generated so that role-workload distribution and role-system task allocation can be examined.

In order to demonstrate how the variables captured in the HV-E product can impact the Human Dynamics model, a methodology to implement reachback in an IMPRINT model was devised. First the entity types of the Human Network diagram were identified. Secondly, the attributes of the entity that vary under a reachback or co-located situations were described, along with their impact on performance. Finally, based on network theory research, workload demand differences that can be manipulated in the IMPRINT model were identified.

As an example of how reachback can be incorporated in the IMPRINT model and affect the performance outcomes, the two "collaborative team" entity attributes from Table 1 were explored further. Research on network theory has characterized Command, Control, Communications and Intelligence (C3I) architectures by the elements of People, Knowledge and Tasks and has defined the set of six relationships linking them, defined as a Meta-Network [3]. The Meta-Network also defines measures to assess each of the six defined relationships. By mapping the entity attributes of Shared Awareness and Team History to the Meta-Network, the resulting measures can be used as surrogate variables in the IMPRINT model to represent differing degrees of reachback. Shared Awareness involves knowing what knowledge is needed for a task; this is

represented by the Knowledge-to-Task relationship. Team History involves knowing who knows what; this is represented by the People-to-Knowledge relationships; see Table 2.

**Table 2:** People, Knowledge & Tasks Network [2]

	People	Knowledge	Tasks
People	Who knows who	Who knows what: <b>Team History</b>	Who does what
Knowledge		What informs what	Knowledge needed: <b>Shared Awareness</b>
Tasks			What task has precedence

Both the People-to-Knowledge and Knowledge-to-Tasks relationships effect workload outcomes and can be assessed using components of cognitive load. Cognitive load is increased when the individual roles are more interdependent (increased cognitive demand) and require more interaction between roles (increased communication demand); this is more pronounced in a reachback situation. Therefore, the cognitive and speech/auditory parameters within IMPRINT can be used as surrogate variables to represent the presence of reachback in a simulation model.

Reachback can be implemented in the model by adjusting these parameters as shown in Table 3. The IMPRINT simulation can then be configured to use the correct parameter when the interactions between team members are in either a co-located or a reachback condition. The impact of the difference will be reflected in the overall role workload, which in turn impacts the timeliness and accuracy of the work process being simulated. For example, as shown in Figure 2, there is a reachback condition between the Intelligence Center and the Future Plans Center. When these nodes communicate additional workload is added to the communication function in the model. This increases the overall workload of each role, and if it surpasses a set workload limit, it will cause a detriment in the performance of the role's functions. Several variables will impact the severity of the performance impact, including the workload threshold, the timeliness penalty, and the workload management strategy. This may include dropping tasks, off-loading tasks and/or delaying tasks.

**Table 3: Reachback adjustments to IMPRINT Parameters**

Entity / Attribute	Meta-Network Relationship	IMPRINT Workload Demand Settings*	
		Without Reachback	With Reachback
Collaborative Team: Shared Awareness	Knowledge-to-Tasks	Cognitive = 4.6 Eval/Judge/ Consider Single Aspect	Cognitive = 6.8 Eval/Judge/Consider Several Aspects
Collaborative Team: Team History	People-to-Knowledge	Speech = 2.0 Simple Auditory = 3.0 Simple	Speech = 4.0 Complex (Sentence) Auditory = 6.0 Interpret Speech Complex (Sentences)

**6. CONCLUSION**

For network-based operations, such as NATO NNEC, the collaboration requirements between distributed roles and the resulting communication patterns are of particular importance. The NATO Human View product, Human Networks, focuses on capturing the parameters and variables that characterize the human communication processes and can provide the necessary data for a simulation model for evaluation of network-based systems.

The NATO Human View is currently being evaluated for integration into the NATO Architecture Framework (NAF) 3.0 through the Human Views extension to Enterprise Architecture project. Through this process, the NNEC elements for the Human View products have been defined, the integration of these products into the overall framework has been designed, and example Human View products have been created for the NAF Running Example. These documents have been posted on the NATO TIDEPEDIA<sup>4</sup> for comment and evaluation.

By providing a mechanism for capturing the required data to characterize the human interactions, the Human Network provides a valuable tool for designing human centered systems and evaluating the impact of the human component on the overall system. As the NATO militaries transform to a more networked force, the Human View architecture is an important tool that can be used to make recommendations on the proper mix of technological innovations and human roles, interactions, and behaviors.

<sup>4</sup>[http://tide.act.nato.int/mediawiki/index.php/Human\\_Views\\_extension\\_to\\_Enterprise\\_Architecture\\_project](http://tide.act.nato.int/mediawiki/index.php/Human_Views_extension_to_Enterprise_Architecture_project)

**REFERENCES**

1. Bruseberg, A. (2008). The Human View Handbook for MODAF, System Engineering & Assessment Ltd. On behalf of the MoD HFI DTC, First Issue,
2. Carley, K. M. (2002). "Inhibiting Adaptation" In *Proceedings of the 2002 Command and Control Research and Technology Symposium*. Conference held in Naval Postgraduate School, Monterey, CA.
3. Carley, K. M. & Ren, Y. (2001). "Tradeoffs between Performance and Adaptability for C3I Architectures." In *Proceedings of the 2000 International Symposium on Command and Control Research and Technology*, Naval Post Graduate School, Monterey, CA.
4. Handley, H., Sorber, T., & Dunaway, J. (2006). Maritime Headquarters with Maritime Operations Center, Concept Based Assessment, Human Systems Performance Assessment Capability Final Report. Pacific Science & Engineering Group, San Diego, CA.
5. Handley, H. & Smillie, R. (2008). "Architecture Framework Human View: The NATO Approach". *System Engineering* 11(2) 2008, 156-164
6. Mitchell, D. (2005). Enhancing system design by modeling IMPRINT task workload, *Human System Integration Symposium*, Arlington, VA, June 2005.
7. Piccoli, G., Powell, A. & Ives, B. (2004). "Virtual teams: Team control structure, work process and team effectiveness". *Information Technology & People* 17(4), pp. 359-379.

# A Simulation-Based Approach to Training Operational Cultural Competence

Dr. W. Lewis Johnson

*Alelo, Inc.*

*ljohnson@alelo.com*

**Abstract.** Cultural knowledge and skills are critically important for military operations, emergency response, or any job that involves interaction with a culturally diverse population. However, it is not obvious what cultural knowledge and skills need to be trained, and how to integrate that training with the other training that trainees must undergo. Cultural training needs to be broad enough to encompass both regional (culture-specific) and cross-cultural (culture-general) competencies, yet be focused enough to result in targeted improvements in on-the-job performance. This paper describes a comprehensive instructional development methodology and training technology framework that focuses cultural training on operational needs. It supports knowledge acquisition, skill acquisition, and skill transfer. It supports both training and assessment, and integrates with other aspects of operational skills training. Two training systems will be used to illustrate this approach: the Virtual Cultural Awareness Trainer (VCAT) and the Tactical Dari language and culture training system. The paper also discusses new and emerging capabilities that are integrating cultural competence training more strongly with other aspects of training and mission rehearsal.

## 1. INTRODUCTION

There is a growing awareness of the need for intercultural knowledge and skills across a wide range of disciplines. Recent conflicts in Iraq, Afghanistan, and other countries have highlighted the importance of cultural issues in 21<sup>st</sup> century military operations [14]. Multicultural workplaces, health care and educational settings, and emergency response teams are just some of the other contexts in which intercultural knowledge and skills have been found to be important [5], [10], [15], [16], [17].

This need for intercultural knowledge and skills poses a significant training challenge. Experts in the language education community, such as the American Council on the Teaching of Foreign Languages (ACTFL), commonly assert that true cultural competence arises only after years of immersion in the target culture, as part of a language education program [13]. Unfortunately, relatively few specialists get the opportunity to devote that much time to cultural training. For example, many service members deploying to Iraq and Afghanistan get just a few hours of cultural awareness training, and some get none at all [9].

This paper describes a simulation-based approach to cultural competency training realized in the Alelo family of training products. It is intended to help trainees who may not be cultural specialists quickly develop cultural skills they need to be effective in intercultural settings. This includes the knowledge and skills necessary to handle common intercultural interactions, as well as the adaptability needed to cope with unexpected intercultural interactions. Crucially, it utilizes simulations of intercultural situations that trainees are likely to encounter in the course of carrying out their jobs or missions, which helps trainees develop the skills and confidence necessary to apply intercultural skills in those situations. The approach also supports the assessment of cultural competence by testing trainees in simulated intercultural encounters.

Although intercultural skill is an important outcome of this training, it is not the only desired outcome. Researchers in cross-cultural competence such as Abbe, et al. [2], have argued that intercultural competence has knowledge and affective components as well as skill components, and that all three should be promoted in culturally competent individuals. Therefore, the Alelo training method does not rely solely on simulations, but integrates simulation-based training with multimedia instruction and affective elements, employed in a coordinated fashion to help trainees develop robust intercultural competence.

We have developed a cultural analysis and instructional design methodology, called the Situated Culture Methodology (SCM), which facilitates the creation of training courses that utilize this approach. SCM focuses sociocultural data collection and instructional design on situations that trainees are most likely to encounter in the context of their work activities. This provides a necessary focus to the cultural training. Cultural training courses can have a tendency to turn into a litany of facts about the target culture—of interest to the cultural specialist, but of questionable value to the ordinary intercultural practitioner. SCM focuses on culture relevant to the job or mission context, resulting in training that is more coherent and effective in achieving its intended learning outcomes.

This approach has been employed to develop a variety of Alelo training courses, such as the Tactical Language family of language and culture training courses [8] and the Virtual Cultural Awareness Trainer (VCAT) [4]. Tens of thousands of trainees have made use of these courses to date, with demonstrated positive impacts on cultural competence as well as overall operational effectiveness [11].

## 2. EXAMPLE TRAINING MATERIALS

Two courses will serve as examples for following discussion: Tactical Dari and VCAT. Tactical Dari is a PC-based serious game that helps learners acquire

operational knowledge of the Dari language and Afghan culture. VCAT helps trainees develop pre-deployment cultural awareness of the Horn of Africa. Both provide training in language and cultural skills, although Tactical Dari places greater emphasis on language skills and VCAT puts greater emphasis on cultural awareness and operational cultural knowledge.

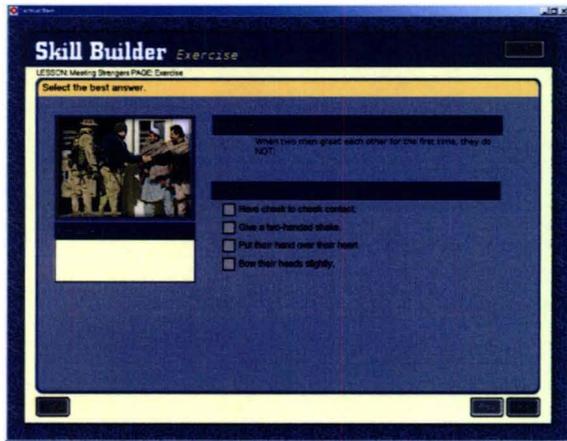


Figure 1. Tactical Dari nonverbal interaction quiz

Tactical Dari includes interactive lessons that focus on common skills relating to interpersonal interaction in a Dari-speaking context, such as greetings, introductions, arranging meetings, discussing business with counterparts, etc. The course addresses all aspects of interpersonal communication in such contexts, including appropriate forms of address, gestures and body language, and social norms for hospitality and relationship building in conducting business. These concepts are all taught in the context of interactive lessons and exercises. Figure 1 shows one such exercise. Here the learner is asked to identify appropriate gestures and body language for men greeting each other for the first time in Afghanistan.

Tactical Dari then gives learners opportunities to practice their intercultural skills in simulated encounters with Afghans. Figure 2 shows such a simulated meeting. The player character, gesturing on the left, is leading a team engaged in discussions with the village leader and other elders (right) about collaborating on a reconstruction project. The trainee plays his role by speaking into a headset microphone in the Dari language and selecting appropriate gestures for his or her character. The built-in speech processing system interprets the trainee's speech in context and causes the non-player characters to generate socially and culturally appropriate responses.

In order to succeed in a scene such as this, trainees must employ a wide range of cultural skills. For example, they should remember to introduce everyone in their team. They should inquire about the elder's family, but not in a way that causes embarrassment (e.g., by inquiring inappropriately about female family members). They should make appropriate use of Dari phrases expressing respect and humility, at socially appropriate times. For example, they should make use of the Dari phrase /naame khudaa/ (Thanks be to God) in acknowledging the leader's good fortune as well as

one's own. Trainees thus learn not just to understand and recognize culturally appropriate behavior, but are able to practice until they become skilled at behaving the right way at the right times.

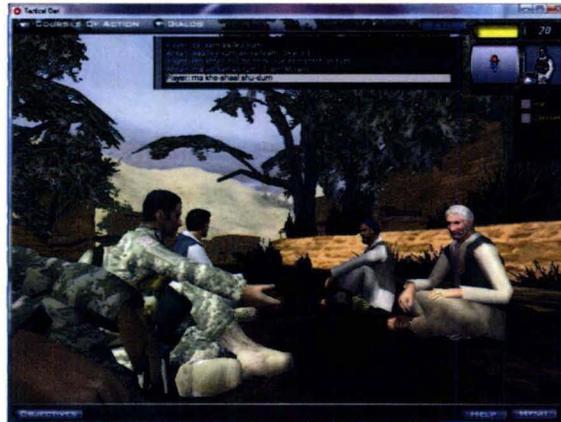


Figure 2. Meeting with Afghan village elders

VCAT is a Web-based training course, accessed via Joint Knowledge Online. When trainees start the course they indicate the particular country they will deploy to, their level of seniority, and the type of mission they are likely to undertake in the region. VCAT then automatically selects a tailored curriculum that focuses on their particular needs.

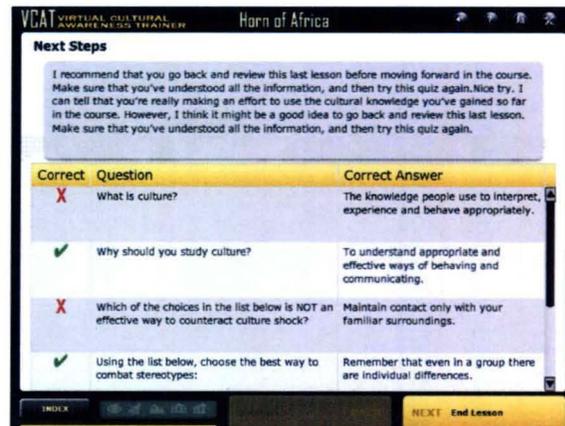


Figure 3. VCAT culture-general feedback

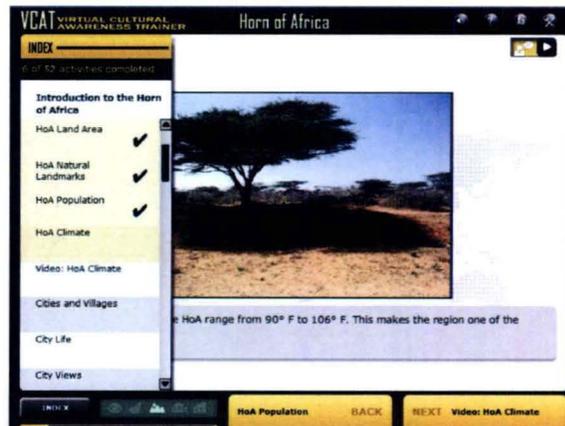


Figure 4. Introduction to the physical environment in the Horn of Africa

At the start of the course trainees are given a general introduction to culture—what it is, and why it is important. They are also introduced to six general dimensions of culture: physical environment, social structure, political structure, economic structure, and cultural perspectives and practices. This provides learners with a conceptual framework for understanding culture and how it applies to their mission and activities in country. Figure 3 shows feedback from a quiz at the end of this section, where the learner is tested on these general cultural concepts.

VCAT then introduces trainees to the culture of their region of interest. Like Tactical Dari, it includes lesson modules that introduce cultural concepts. These touch on the various dimensions of culture relevant to operations in the Horn of Africa. For example, Figure 4 shows part of the course that introduces trainees to the physical environment of the Horn of Africa, particularly the hot climate. Climate is important in HOA in part because it affects when and how people work and conduct business. People are less likely to be available to work at the hottest times of the day, and offering and receiving drinks is an important aspect of hospitality when meeting with people in the region.

As in Tactical Dari, trainees get opportunities to practice their cultural skills in immersive simulations. Trainees are not required to speak in the local language on behalf of their avatar, but they are required to make choices in the context of the situation, and thereby become skilled at behaving appropriately in that situation. Figure 5 illustrates one such scenario. Here the trainee character (right) is engaged in a meeting with the local health minister to discuss a mission to deliver medical supplies. The minister has offered the trainee water to drink, and the trainee must decide whether to accept it. The trainee must consider the health risks of drinking water that may be contaminated, the health risks of *not* drinking fluids in such a hot climate, as well as the risk of offending the health minister by declining the offer. Such situations require trainees to apply their cultural knowledge in complex situations where multiple factors are involved. Thus they are more likely to be prepared when they are required to put their cultural skills into practice.



Figure 5. Practicing cultural skills

### 3. THE SITUATED CULTURE METHODOLOGY

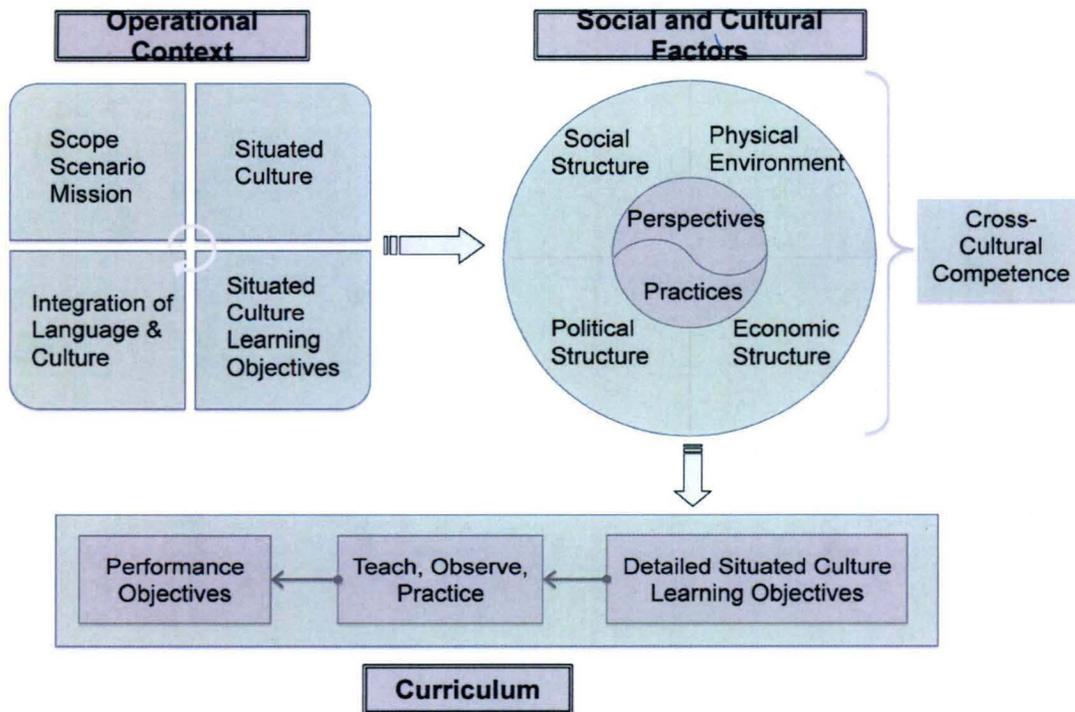
A key challenge in developing such culture courses is determining what cultural information is relevant to the trainee's needs and presenting it appropriately. Cultural literacy approaches (e.g., [3]) tend to present a range of facts about the culture, but do not give much consideration to which of those facts might be relevant to trainees and contribute to useful intercultural skills. Cultures are rich and varied, and so descriptions of a given culture can easily become very large and wide-ranging. An effective cultural training approach needs to take cultural information, which may be interesting in its own right from an anthropological perspective, and focus on the material most relevant to trainee needs in specific contexts.

Military approaches to teaching culture, such as the Marine Corps's notion of operational culture [14], are a step in the right direction. They focus on cultural information that is relevant to military operations. But even that provides insufficient focus for the purpose of intercultural-skills training. In the military case, for example, successful intercultural-skills training requires knowing about culture that is operationally focused as well as *operator-focused*, i.e., focused on what an individual military operator (servicemember in the field) needs to know and be able to do to be effective in the intercultural situations they are likely to encounter in military operations. Similar concerns arise in nonmilitary courses, such as our goEnglish course ([www.goEnglish.me](http://www.goEnglish.me)), developed to teach American English and American Culture worldwide. The focus there is to provide learners with an understanding of culture and communicative skills they are likely to need in everyday situations they are likely to encounter when they come to the United States.

These concerns led Alelo to develop a methodology for designing intercultural competence courses known as the Situated Culture Methodology (SCM). An overview of the SCM approach is shown in Figure 6.

#### 3.1 Focus on operational context

SCM focuses the curriculum on the operational context: the range of situations in which the trainee is expected to apply the cultural knowledge being taught. The factors involved in considering operational context are depicted in the top left of Figure 6. First, it is necessary to determine the desired scope of the course: the size of the region that trainees are being trained for, the degree of cultural proficiency sought, and the range of jobs and missions the trainees are being prepared for. In most training courses, the scope of the course is constrained by the amount of time available to train, which is typically quite limited. For example, trainees who train using the VCAT course are expected to have an average of only four hours of training time. The program of instruction for each trainee must therefore be very narrow in scope.



**Figure 6. Situated Culture Methodology**

To maximize training effectiveness for each trainee while complying with stringent limitations on training time, we generally adopt an approach in which the scope of the course is adapted for each individual trainee through tailored programs of instruction. When trainees begin a course, they complete a brief questionnaire in which they indicate the nature of their job and anticipated overseas assignment. The training software platform then dynamically configures the curriculum to fit those job requirements. In the case of Tactical Dari, the servicemember's rank and specialty help to determine the scope of the curriculum. In the case of VCAT, seniority helps to determine the curriculum focus, as well as the specific country in the Horn of Africa that the servicemember is deploying to. This means that the curriculum designer needs to plan for not just one course scope, but multiple scopes, each of which may involve different learning objectives.

To determine the cultural knowledge associated with a particular scope, instructional designers identify scenarios and missions that are typical for a given rank or specialty in the cultural region of interest. These, in turn, suggest common situations that trainees are likely to encounter, e.g., meetings with local leaders and counterparts, chance encounters with children in the street, or patient interviews in a medical clinic. These situations are what provide the primary context and focus for the courses. The cultural training course is successful if it can properly train people to be effective in the intercultural exchanges they are likely to encounter in those situations. Those situations also help to determine the specific learning objectives that are captured in the curriculum design. These typically include both cultural-competence objectives as well as language-skill objectives, since both may be necessary to cope with a given scenario.

Although the scope of the curriculum depends upon the responsibilities of the individual trainee, there is typically a significant amount of overlap in each individualized curriculum. Some cultural skills, such as culturally appropriate greetings, are likely relevant regardless of the trainee's job. Some skills, particularly cross-cultural competence skills, are relevant regardless of the target culture. In some cases the trainees' job responsibilities cannot be anticipated with precision. Moreover, in practice there are limitations to the amount of individualized training materials that can be authored and developed. These factors serve to limit the variability in the situated-culture training objectives.

### 3.2 Organize around cultural dimensions and factors

Once the situated-culture learning objectives are identified, the next step is to identify cultural information that addresses those learning objectives, and organize it for inclusion in the training course. To help make it easier for trainees to understand what cultural information is relevant for operational purposes, we organize the cultural material along six dimensions: social structure, physical environment, political structure, economic structure, perspectives, and practices. The social, physical, political, and economic factors are concerned more with the macrosocial aspects of culture. Perspectives and practices are concerned more with factors relating to the microsocioal, or one-on-one interactions, and provide more of an individual- or operator-oriented view. Perspectives include time orientation, individualism vs. collectivism, task-oriented vs. relationship-oriented working relations, and other attitudes toward personal relations. Practices include discourse genres, conversational culture, nonverbal communication, politeness norms, formality vs.

informality, and other factors influencing cultural action.

To ensure that the cultural material under each cultural dimension is relevant to the operational context, each cultural dimension is subdivided into cultural factors, which are a set of cultural topics that are frequently relevant to operational applications. For example, the physical environment dimension includes the following topics: division of the terrain into cultural regions, patterns of land use across the region, patterns of human movement and contact between regions, access to drinking and irrigation water, and access to energy and fuel. By working through the target scenarios with subject matter experts, it is possible to identify cultural factors that are likely to be relevant. For example, in a humanitarian assistance scenario in VCAT, access to potable water and electrical power turned out to be critical factors in locating a site for a medical relief station.

To further focus the research and identify relevant cultural information, we consider *operational cultural questions* associated with each factor. Operational culture questions are questions pertaining to a cultural factor that commonly arise in the context of a given type of operation. We have gathered operational culture questions from military culture resources, (e.g., [14]), and extend the set of such questions as needed to serve the goals of the course curriculum. Using these questions in interviews with subject matter experts can help uncover additional cultural factors that need to be considered in the course. For example, a number of operational culture questions pertain to water and power, such as who has access to it, who provides or controls access to it, and how local people deal with shortages of it.

In addition to identifying relevant cultural factors and skills, we seek to identify optimal metacultural skills – knowledge and skills that are useful in any cross-cultural situation. The successful use of these skills is commonly referred to as cross-cultural competence. Cross-cultural competence can come into play in explaining why particular cultural factors pertain in a particular situation and to draw lessons that learners can apply in the future to similar situations. Consider for example the case of a humanitarian relief scenario in which local officials make requests or proposals that, from an American perspective, are considered inappropriate, such as asking for preferential treatment for their family or tribe. This affords the trainee opportunities to reinforce metaskills such as perspective taking and not being judgmental about differences in social norms and attitudes.

### **3.3 Design performance-oriented curriculum**

Once the cultural learning objectives have been identified and the cultural content has been researched and organized, we then design the cultural curriculum materials. Based on the analysis of relevant cultural factors described above, a detailed set of situated-culture learning objectives are defined.

Most social and cultural factors involve a combination of knowledge-oriented learning objectives (e.g., the ability to recall cultural facts about the region of interest) and skill-oriented learning objectives (the ability to apply cultural knowledge in specific settings to achieve particular objectives). This typically results in a coordinated set of learning materials to cover each cultural topic: presentation materials and exercises that help ensure that trainees understand the relevant cultural concepts and how they apply to task objectives, and dramatized scenarios that help learners to develop the necessary skills and the ability to apply them in typical situations at a desired performance level.

Simulation therefore plays an essential role in the resulting courses, particularly in developing and reinforcing the intercultural skills. However, simulation-based activities must be supported by other learning activities that help develop the underlying cultural knowledge. This motivates the hybrid instructional designs of training products such as Tactical Dari and VCAT, which interleave interactive multimedia instruction and simulation-based training.

Although the emphasis of this discussion has been on learning cultural skills, it should be evident that language skills are involved as well. Many of the detailed skills involve communicating with host nationals to achieve particular objectives. For this reason, foreign language skills are an essential part of the courses that we develop. This is a controversial point in some circles; for example, some in the military have argued that military operators can rely on interpreters, or even translation devices, and therefore have little need for language skills beyond some minimal vocabulary [1]. Recent experience in operations in Iraq and Afghanistan give reason to call that position into question. For example, a Marine Corps Center for Lessons Learned (MCCLL) study of the use of Tactical Iraqi training by the 3<sup>rd</sup> Battalion, 7<sup>th</sup> Marines indicated that there were not enough interpreters assigned to the battalion to support every encounter between Marines in the unit and host nationals, so that it was beneficial to make sure that every squad had at least one or two individuals with Arabic language training. The MCCLL study [11] showed also that even basic language skills facilitated operational culture objectives, because it demonstrated that the Marines were knowledgeable about the host nation culture and were open and receptive to intercultural interchange.

We believe that one reason why the language training in Alelo's courses is effective is because, unlike most conventional language courses and language learning software, the language-learning objectives are so closely tied to cultural-learning objectives and situated-task objectives. The language-learning objectives are in support of particular intercultural communicative skills. Popular language-learning software packages, such as Rosetta Stone, deliberately disregard cultural factors to make it easier to create a family of language courses that all teach in a uniform way. Alelo courses also contrast with language courses that promote general language proficiency, regardless of task application.

The design methodology described above assumes that when training time is limited, a more efficient training approach may be to focus on language skills that are involved in the specific communicative skills of interest.

Because intercultural skills are applied in particular situational and task contexts, it is appropriate to consider integrating cultural-skills training with other skills training. A recently developed capability, named Virtual Role Players (VRP), makes precisely that possible. Small units may practice missions in immersive simulations where they can practice their culture and language skills in encounters with simulated non-player characters. Once this capability is more broadly put to use, intercultural communication skills will simply become part of the broader range of skills that trainees employ to carry out their training exercises.

The SCM methodology is an iterative design approach. Once early versions of cultural training products are developed, they undergo pilot testing with representative users. Revisions are then made to scenarios, situations, learning objectives, or learning content to reflect end-user needs and feedback, as appropriate.

#### **4. THE ROLE OF SUBJECT MATTER EXPERTS**

Subject matter experts (SMEs) play critical roles throughout the SCM development process. There are three main types of subject matter experts: task experts, culture experts, and language experts. Task experts are familiar with how to conduct the scenario or mission; typically they have experience conducting missions in the area of focus, and understand how the local culture can influence the conduct of the mission. Culture experts are native to the region, and ideally have some expertise in analyzing and explaining culture. Language experts have native or near-native proficiency in the target language, and ideally have the metalinguistic skills to reflect on language forms and the language-teaching skills to explain them to learners who are unfamiliar with the language. Sometimes a SME may play more than one role at once, e.g., a culture SME may have linguistic knowledge, or may have worked as an interpreter in the region of interest and therefore may have some mission experience.

The more specific and remote the region of interest, the more difficult it can be to find culture and language SMEs who have expertise that is accurate and up-to-date, and also have the ability to reflect on and explain it. To cope with this problem, we typically rely on multiple SMEs with overlapping knowledge, and triangulate between SMEs, as well as between SMEs and other resources. We typically start with one or more SMEs who have broad knowledge of the general area and good ability to articulate and explain it. These individuals can help provide general background, as well as identify important dimensions of cultural variability and diversity in the region. We then refine this with specific information obtained from SMEs with more local knowledge. We try to interview both male

and female SMEs, in order to get a well-rounded picture of the target culture. Triangulating between SMEs is particularly helpful in developing learning materials for courses with multiple overlapping scopes (e.g., multiple countries or local regions within the same area) or whose scope overlaps with that of a previous course (e.g., cultural factors in the Tactical Dari course overlapped with those in an earlier Tactical Pashto course, since the two languages are spoken in the same region).

In the case of task expertise the requisite knowledge is more widely available and tends to be well documented, at least for military task expertise. However, SMEs who have the most specific and in-depth knowledge are often active-duty personnel with limited availability. So we typically try to rely on a combination of multiple SMEs here as well. Task SMEs with the most current and accurate knowledge help define the task requirements and validate scenario designs for accuracy. Then we may rely on other task SMEs who are retired and no longer currently involved with such missions, but who have access to current resources and reports, to add further detail to the task descriptions, as needed.

Work with SMEs progresses in stages, to inform the design process as well as possible and to aid in validation of the content. In the first stage, the focus of discussion with SMEs is on outlines of the task and scenarios (in the case of task SMEs) and cultural and linguistic topics (in the case of culture and language SMEs, respectively). This may initially consist only of a summary of common subtasks to perform, and common phrases and communication requirements. This is used as the basis for creating outlines of the curriculum scenarios. We then ask SMEs to review and approve these outlines. The culture researchers on the team then conduct preliminary research from Internet and library resources, and then follow up later with the SMEs for additional information gathering and review.

The primary method for gathering detailed cultural information from cultural SMEs is ethnographic interviews. In principle, ethnographic observation in real-world contexts is desirable, but is often impractical, particularly in hazardous overseas locations. Role-playing exercises are also helpful to elicit further details about cultural practices and this method is used for targeted information gathering. This method is particularly valuable in gathering information about nonverbal communication and other aspects of culture that are implicit or tacit and need to be made more explicit. We also ask SMEs to write dialogs in the foreign language that are typical for the target scenario; this helps to clarify what specific language skills and cultural skills are involved.

During the authoring phase, it is desirable to have SMEs available on an ongoing basis to answer specific questions and review authored content for accuracy. Often some of these SMEs become members of the authoring team, and help edit and review material.

## 5. ADDRESSING CROSS-CULTURAL COMPETENCY STANDARDS

Cultural competency training is an evolving field, and broadly accepted standards are not yet established. However, cultural training standards are emerging in the military training arena, and so it is worthwhile to compare military courses developed using the SCM approach against these standards. Other disciplines such as medicine and education have their own cultural competency standards, and so courses developed in those fields would need to be compared against those standards.

The Defense Regional and Cultural Capabilities Assessment Working Group (RACCA WG) [12] has identified forty cultural competencies relevant to military training. These were intended to be oriented to the cultural training needs of junior military personnel, and so are not precisely suited for all of Alelo's courses. Nevertheless, they provide a useful basis for comparison.

VCAT, in particular, does a fairly good job of addressing the RACCA learning objectives. Of the forty cultural competencies identified by the RACCA WG, VCAT addresses twenty-six of them, and partially addresses an additional nine. The remaining objectives are primarily culture-general and not oriented toward foreign military operations (e.g., focusing on American military culture).

The ability to meet the full range of RACCA WG learning objectives within a given course is constrained by the amount of available training time. VCAT, in particular, is designed to be completed in a short period of time and this necessitates focusing strongly on cultural skills that can be put to immediate use in overseas deployments. If, in the future, cultural training is given greater emphasis and, therefore, training time, it will be possible to incorporate more cross-cultural competence skills within the SCM framework.

Older Tactical Language courses, such as Tactical Iraqi and Tactical Dari, cover a more limited range of cultural skills, in part because they give greater emphasis to language skills. However, we are taking lessons from VCAT and supporting a wider range of cultural learning objectives in current language and culture courses. For example, the new Operational Indonesian course includes a module on culture and metacultural skills, and provides learners with a framework for understanding culture in general as they learn about the particular cultures and languages of Indonesia.

## 6. TECHNOLOGY IMPLICATIONS

A range of software tools and instruction delivery technologies can be employed to support the methodology described above. Alelo has already developed many of these technologies and others are the subject of ongoing research and development.

- Cultural information management tools are needed to gather, annotate, and organize cultural information. We currently use Google

Notebook for initial data collection and Fedora for managing media assets.

- Authoring tools are needed to specify the content to be delivered. These should support collaborative authoring, including participation by SMEs. It should support both interactive multimedia authoring and interactive simulation authoring. We have developed a Web-based authoring portal named Kona, to meet this need [7].
- The authoring tools and content delivery tools should support tailoring of the content for the needs of individual trainees. We have therefore incorporated such tailoring functions into each of our content delivery systems [8]. For courses that are delivered using SCORM-compliant learning management systems, such as VCAT, we rely on the advanced sequencing functions in SCORM to tailor the curriculum.
- Since trainees typically have limited time to train, it is desirable to provide trainees options for continuing their training and maintaining their skills. Ideally, these should provide trainees the option of training anywhere, anytime, to maximize their available training time. To meet this need, we have developed multi-platform content delivery systems to deliver content on whichever delivery platform trainees find most convenient, including the handheld platforms [6].
- Since cultural skills training developed using the SCM method is closely aligned with task training, it is useful to provide the option of training intercultural skills and task skills together. Alelo's Virtual Role Player training capability, that integrates artificially intelligent virtual role players into multiplayer training systems, helps to meet this need.

## 7. SUMMARY

This paper has presented an approach for cultural competency training, which is designed to help trainees quickly acquire useful intercultural skills. An authoring methodology, Situated Culture Methodology, has been created to develop these courses. This methodology results in courses that are tailored to the needs of individual trainees, and provides good coverage both of regional cultural competency and cross-cultural competency, in a limited amount of training time.

Courses designed using this method are available for download from the Alelo support Web site, as well as on Joint Knowledge Online.

## 8. ACKNOWLEDGMENTS

The author wishes to acknowledge various members of the Alelo team who contributed to the work described here, including Dr. Suzanne Wertheim, Dr. Andre Valente, Michelle Flowers, Rebecca Row, LeeEllen Friedland, and Kristen Russell. Thanks to LeeEllen Friedland for her comments on this article.

This work was funded by the US Joint Forces Command, USMC Program Manager for Training Systems, the Office of the Secretary of Defense, and the Defense Language Institute. Opinions expressed in this article are those of the author, and not of the US Government.

## 9. REFERENCES

1. Abbe, A. (2008). *Building Cultural Capability for Full-Spectrum Operations*. U.S. Army Research Institute technical report, study report 2008-04.
2. Abbe, A.; Gulick, L.M.V.; & Herman, J.L. (2008). *Cross-Cultural Competence in Army Leaders: Conceptual and Empirical Foundation*. U.S. Army Research Institute technical report, study report 2008-01.
3. Arvizu, S.; & Saravia-Shore, M. (1990). "Cross-cultural literacy, An anthropological approach to dealing with diversity". *Education and Urban Society* 22 (4) 364-376.
4. Camacho, J. (2009). "Next Generation JKO". *Military Training Technology* 14 (5), September-October. Available at <http://www.military-training-technology.com/mt2-archives/197-mt2-2009-volume-14-issue-5-/1925-next-generation-jko.html>.
5. Earley, P.C. (1987). "Intercultural training for managers: A comparison of documentary and interpersonal methods". *The Academy of Management Journal* 30 (4), 685-698.
6. Johnson, W.L.; Valente, A.; & Heuts, R. (2008). "Multi-platform delivery of game-based learning content." *New Learning Technologies 2008 Orlando SALT Conference*.
7. Johnson, W.L. & Valente, A. (2008). "Collaborative Authoring of Serious Games for Language and Culture". In *Proceedings of SimTecT 2008*.
8. Johnson, W. L. & Valente, A. (2009). "Tactical Language and Culture Training Systems: using AI to teach foreign languages and cultures". In *AI Magazine* 30 (2), 72-84.
9. Khatchadourian, R. (2009). "The Kill Company". *The New Yorker*, July 9, 2009.
10. Kosoko-Lasaki, S., Cook, C.T., & O'Brien, R.L. (2008). *Cultural Proficiency in Addressing Health Disparities*. Boston: Jones & Bartlett.
11. MCCLL (2008). "Tactical Iraqi Language and Culture Training System" *Marine Corps Center for Lessons Learned Newsletter* 4 (8), 4.
12. McDonald, D.P., McGuire, G.; Johnson, J.; Selmeski, B.; & Abbe, A. (2008). Developing and managing cross-cultural competence within the Department of Defense: Recommendations for learning and assessment. Technical report, RACCA WG.
13. National Foreign Language Center (NFLC) (2007). Cultural Proficiency Guidelines (3.2). On line at [www.nflc.org/culture-prof-guide\\_3-2.pdf](http://www.nflc.org/culture-prof-guide_3-2.pdf).
14. Salmoni, B.A. & Holmes-Eber, P. (2008). *Operational Culture for the Warfighter: Principles and Applications*. Quantico, VA: Marine Corps University Press.
15. Schneider, P. & Sadowski, D. (2009). "The effects of intercultural collaboration strategies on successful PhD education". *Proc. of IWIC 2009*. New York: ACM Press.
16. Sorcher, M. and R. Spence. 1982. "The interface project: Behavior modeling as social technology in South Africa". *Personnel Psychology* 35(3), 557-581.
17. US Dept. of Health and Human Services (USDHHS) (2008). Cultural competency curriculum for disaster preparedness and crisis response. On line at <https://cccdpccr.thinkculturalhealth.org/>

# Propagation of New Innovations - An Approach to Classify Human Behavior and Movement from Available Social Network Data

**Mahmud, Faisal; Hasan, Samiul<sup>1</sup>**

Civil & Environmental Engineering Department  
Old Dominion University, Norfolk, Virginia, USA

Massachusetts Institute of Technology<sup>1</sup>, Cambridge, Massachusetts, USA

*E-Mail: [fmahm001@odu.edu](mailto:fmahm001@odu.edu), [shasan@MIT.EDU](mailto:shasan@MIT.EDU)*

**Abstract.** It is interesting to observe new innovations, products, or ideas propagating into the society. One important factor of this propagation is the role of individual's social network; while another factor is individual's activities. In this paper, an approach will be made to analyze the propagation of different ideas in a popular social network. Individuals' responses to different activities in the network will be analyzed. The properties of network will also be investigated for successful propagation of innovations.

## 1. INTRODUCTION

In this world of science and technology, online social networking gained vast popularity for chatting, photo/ video/ file sharing as well as communicating with friends, family members, colleagues and others. Recently, websites like Facebook, Flickr, Youtube and Myspace turned into the media of communication and connect users to each other with the purpose of finding and exchanging content. As people are passing their more times in these sites, these websites become a marketing platform for companies' products, in some cases political campaigns. From movie trailers to amateur fun video, everything can reach to millions of online users through these entire social networks.

In this paper, the authors proposed a propagation concept on new innovations through which human behavior can be classified upon using popular social network sites. For data analysis and conceptualization of data spread in the social networks, only Facebook, a popular social network site is considered. Thus we have divided our paper into three main parts- Part A: Research questions and probable solutions, Part B: Proposal of a propagation theory and Part C: Analysis and link establishment between social network to real world. This paper begins with a background to the related work in the very concept of propagation theory. Then Part A to Part C is discussed following the conclusions and future works.

### 1.1 Literature Review

There have been various works related to social networking and characterizing. One of the distinguishing features of online social networks is information dissemination along social links. Content in the form of ideas, products, and messages spreads across social connections like a virus: one person discovers new content and shares it with a few of their friends, who share it with a few of their friends, and so on. Alan et al. called this spreading of a piece of content along links in a social network a "social cascade" [1]. Seminal work on persuasive communication, the branching process, and the diffusion of innovations spawned extensive literature in sociology, economy, social psychology, political science, marketing, and epidemiology [1, 2, 3, 4, and 5].

In order to get the path of propagation, we had to obtain the traces of content dissemination from Facebook [6]. Before the data analysis in Part: C, Part: A, and Part: B are discussed with some research questions, probable answers and establishment of a propagation theory.

## 2. Detailing of Analysis

**Part: A**  
**Research questions and probable answers**

All type of research works and projects are based on some questions and the goal of a project work or research is to find out the answers of these questions either with an explanatory remark or not. In this research work, we had to face some research goals with the following questions:

1. What is the significance of utilization of a social network data and what should it meant for?
2. Why we should consider social networking as an important media of communication?
3. If social networking is an important media of communication, then how we can consider it for our real life human behavior classification?

By getting the explanatory answers to these questions, we can make a decision- a social network is such a new innovation that can help to classify human behavior.

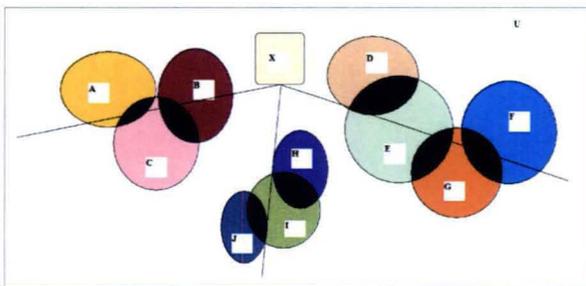


Figure 1: Visualization of common interest.

In Facebook, friends, family members, colleagues and sometimes unknown persons relate to each other through the friend list. No matter whether any two person knew each other or they have common interest, just to increase one's own friend list, he/she may add other people in his/her friend list. This is a case where people add randomly a person into his/her network. In other case, suppose A, and C are friends (figure 1) whereas B and C are friends. A may not know B and vice versa. But, if all these A, B and C are the members of a common group, say X, then they can know each other. The only other way to relate A, and B is possible when C can allow both A and B to see his/her friend list. The situation is similar in case of D, E, F, G, H, I and J. If and only if they can relate to each other under a common group X where they all are the members of this group X. so, if we can trace a group's propagation into a network, we can find out a social network link where each and every entity corresponds leaf of a big tree, the tree thus be named as master group. Under this master group there lies hundreds to millions of people of common

interest. Thus the significance of utilization of a social network data is that it reveals a real-life scenario of human behavior, the behavior that is meant for whether a person feels a common interest for something or not, whether he/she feels in the same way that other people feel. In total, this network data is a master platform of human propagation.

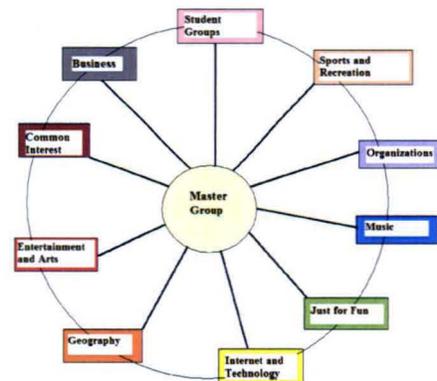


Figure 2: Wagon wheel of Facebook group.

There are 10 different group types in Facebook. Any individual can create a new group under these 10 different types. As a result, if it is possible to create a master group so that these 10 different groups can be a member of that master group, we can easily find out a link of propagation theory. The work is not so difficult indeed, the Facebook itself is the master group, so if we can sort out individual group existence and the members' propagation, we can find out a huge link or network. This work is placed for future research.

We should consider social networking as an important media of communication as it reflects our original life picture. A person's behavior can reflect through a social network. In Facebook, you can send a message to a friend or someone, can chat, can upload pictures, invite to join a party, support a person or not as well as join in a group to exchange your views and ideas. You can even know about a person by the information part of his/her profile in the Facebook, can know every other details if you are not enlisted to limited profile list or the person prevents you to know about him/her by privacy settings. Is it possible to gather/ know/ collect the information about a person by other means? Is it possible to share such a huge thing through other way of communication? The answer is "No", may be you can know by other means, but it will cost you time and money. From this perspective, a social

network like Facebook now requires too many security options- otherwise the network which once you trusted just for communication and thoughts sharing, may turn a suicidal weapon to you.

**Part: B**  
**Proposal of a propagation theory**

All types of social network sites are created with new innovative ideas. The sites vary from each other by contents, types, purpose as well as features. Some sites are created just to share the photographs, some to chat only, some for video and some other for everything in a mix-up condition. We have chosen Facebook for its featuring contents.

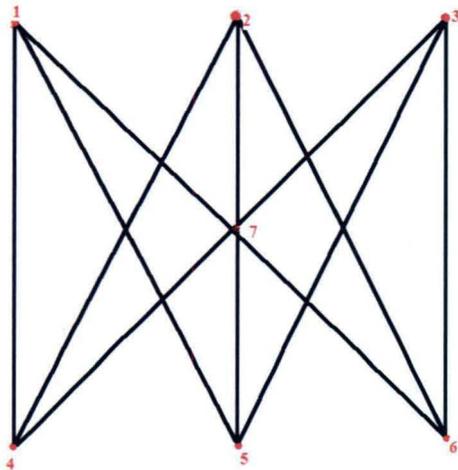


Figure 3: Node connectivity.

In figure 3, node connectivity diagram is shown with 7 different nodes. It is actually a complete bipartite graph [7]. Each node represents an entity of a person. The lines thus represent the connecting group through which an entity can travel to its designated location upon his/her wish. The path of travel of a person thus distributed through different ways and different locations, in some cases with common interest to the same location. If we consider the node connectivity figure in a space and rearrange the nodes with new connecting lines, it will make a new propagation approach within a common field, but with new innovation. So, there are the following options of connectivity:

1. A person can be linked with one or more groups with common interest and
2. A person can be linked without common interest but within the same path of travel.

**Part: C**  
**Analysis and link establishment between social networks and real world**

When a group is established by using a social network (hereinafter Facebook), its existence and popularity changes with the passage of time.

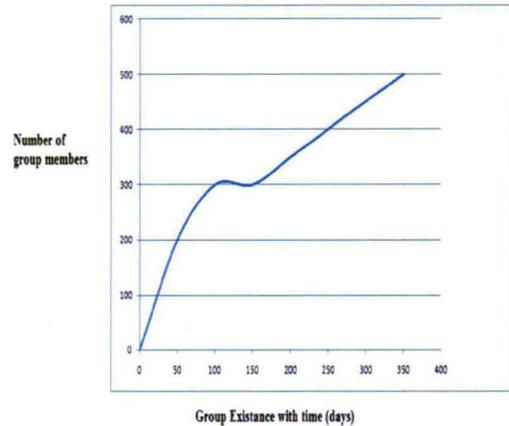


Figure 4: Growth pattern of a group (popular).

A group can gain its popularity within a very short time and also can lose the popularity. It depends on how many people are joining in that very group with common interest. In order to get the group popularity within a social network, we posted a video link about a train accident for trial basis and counted its popularity as the days passed. In the same manner we observed the fan's number for different group activities, comments and photographs. Based on this type of observation and research, we created and simulated some growth pattern graphs. In figure 4, a schematic graph of group growth and popularity is visualized where group popularity increases with time. The simulated figure with different other options of group popularity is shown in figure 5 and 6. In figure 5, the growth popularity changes from popular to unpopular. In figure 6, the growth popularity shows a stable condition. The reasons behind popularity, unpopularity or stability depend on many more things. These reasons' findings are kept for future works.

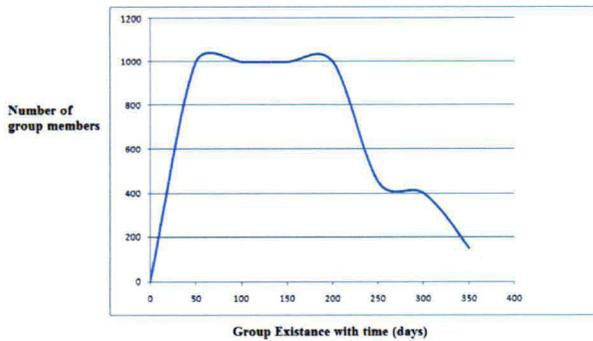


Figure 5: Growth pattern of another group (popular to unpopular).

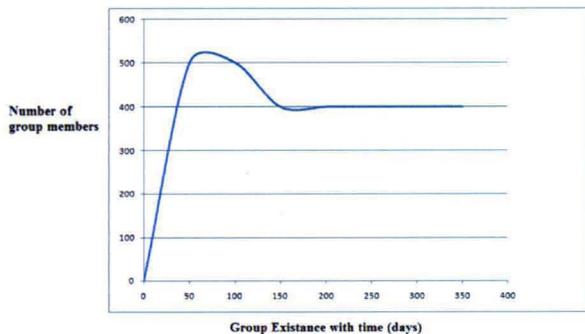


Figure 6: Growth pattern of another group (popular to stable).

The essence of using social network is to be connected with the other people either for reasons or no reasons. In actual life this same thing is happened with a live perspective. Now, social networking sites become so innovative that we can do almost everything through these sites. Our likings, disliking, beliefs, supports, aims and so other things are reflected through these sites. We don't need to go anywhere to wish, support, prevent or say anything by car or by person; we can easily do that with more lively through the sites. Thus social networks are taking the place of real life communication.

### 3. Conclusion and Remarks

The aim of this paper is to analyze a popular social network and classify the human behavior for getting an overall idea about propagation theory. The propagation theory to match with the real life will only make a sensible research work about human behavior classification; we believe to have done this

work. The authors wish to be involved this type of work in future to establish a rigid base theory of human behavior by utilizing existing social networks.

### 4. REFERENCES

- [1] Mislov, Alan and et al. "Characterizing social Cascades in Flickr".
- [2] P. Dodds and D. Watts. A Generalized Model of Social and Biological Contagion. Journal of Theoretical Biology, 2005.
- [3] D. Kempe, J. Kleinberg, and E. Tardos. Maximizing the Spread of Influence through a Social Network. In proc. of ACM SIGKDD, 2003.
- [4] R.M. May and A.L. Lloyd. Infection Dynamics on Scale-Free Networks. Physics review E, 2001.
- [5] R. Pastor-Satorras and A. Vespignani. Epidemics and Immunization in Scale-Free Networks. Wiley, Berlin, 2005.
- [6] Facebook. <http://www.facebook.com>.
- [7] [http://en.wikipedia.org/wiki/File:Complete\\_bipartite\\_graph\\_K3,3.svg](http://en.wikipedia.org/wiki/File:Complete_bipartite_graph_K3,3.svg).

# Effects of Secondary Task Modality and Processing Code on Automation Trust and Utilization during Simulated Airline Luggage Screening

Rachel Phillips; Poornima Madhavan  
Old Dominion University  
rphil019@odu.edu; pmadhava@odu.edu

**Abstract.** The purpose of this research was to examine the impact of environmental distractions on human trust and utilization of automation during the process of visual search. Participants performed a computer-simulated airline luggage screening task with the assistance of a 70% reliable automated decision aid (called DETECTOR) both with and without environmental distractions. The distraction was implemented as a secondary task in either a competing modality (visual) or non-competing modality (auditory). The secondary task processing code either competed with the luggage screening task (spatial code) or with the automation's textual directives (verbal code). We measured participants' system trust, perceived reliability of the system (when a target weapon was present and absent), compliance, reliance, and confidence when agreeing and disagreeing with the system under both distracted and undistracted conditions. Results revealed that system trust was lower in the visual-spatial and auditory-verbal conditions than in the visual-verbal and auditory-spatial conditions. Perceived reliability of the system (when the target was present) was significantly higher when the secondary task was visual rather than auditory. Compliance with the aid increased in all conditions except for the auditory-verbal condition, where it decreased. Similar to the pattern for trust, reliance on the automation was lower in the visual-spatial and auditory-verbal conditions than in the visual-verbal and auditory-spatial conditions. Confidence when agreeing with the system decreased with the addition of any kind of distraction; however, confidence when disagreeing increased with the addition of an auditory secondary task but decreased with the addition of a visual task. A model was developed to represent the research findings and demonstrate the relationship between secondary task modality, processing code, and automation use. Results suggest that the nature of environmental distractions influence interaction with automation via significant effects on trust and system utilization. These findings have implications for both automation design and operator training.

## 1. INTRODUCTION

Automation, as found in many work environments, is seldom used in isolation. For example, luggage screeners are, by the nature of their task, exposed to the sights and sounds associated with large groups of people. Similarly, fighter pilots are often responsible for monitoring many different systems in various locations. Therefore, it is important to understand the influences of concurrently performed tasks (or secondary tasks) on performance, as well as the impact of specific kinds of tasks on automation use. Achieving this objective would allow for the development of training programs, work environments, and system designs which would maximize human-automation potential.

Multiple Resource Theory (MRT) deals with the theory of multiple task performance and derives importance from the prediction of interference between concurrently performed tasks. The most recent version of the model proposes four categorical and dichotomous dimensions: processing stages, perceptual modalities, visual

channels, and processing codes [1], of which we will discuss only the most relevant. MRT postulates that if tasks share a dimensional level, there is greater interference and performance decrement than if the tasks utilize different levels of the same dimension [1], [2], [3].

The perceptual modalities of the MRT are composed of visual and auditory input. It has been found that tasks from different modalities (cross-modal) cause less interference than tasks from the same modality (intra-modal) [4], [5], and that people respond differently to long-term monitoring tasks presented in these two modalities [6]. Processing codes distinguish between analogue/spatial and categorical/symbolic processes. In the model, these modalities are represented as verbal and spatial [1]. The spatial modality is comprised of shape and motion detection [7] as well as sounds, whereas the verbal modality is defined by comprehension of verbal stimuli (either visual or auditory).

Arousal hypothesis basically states that observer alertness is dependent on sensory stimulation [8];

this relationship is generally characterized as an inverted U, with low and high arousal causing performance decrement and moderate arousal leading to the best performance [9]. Over time, if a target occurs rarely among frequent stimuli, accuracy and/or speed of detection generally deteriorates, a phenomenon known as vigilance decrement [10]. In automation interaction, non-vigilance generally manifests as complacency, whereby users assume (incorrectly) satisfactory system state, function, and/or performance, and behave accordingly. Complacency is thought to be influenced by trust, reliability, and confidence [11].

Although automation was developed with the intention of improving human performance, in many instances it has changed the nature of user interaction. Whereas previously, users were responsible for executing a variety of behaviors, now they have been reduced primarily to monitoring positions, interceding only in the event of a problem. Excessive automation has been found to contribute to experiences of sleepiness and fatigue in factory workers [12], as well as decreased arousal and increased frustration in drivers [13]. Decreased arousal is problematic because it is associated with a lack of focus [9], and a decrease in the availability of attentional resources [14], [15].

As stated earlier, vigilance decrement is thought to be a result of lack of sensory stimulation. Sensory stimulation should, therefore, reduce vigilance decrement. Based on this logic, the findings that automation with variable reliability led to greater performance than automation with constant reliability [11] and that adding haptic (tactile) feedback improved performance [16] is not surprising.

Participants completed a luggage screening task with a 70% reliable aid with and without distraction. We hypothesized that tasks which shared two levels (processing code and perceptual modality) with either the primary task or automated aid would cause greater interference than those which shared only one. We further hypothesized that different types of tasks would result in different kinds of interference and this would manifest in different interaction patterns with the automation.

## 2. METHOD

### 1. Participants

Eighty-one undergraduate ODU students participated for partial fulfillment of course credit. All participants were 18 years of age or older and had normal or corrected to normal vision and hearing.

## 2. Materials

### *Visual Search Task*

Participants completed a computer simulation, on two consecutive days, in which they played the role of airline luggage screeners. X-ray images of luggage were presented on a 17 inch color monitor placed approximately 17 inches from the edge of the desk. The computer simulation was developed using Visual Basic for Windows and presented the image, diagnosis of the aid, opportunity for participant input, and feedback. The x-ray images were created using Adobe Photoshop and were comparably cluttered with everyday items (toys, clothes, accessories, etc.). A subset of 20% of the images had one of eight possible knife images digitally superimposed. The participant's task was to indicate which bags contained weapons for one training block (Day 1) of 100 luggage images and two test blocks (Day 2) of 200 luggage images each.

### *Secondary Tasks*

In addition to performing the luggage screening portion of the experiment, for one test block on Day 2, participants were assigned to a secondary task presented in one of two perceptual modalities (visual or auditory) and in one of two processing codes (verbal or spatial). Participants either listened to music (auditory) or read text presented in a text box to the right of the primary task (visual). Participants in the verbal condition were instructed to count the number of times they heard or saw (depending on the modality condition) the word *me*. Participants in the spatial condition were instructed to count the number of times a specific sound occurred in the auditory condition or, for the visual condition, the number of times a specific symbol occurred. Songs were chosen based on the unlikelihood of familiarity with the artist, the catchiness of the tunes, and the relative clarity of the lyrics. The visual condition presented transcriptions of the auditory-verbal condition (in the visual-spatial condition, symbols were substituted for words). All conditions were appropriately counterbalanced.

### *Trust Questionnaire*

The System Trust Scale (STS) (Jian, Bisantz, & Drury, 2000) was utilized to determine how accurate and dependable participants found the automation. The questionnaire was administered at the end of each block. Participants were asked to respond to twelve statements regarding their feelings for the automation on a scale of 1 (strongly disagree) to 10 (strongly agree).

### 3. Procedure

Participants were given an informed consent form which explained their rights as participants and further explained that the study lasted for two days. They completed an entrance questionnaire which obtained demographic information such as age, gender, program of study, experience with computers, etc.

#### *Day 1 – The “Training Phase”*

Participants completed a luggage screening task in which they attempted to detect 20 hidden knife images in 100 x-ray images of luggage. Prior to beginning, participants were shown an example luggage image and the eight knife images they were searching for. Participants were informed that the experiment was timed and that the computer would be keeping score. On each trial, an x-ray image of passenger luggage appeared for three seconds. After the image disappeared, participants either clicked on “stop bag” if they thought a knife was present or on “pass bag” if they thought the knife was absent. They then rated their confidence in the decision on a scale of 1 (not confident at all) to 5 (extremely confident). After completion of the luggage task, participants were reminded to return the following day.

#### *Day 2 – The “Test Phase”*

Exactly 24 hours after the initial portion of the experiment, participants returned to their seats from the previous day. Participants were informed that they were, once again, playing the role of luggage screeners and that the task and targets were the same. It was explained that we were interested in their ability to multi-task and therefore, either in the first or second half of the experiment they would complete a secondary task and the luggage screening task simultaneously. Participants completed two test blocks of 200 images each. The secondary tasks were counterbalanced so half of the participants performed the secondary task in the first test block and the other half completed the secondary task in the second.

In addition, on this day, the participants were assisted by a text-based automated decision aid. The aid provided a diagnosis of knife presence or absence in the form of a text message at the top of the screen at the end of each trial prior to participant input. Unbeknownst to participants the aid was designed to be only 70% accurate.

After the completion of the first trial block of 200 images, participants completed the STS and Secondary Task Questionnaire as appropriate. After a short break, participants resumed the screening

task. After completing the second set of 200 images, participants again completed the scales as appropriate. They were thanked and debriefed before leaving.

### 3. RESULTS

Day one was included in the experiment primarily to facilitate similar baseline levels of performance for the second day. Although training is undoubtedly an interesting and important area of research, it is beyond the scope of this paper.

#### *System Use Measures*

Participants were measured for both compliance (probability of agreeing with the aid when it said target present) and reliance (probability of agreeing with the aid when it said target absent). A 2 (distraction: distracted vs. undistracted) x 2 (modality: visual vs. auditory) x 2 (code: verbal vs. visual) mixed ANOVA for compliance revealed a significant three way interaction,  $F(1, 77) = 3.84, p = .054$  indicating that compliance levels were influenced by distraction differently depending on the distractor modality and processing code. Interestingly, those in the auditory-verbal distraction condition demonstrated a decrease in compliance when distracted ( $M = .703, SE = .037$ ) versus undistracted ( $M = .740, SE = .038$ ) which was contrary to the increase demonstrated by the auditory-spatial (distracted:  $M = .717, SE = .037$ ; undistracted:  $M = .646, SE = .038$ ), visual-verbal (distracted:  $M = .648, SE = .037$ ; undistracted:  $M = .578, SE = .038$ ), and visual-spatial (distracted:  $M = .682, SE = .036$ ; undistracted:  $M = .659, SE = .037$ ) conditions.

A 2 (distraction: distracted vs. undistracted) x 2 (modality: visual vs. auditory) x 2 (code: verbal vs. spatial) mixed ANOVA for reliance yielded a marginally significant three-way interaction between distraction, modality, and processing code,  $F(1, 77) = 3.59, p = .062$ . Participants in the auditory-verbal and visual-spatial condition decreased reliance when distracted (auditory-verbal:  $M = .668, SE = .014$ ; visual-spatial:  $M = .686, SE = .013$ ) versus undistracted (auditory-verbal:  $M = .702, SE = .016$ ; visual-spatial:  $M = .689, SE = .015$ ), but those in the auditory-spatial and visual-verbal conditions increased reliance when distracted (auditory-spatial:  $M = .702, SE = .014$ ; visual-verbal:  $M = .678, SE = .014$ ) versus undistracted (auditory-spatial:  $M = .696, SE = .016$ ; visual-verbal:  $M = .659, SE = .016$ ).

#### *Confidence*

We divided confidence into two variables, confidence when agreeing with the aid and

confidence when disagreeing with the aid. A 2 (distraction: distracted vs. undistracted) x 2 (modality: visual vs. auditory) x 2 (code: verbal vs. spatial) mixed ANOVA for confidence when agreeing indicated a significant main effect for distraction,  $F(1, 77) = 4.89, p = .030$ . Participant confidence when agreeing with the aid actually decreased when distracted ( $M = 2.35, SE = .09$ ) as compared to undistracted ( $M = 2.45, SE = .09$ ).

A 2 (distraction: distracted vs. undistracted) x 2 (modality: auditory vs. visual) x 2 (code: verbal vs. spatial) mixed ANOVA for confidence when disagreeing revealed a significant interaction between distraction and the modality of the secondary task,  $F(1, 77) = 5.13, p = .026$ . This indicated that participant confidence when disagreeing varied as a function of distraction and modality. For the auditory condition, participant confidence levels remained relatively stable when distracted ( $M = 2.26, SE = .13$ ) versus undistracted ( $M = 2.20, SE = .12$ ). However, in the visual condition, participant confidence when disagreeing decreased with the addition of the distraction ( $M = 2.04, SE = .13$ ) as compared to undistracted confidence levels ( $M = 2.20, SE = .12$ ).

#### *Trust Measures*

A 2 (distraction: distracted vs. undistracted) x 2 (modality: auditory vs. visual) x 2 (code: verbal vs. spatial) mixed ANOVA revealed a significant three-way interaction between distraction, distractor modality, and processing code,  $F(1, 72) = 6.32, p = .014$ . For the visual condition, participant trust in the aid increased with the addition of the verbal stimuli (distracted:  $M = 49.9, SE = 4.44$ ; undistracted:  $M = 47.0, SE = 4.58$ ), but decreased with the addition of spatial stimuli (distracted:  $M = 49.3, SE = 4.22$ ; undistracted:  $M = 54.0, SE = 4.36$ ), a pattern directly opposite that demonstrated by those in the auditory condition [spatial: (distracted:  $M = 50.6, SE = 4.44$ ; undistracted:  $M = 46.2, SE = 4.58$ ); verbal: (distracted:  $M = 59.8, SE = 4.69$ ; undistracted:  $M = 63.5, SE = 4.84$ )].

Participants were asked to estimate the reliability of the aid when the target was present. A 2 (distraction: distracted vs. undistracted) x 2 (modality: auditory vs. visual) x 2 (code: verbal vs. spatial) mixed ANOVA revealed a significant interaction between distraction and distractor modality for estimated system reliability,  $F(1, 73) = 4.26, p = .043$ . When distracted, participants in the visual condition estimated slightly higher accuracy ( $M = 56.75, SE = 3.75$ ) than when undistracted ( $M = 55.20, SE = 3.61$ ) and had higher estimates than distracted participants in the auditory condition ( $M =$

$50.17, SE = 4.00$ ) but slightly lower estimates than the auditory participants when undistracted ( $M = 56.94, SE = 3.86$ ).

Participants were also asked to estimate the reliability of the system when the target was absent. A 2 (distraction: distracted vs. undistracted) x 2 (modality: visual vs. auditory) x 2 (code: verbal vs. spatial) mixed ANOVA for estimated reliability when the target was absent revealed no significant main effects or interactions.

#### **4. DISCUSSION**

Performance is impacted by arousal [9]. Over time, vigilance tasks become boring and repetitive, leading to vigilance decrement [10] which manifests with automation as complacency [11]. Complacency can result in potentially dangerous human-automation interaction and is thought to be influenced by trust, system reliability, and user confidence [11]. Our findings suggest that this may be mediated by resource demands resulting from job requirements (simultaneous multiple task performance) or environmental factors (background music, noise, etc.). This would suggest the interaction model shown in Figure 1.

As hypothesized, the perceptual modality and processing code of the secondary task influenced automation use. Based on the MRT [1] we expected that the visual distractor tasks would be more detrimental to performance than the auditory. As indicated by the decrease in confidence when disagreeing with the aid and increase in the perceived reliability of the system with the addition of a visual distractor (see Figure 1), this seems to be the case. Especially when one considers that participants in the auditory conditions actually demonstrated a decrease in perceived system reliability, and an increase in confidence when disagreeing with the aid.

The processing codes of the secondary task further influenced performance depending on the presentation modality. The spatial task was expected to interfere with primary task performance. Contrary to our predictions, participant trust decreased with the addition of the visual-spatial distractor but increased with the addition of the auditory-spatial distractor. It is possible that the visual-spatial task was challenging enough to prevent complacency, increasing participant awareness of automation errors and thus leading to decreased trust. On the other hand, the auditory-spatial condition interfered with primary task performance but not with complacency, leading to increased system trust.

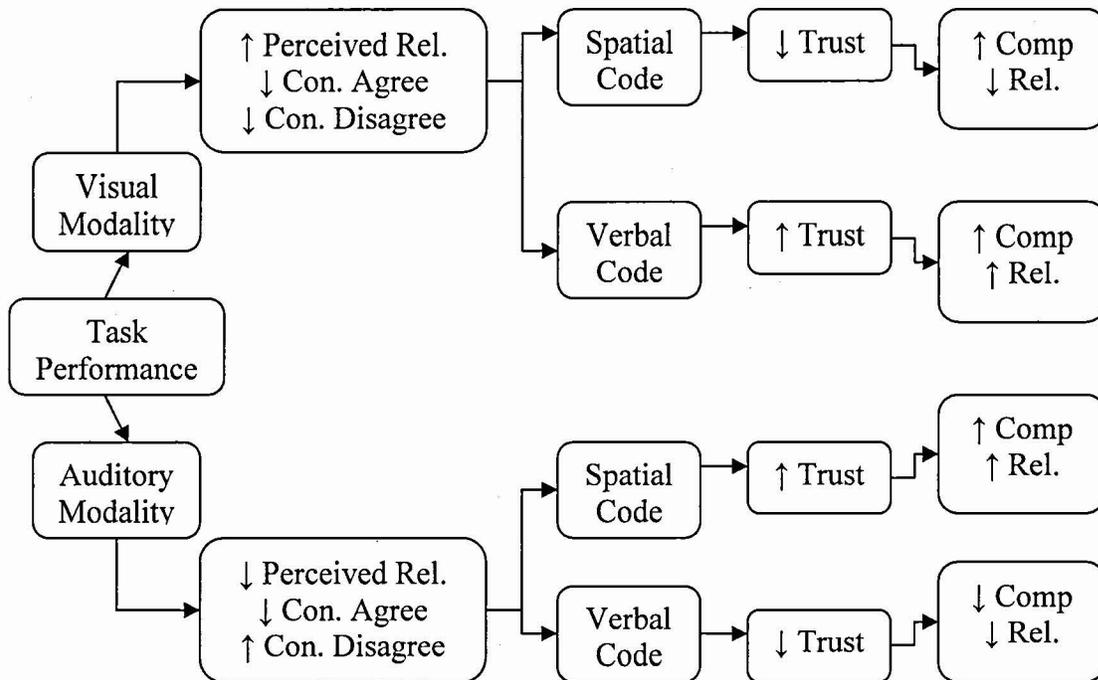


Figure 1. Suggested model for the influence of processing codes and perceptual modality on automation use.

The verbal condition was expected to interfere with aid comprehension. In the visual-verbal condition, participant trust increased, while in the auditory-verbal condition, participant trust decreased (see Figure 1). It is again possible that the visual-verbal condition interfered with task performance but not complacency, leading to increased trust, but the auditory-verbal condition interfered with both task performance and complacency, thereby decreasing trust.

It should be noted that the aid made two kinds of errors, misses and false alarms; however, because we had a 20% weapon base-rate, it actually presented a greater number of false alarms than misses. Additionally, the aid said "target absent" much more than it said "target present" giving participants greater opportunity to demonstrate reliance. It is possible that the difference in compliance (probability of agreeing with the aid when it said target present) and reliance (probability of agreeing with the aid when it said target absent) patterns between the auditory-verbal and visual-spatial conditions may have been due to the nature of the task. However, since reliance decreased (as illustrated in Figure 1) it seems more likely that the results were a product of interactions between perceived reliability, confidence, and trust.

In conditions for which trust increased with the addition of the secondary task (visual-verbal and auditory-spatial), compliance and reliance both

increased. For the visual-spatial condition, compliance increased while reliance decreased suggesting that participants noticed the misses more than the false alarms. The combination of decreased confidence and trust may have also played a role. Because participants were less confident in their own abilities they may have been unwilling to disagree with the aid when it said "target present" leading to increased compliance. However, since they distrusted the aid (despite finding it more reliable), they were more likely to disagree with the "target absent" diagnosis resulting in decreased reliance. In the auditory-verbal condition, participants decreased in both compliance and reliance indicating that they were more leery of the system overall, and more confident in their own abilities.

As shown clearly in the model, participant interaction with the automation differed not only as a function of secondary task modality (visual or auditory) but also as a function of the processing code (verbal or spatial) utilized by the secondary task. Regardless of the reason for the resulting differences in perceived reliability, confidence when disagreeing, trust, and compliance and reliance patterns, these differences demonstrate the inaccuracy of assumptions regarding the stability of human-automation interaction. In order to be used correctly, different automation systems may require a specific pattern of use. Once validated, the model would be useful for designing workspaces and

responsibility combinations to elicit the desired interaction pattern, thus improving human-automation performance.

#### REFERENCES

1. Wickens, C. D. (2002) "Multiple resources and performance prediction," *Theoretical Issues in Ergonomics Science*, vol. 3, no. 2, Apr-Jun, pp. 159-177.
2. Leigh, J. H. (1992) "Modality congruence, multiple resource theory, and intermedia broadcast comparisons: An elaboration," *Journal of Advertising*, vol. 21, no. 2, Jun, pp. 55-62.
3. Smith, R. E. & Buchholz, L. M. (1991) "Multiple resource theory and consumer processing of broadcast advertisements: An involvement perspective," *Journal of Advertising*, vol. 20, no. 3, Sep, pp. 1-7.
4. Sklar, A.E. & Sarter, N.B. (1999) "Good vibrations: Tactile feedback in support of attention and human automation coordination in event driven domains," *Human Factors*, vol. 41, no. 4, Dec, pp. 543-552.
5. St. John, M.; Kobus, D.A.; Morrison, J.G.; & Schmorrow, D. (2004). "Overview of the DARPA augmented cognition technical integration experiment," *International Journal of Human-Computer Interaction*, vol. 17, no. 2, pp. 131-149.
6. Szalma, J. L.; Warm, J. S.; Matthews, G.; Dember, W. N.; Weiler, E. M.; Meier, A.; & Eggemeier, F. T. (2004) "Effects of sensory modality and task duration in performance, workload, and stress in sustained attention," *Human Factors*, vol. 46, no. 2, Sum, pp. 219-233.
7. Weinstein, L. F. & Wickens, C. D. (1992) "Use of nontraditional flight displays for the reduction of central visual overload in the cockpit," *International Journal of Aviation Psychology*, vol. 2, no. 2, pp. 121-142.
8. Davenport, W. G. (1972) "Vigilance and arousal: Effects of different types of background stimulation," *The Journal of Psychology*, vol. 82, no. 2, Nov, pp. 339-346.
9. Bodner, E.; Gilboa, A.; & Amir, D. (2007) "The unexpected side effects of dissonance," *Psychology of Music*, vol. 35, no. 2, Apr, pp. 286-305.
10. Caggiano, D. M. & Parasuraman, R. (2004) "The role of memory representation in the vigilance decrement," *Psychometric Bulletin & Review*, vol. 11, no. 5, Oct, pp. 932-937.
11. Singh, I. L.; Molloy, R.; & Parasuraman, R. (1993) "Individual differences in monitoring failures of automation," *The Journal of General Psychology*, vol. 120, no. 3, Jul, pp. 357-373.
12. Persson, R.; Garde, A. H.; Hansen, A. M.; Ørbæk, P.; & Ohlsson, K. (2003) "The influence of production systems on self-reported arousal, sleepiness, physical exertion, and fatigue—consequences of increasing mechanization," *Stress and Health*, vol. 19, no. 3, Aug, pp. 163-171.
13. Stanton, N. A. & Young, M. S. (2005) "Driver behavior with adaptive cruise control," *Ergonomics*, vol. 48, no. 10, Aug, pp. 1294-1313.
14. Chase, B.; Irwan-Chase, H.; & Sonico, J. (2004) "A methodology to control for individual differences in dual-task performance," *Theoretical Issues in Ergonomic Science*, vol. 5, no. 5, Sep-Oct, pp. 445-451.
15. Young, M. S. & Stanton, N. A. (2002) "Malleable attentional resources theory: A new explanation for the effects of mental under load on performance," *Human Factors*, vol. 44, no. 3, Fall, pp. 365-375.
16. Griffiths, P. G. & Gillespie, R. B. (2005) "Sharing control between humans and automation using haptic interface: Primary and secondary task performance benefits," *Human Factors*, vol. 47, no. 3, Fall, pp. 574-590.

# Automated Discovery and Modeling of Sequential Patterns Preceding Events of Interest

Kurt Rohloff  
BBN Technologies  
Cambridge MA, 02138  
krohloff@bbn.com

**Abstract.** The integration of emerging data manipulation technologies has enabled a paradigm shift in practitioners' abilities to understand and anticipate events of interest in complex systems. Example events of interest include outbreaks of socio-political violence in nation-states. Rather than relying on human-centric modeling efforts that are limited by the availability of SMEs, automated data processing technologies has enabled the development of innovative automated complex system modeling and predictive analysis technologies. We introduce one such emerging modeling technology – the sequential pattern methodology. We have applied the sequential pattern methodology to automatically identify patterns of observed behavior that precede outbreaks of socio-political violence such as riots, rebellions and coups in nation-states. The sequential pattern methodology is a groundbreaking approach to automated complex system model discovery because it generates easily interpretable patterns based on direct observations of sampled factor data for a deeper understanding of societal behaviors that is tolerant of observation noise and missing data. The discovered patterns are simple to interpret and mimic human's identifications of observed trends in temporal data. Discovered patterns also provide an automated forecasting ability: we discuss an example of using discovered patterns coupled with a rich data environment to forecast various types of socio-political violence in nation-states.

## INTRODUCTION

A major challenge in analyzing complex systems is identifying patterns of behavior which are symptomatic precursors to Events of Interest (Eols) such as onsets or terminations of socio-political violence in nation-states. By identifying patterns of behavior that precede Eols such as onsets or terminations of socio-political violence in nation-states we can begin to both understand the underlying causal structures which drive these events to occur and consequently forecast these events.

In this paper we discuss a generalizable sequential pattern concept based on the supposition that the phenomena which cause (or at least are related to) the occurrences of Eols exhibit similar symptomatic behaviors across multiple Eol occurrences. For example, countries experiencing rebellions driven by the desire for freedom by internal ethnic groups commonly exhibit increasing ethnic tension and violence before the occurrence of ethnic rebellions. We formalize our sequential pattern concept using a finite-state machine model of countries' behaviors and use

collections of sampled factor data to define the "states" of a complex system such as a country.

The sampled factor data represents quantifiable measurements of systems such as countries at discrete, regular points in time. We use a discrete clock-tick formalism to model the updating of state locations. Example factors from our socio-political domain include GDP, the rates of occurrence of various words in the national press, the average caloric intake, Goldstein measures of conflict/cooperation between governmental entities, etc. These example factors change continuously over time which motivates our use of sampled data. We map the sampled factor data to observed "trends" in this factor data where a factor's sampled measurement can be increasing, decreasing, or fluctuating over either the short-term or the long-term.

Although there are numerous published works on pattern discovery, the innovation in our approach to pattern discovery comes from our handling of approximate matches necessitated by noisy data in an application context where rigorous matching may not

always be relevant. In particular, we have developed technology to:

- Define loose matching of observed data trends as part of pattern discovery and matching.
- Numerically optimize pattern matching parameters that are Eol-independent for improved forecasting.
- Discover an algorithm to quickly identify loosely matching patterns

Taken together, these innovations enable our pattern discovery and forecasting approach.

In our motivating context of socio-political violence we are interested in patterns that match the trends of observed behaviors preceding at least two instances of Eol occurrences and which are not present in countries when an Eol does not occur over historical data. We have a generalizable, computationally efficient branch-and-bound back-chaining method to identify the set of factors which define a state space in patterns that match the behavior preceding Eol occurrences in at least two countries from historical data. The backwards chaining methodology permits us to identify which factors change similarly for multiple countries for several time steps leading up to the socio-political violence onset or termination in selected countries in a computationally efficient manner.

As a result of our hypothesis that the phenomena which cause (or at least are related to) the onsets and terminations of socio-political violence exhibit similar symptomatic behaviors across multiple onsets and terminations of socio-political violence, we can generate real-time early-warning forecasts of Eols if early portions of the patterns are observed in a specific country. This forecasting process is based around the notion of *matching* a country's behavior to early parts of historical patterns. If the country's behavior matches the early parts of the pattern then we forecast that onsets or terminations of socio-political violence will occur in the country in the near future. We found that this approach to forecasting using single patterns is inadequate in practice because individual patterns provide a limited representation of the full breadth of all possible behaviors that

may precede onsets and terminations of socio-political violence. This motivates our need to generate libraries of patterns that provide a broader representation of the observed preceding dynamics associated with the occurrence of onsets and terminations of socio-political violence.

We demonstrate our pattern discovery and forecasting methodologies over data of onsets and terminations of ethnic-religious violence in Pacific-region countries from 1998-2006. We show that by discovering patterns for ethnic-religious violence onset and ethnic-religious violence termination over Pacific-region countries from 1998-2004, we can use these patterns to forecast ethnic-religious violence onset and ethnic-religious violence termination over 2005-2006 with a very low false-alarm rate.

Previous versions of our pattern discovery approach are presented in [2], [3]. A more in-depth version of the work presented in this paper is provided in [4]. An introduction to our underlying rich data environment to support the experiments discussed here is provided in [1], [5].

## PATTERNS

We define our patterns to be sequences of trends of behaviors observed in factors before Eols. We look for trends in sampled factor data where the sampled factors either increase, decrease or fluctuate over either the short-term or the long-term. For instance, in India in the quarters preceding the onset of ethnic-religious violence in early 2002, we see that the level of cooperating expressed by the government towards opposition parties (as measured by a Goldstein metric) holds fluctuating for several quarters before increasing over a short term and then decreasing shortly before the onset of violence.

With this in mind, we formally define the 6 possible types of trends that can be observed in factors as:

- Long-Term Increasing
- Short-Term Increasing
- Long-Term Fluctuating
- Short-Term Fluctuating
- Long-Term Decreasing
- Short-Term Decreasing

Our definition of patterns around these observed factor trends is one of our innovations in our pattern definition. This approach to pattern definition allows for the loose matching of patterns to observed factor data. This loose matching procedure is generally simple for humans, but exceedingly difficult to automate in a computation environment.

For our application context of socio-political violence we define *short-term* trends as those occurring over 3 quarterly time samples or less, and *long-term* trends occurring over 3 quarterly time samples or more. We allow the definitions of increasing, decreasing, and fluctuating to be system- and factor-specific. In our socio-political violence context, the definitions of increasing, decreasing, and fluctuating vary from country to country and factor to factor. Our motivation for this intuition is that "normal" observed factor behaviors change differently not only from factor to factor (as may be intuitive because different factors measure different phenomena), but that "normal" observed factor behavior varies from country to country for the same factor. As an example, any small change in the level of cooperation expressed by the Chinese government towards potential opposition parties is unusual and significant, but relatively dramatic observed changes in the level of cooperation expressed by the Indian government towards opposition parties is fairly routine.

To map observed changes in factor data to increasing, decreasing or fluctuating trends, we use a weighted threshold test based on the standard deviation of the changes in the factor over a set of training data. This parameter is Eol-independent and is our method for finding this threshold is one of our other innovations in this pattern approach. Our general approach to setting the increasing, decreasing, and fluctuating thresholds is to find the thresholds that would result in maximum forecasting performance over some set of training data.

## FORECASTING

After discovering a set of patterns that precede Eols in complex systems over some training data, we can use these patterns to make out-of-sample forecasts for the Eols over test data. We found that a

relatively simplistic approach to forecasting is generally very effective - we used a weighting voting mechanism where the discovered patterns matched out-of-sample observations in the test data to generate forecasts.

To implement our weighted voting mechanism in our socio-political violence domain for a given country at a given time, we determine which patterns match the observed factor data leading up to that time. If the number of patterns matching the data exceeds a voting threshold  $v$ , then we forecast the onset/termination of socio-political violence in that country at that time. Similar to the weight threshold for increasing, decreasing and fluctuating, we compute  $v$  to maximize forecasting performance over some training data that wasn't also used for pattern discovery.

## EXAMPLES OF FORECASTING ONSET AND TERMINATION OF ETHNIC-RELIGIOUS VIOLENCE

Using the trend weight threshold  $w$  and the voting threshold  $v$  that maximized the  $f$ -measure of forecasts for the onset of coups in our training data, we applied our approach to forecast the onset and termination of ethnic-religious violence in Pacific-region countries. We ran this experiment to forecast the onset and termination of ethnic-religious violence using a set of quarterly sampled Goldstein metric factor data that expressed the relative levels of conflict/cooperation between political groups operating in the countries (such as the government, opposition parties, international organizations, etc...).

We split our data into training and test sets. The training data ran from 1998-2004 and the test data ran from 2005-2006. Over the training data there were onsets of ethnic-religious violence in the following countries at the following times:

- China Q1-2004
- India Q1-2002
- Indonesia Q1-1999
- Solomon Islands Q1-2000
- Solomon Islands Q1-2003
- Sri Lanka Q1-2003

Similarly, there were terminations of ethnic-religious violence in the following countries at the following times:

- India Q1-2004
- Solomon Islands Q1-2001
- Solomon Islands Q1-2004

Using these two sets of events and the  $w = 0.2$  threshold, we discovered 55 single-factor patterns for the onset of ethnic-religious violence and 19 single-factor patterns for the termination of ethnic-religious violence. When then used these patterns to forecast the onset/termination of ethnic-religious violence using our threshold voting mechanism.

For the onset of ethnic-religious violence, we generated the following forecasts:

- India Q1-2005
- Nepal Q2-2005
- Taiwan Q4-2006

Over the 2005-2006 test data, the only onset of ethnic-religious violence is in India in the beginning of 2005. There are no true ethnic-religious violence outbreaks in Nepal or Taiwan so we generated two false-positive forecasts. It is interesting to note however, that in early to mid-2005 in Nepal there was an uptick in the level of violence associated with the smoldering Maoist insurgency in that country.

For the onset of ethnic-religious violence, we generated the following forecasts:

- \item China Q1-2005
- \item Sri Lanka Q1-2006
- \item Sri Lanka Q3-2006

Over the 2005-2006 test data, the only termination of ethnic-religious violence is in China in the beginning of 2005. There are no true ethnic-religious violence terminations in Sri Lanka so we generated two false-positive forecasts. It is similarly interesting to note however there was a dip in the ongoing ethnic Tamil insurgency in the beginning of 2006 in Sri Lanka, but this violence picked up again several months later.

For both forecasting both the onset and termination of ethnic-religious violence, we were able to forecast both occurrences of onset/termination correctly along with a reasonably low false-positive rate. Because

we were forecasting over 29 countries and two years, the false-positive rate is approximately one false alarm every 100 country-quarters.

### **A RICH DATA ENVIRONMENT TO SUPPORT PATTERN DISCOVERY AND FORECASTING**

In implementing our socio-political violence forecasting methodology, our experimentation environment was supported through the application of emerging information technologies to run our pattern discovery and forecasting methodology. Key technologies in our environment include automated data collection, knowledge representation, model integration and data visualization. To support our pattern discovery and forecasting activities we constructed an end-to-end distributed knowledge system that supports:

- Automated collection and classification of unstructured data (such as raw news feeds and communications data); and collection of structured data.
- Automated fusing of extracted information with historical structural and geospatial datasets using Semantic Web technologies to support distributed modeling and analysis.
- Context-dependent data visualizations, including faceted browsing and spatial-temporal displays, to reveal underlying structures, patterns, and correlations.

Our system utilizes technologies for automated collection and classification of unstructured data including a set of named entity, relationship, and event extraction capabilities that operate over the entire content of articles and can "learn" or evolve over time. We applied these technologies to analyze a voluminous corpus of news feed data that covered a wide geographic region over a period of ten years. The extracted information formed the basis of theory-based independent variables (such as general tension metrics, non-state actor attributes, and leadership characteristics) as well as augmenting the more stale historical factors extracted from existing social science and econometric datasets.

Using Semantic Web technologies, our knowledge system fuses information extracted via the natural language tools

described with data from numerous social science datasets to develop a knowledge environment. This capability also provides a basis for automated reasoning and inference for model and analysis results integration. By directly encoding the semantics typically stored in dataset codebooks, the system fuses multiple datasets that goes beyond superficial dataset alignment (such as merely sorting data from various datasets by county and year). Our system's knowledge infrastructure supports model-agnostic access to the stored data for further manipulation and analysis. Modeling and analysis results can be fed back into the knowledge system for access by other models.

The capstone interface of our knowledge system employs data visualization techniques to display data analysis results and provide interactive "drill-down" capabilities to better study results. Faceted browsing of factors and patterns based on these data values allows a user to select different events and the associated variables associated with the events in various countries.

## DISCUSSION

The sequential pattern methodology is an approach to identify temporal patterns of behavior that precede Eols such as multiple occurrences of socio-political violence such as riots, rebellion onset, coups, etc. Our current development of the sequential pattern methodology mines regularly sampled factor data collected from our data collection processes and aggregated by our Semantic Web technology. This methodology identifies commonalities preceding multiple events of interest occurrences but not before non-occurrences. The patterns rely on the pre-defined of equivalence classes of factor values. Our latest experimentation with the sequential pattern methodology relies on equivalence classes defined by both static and dynamic quantization operations.

The sequential pattern methodology is a groundbreaking approach to pattern discovery because it generates easily interpretable patterns based on direct observations of sampled factor data. The resulting patterns are easily to visually interpret as timed finite-state-machine

models. Additionally, they permit a clear audit trail to guide the replication of discovered patterns and for the forecast of political instability based on already discovered patterns. We foresee additional development of our sequential pattern technology to natively incorporate discrete-event behaviors to capture the underlying "mode-switching" behavior inherent in many processes.

An important property of our sequential pattern approach is its generalizability. The sequential pattern methodology can be applied in other complex system application contexts where patterns of behavior in sampled factor data preceding events needs to be identified and represented.

## REFERENCES

1. Battle, R., D. Reid & K. Rohloff. "CWEST: Disruptive Integration of Computation Technology for Data Analysis and Visualization." *Visualizing the Past: Tools and Techniques for Understanding Historical Processes*, February 2009.
2. Rohloff, K. & A. Victor. "The Identification of Sequential Patterns Preceding the Occurrence of Political Events of Interest." *Second International Conference on Computational Cultural Dynamics*, September 2008.
3. Rohloff, K. & A. Victor. "Computational Methods to Discover Sets of Patterns of Behaviors that Precede Political Events of Interest." *AAAI Spring Symposium on Technosocial Predictive Analytics*, March 2009.
4. Rohloff, K. "Trend Pattern Approach to Forecasting Socio-Political Violence." *Third International Conference on Computational Cultural Dynamics*, September 2009.
5. Rohloff K. and W. Thornton. "A Knowledge Environment for Social Science Exploration." *Human Behavior-Computational Intelligence Modeling Conference*, June 2009.



# Representation of Cultural Role-Play for Training

Thomas Santarelli, Aaron Pepe, Ph.D., Larry Rosenzweig, John Paulus, Ahn Na Yi  
*CHI Systems, Inc.*

{*tsantarelli, apepe, lrosenzweig, jpaulus, ahnnayij*} @chisystems.com

**Abstract.** The Department of Defense (DoD) has successfully applied a number of methods for cultural familiarization training ranging from stand-up classroom training, to face-to-face live role-play, to so-called smart-cards. Recent interest has turned to the use of single and multi-player gaming technologies to augment these traditional methods of cultural familiarization. One such system, termed CulturePad, has been designed as a game-based role-play environment suitable for use in training and experimentation involving cultural role-play scenarios. This paper describes the initial CulturePad effort focused on a literature review regarding the use of role-play for cultural training and a feasibility assessment of using a game-mediated environment for role-play. A small-scale pilot involving cultural experts was conducted to collect qualitative behavioral data comparing live role-play to game-mediated role-play in a multiplayer gaming engine.

## Introduction

The recent shift in DoD focus from high intensity conflicts to the preparation for Stability, Security, Transition, and Reconstruction (SSTR) operations has many profound consequences on how we train U.S. forces for present and future missions. To support this shift, the U.S. military has recognized the importance of cultural and negotiations training and has applied several training approaches, such as live role-play exercises that typically utilize native Arabic role-players and focused training classes. These training methodologies have been used at the Joint Readiness Training Center (JRTC) at Ft Polk, National Training Center (NTC) at Ft. Irwin, Mojave Viper at Twenty-nine Palms, and Camp Pendleton.

Recent game-based cultural training approaches have focused on two types of computer-based interactions. The first provides cultural familiarization through role-play with artificial intelligence (AI)-controlled cultural avatars [1, 2, and 3]. The AI method has the benefit of making cultural familiarization training accessible on demand without requiring cultural experts to be present during training. However, the AI-controlled characters have a limited ability to adapt to novel situations.

The second computer interaction approach has been to create multiplayer-enabled simulation environments. These enable virtual role-play among multiple human role-players [4, 5] and make it possible for each user to have an avatar representing their own persona to engage in distributed role-play.

Of concern with this recent interest in using game based technologies is the limited research

comparing the effectiveness of a game-mediated system versus traditional (i.e., live face-to-face) role-play, particularly in a cross-cultural training context. To address this need, our effort was divided into four steps:

- Conduct a literature review to identify potential risks and limitations involving the use of role-play to explore complex cross-cultural issues irrespective of mode (i.e., face-to-face or game-mediated), as well as issues specifically involving game-mediated cultural interactions.
- Develop effective role-play strategies to mitigate the challenges involved in using a game-based platform for cultural training.
- Develop training materials to support live and virtual role-play and modify a game-based environment to enable effective culturally focused interaction.
- Conduct a pilot study to compare trainee satisfaction with the game-based role-play simulation versus satisfaction with live interaction.

In the following pages, we provide a synopsis of the literature review, the process for conducting the pilot, and a brief review of the pilot results. We end the article by providing conclusions derived from the pilot study results and offering guidance for revision and further testing of key relevant issues.

## Use of Role-Play for Cultural Training

We began the literature review by identifying four areas in which to focus our investigation. The first three involve general issues regarding

use of the role-play method, while the forth concerns our specific interest in employing a game-based platform for role-play. These areas of investigation include: Use of role-play to elicit cross-cultural behaviors, sociological and cultural issues related to group interactions, interactions involving verbal and non-verbal forms of communication, and the use of a game-environment as a mediating form of interaction.

We developed literature-based mitigation approaches which guided implementation and testing of the game-based cultural role-play software. A summary of the key issues identified from the literature review is provided in the tables below. Table 1 identifies key issues involved with role-play in general, while Table 2 describes issues inherent in game-based role-play. In both tables the first column enumerates the salient issues identified during the review, while the second column includes our current risk mitigation approach or hypothesis, along with links to supportive literature.

It became clear through the literature review that our mitigation strategies would focus around two key themes. The *first* theme involved mitigating potential issues regarding use of role-play to explore cross-cultural issues and behaviors in both face-to-face and game-mediated role-play. This included such issues as: cognitive load imposed on role-play participants, factors related to optimal role-play, cross-cultural teams participating in a role-play context to explore cross-cultural scenarios, and typical inhibitions while participating in role-play related to group size or lack of anonymity. The *second* theme involved the need to provide effective user-interface metaphors to facilitate effective role-play and embodiment in game-mediated role-play [6]. This included specific issues such as complexity of mapping user inputs to non-verbal and other avatar behaviors, questions regarding required level of avatar fidelity required to convey subtle non-verbal communication, and mitigating cognitive load imposed by the game-environment.

An important finding from the literature review indicated that the degree of preparation has a direct, positive influence on the effectiveness of role-play [7, 8]. An integral component of preparation includes the creation of the role-play scenario itself including: specification of underlying scenario objectives, roles, motives, background information, and other features of the role-play scenario.

### Pilot - Live vs. Game Cultural Role-Play

From these scenario specifications, collateral materials can be created and provided to role-play participants as 'role-play' guides. Further, the need to support directorial guidance to facilitate effective role-play was borne-out in the literature [9]. An important objective was to derive a framework that could be used to form the basis of standardized CulturePad scenario authoring guidelines. These guidelines would ultimately direct the creation of scenario-specific 'role-player guides' to drive role-play within the pilot but would also generalize as a descriptive model for application in other role-play contexts.

For the purposes of this cultural simulation, we designed a 20-minute scenario that required four participants, one for each role within the scenario. Further, each of the participants engaged in the role-play scenario in each of the two conditions (i.e., live and game-mediated).

The context for the role-playing in both conditions involved a scenario that requires proper Mosque etiquette. Within the scenario, a U.S. military representative, "Lt. Armstrong," was required to obtain intelligence from two local residents "Sheikh Mahmood," and his nephew, "Salim." Lt. Armstrong's objective was to obtain intelligence information regarding a person of interest with possible ties to recent violence within the city, while maintaining culturally sensitive etiquette.

The live role-play (see Figure 1) was conducted at a mosque located in southern Florida. Three of the participants had expertise in Arabic culture, but the fourth did not, thereby setting up a context in which cultural 'friction-points' could more realistically surface



Figure 1: Pilot - live role-play.

**Table 1: Issues related to role-play in general**

General Issues	Mitigation Hypotheses & Supportive Literature
Role-play between cross-cultural users may require more preparation or practice than a culturally homogeneous group.	Integrate team-building exercises into the role-play exercises. Vogel et al. [10] suggested that for intercultural computer-supported collaboration, team-building and workshop exercises increase collaboration effectiveness.
Lack of anonymity in face-to-face role-play may prevent engagement.	Providing a game-mediated approach can support an inherent form of anonymity during exploration of controversial cultural issues. Games can provide anonymity in role-play simulations [11, 12].
Large group size may cause social inhibition during role-play.	Minimize the size of the role-play groups. Given typical group dynamics, some have suggested that four to five players is the most preferred and often-cited size for a role-play group [13, 14]
Role-play may represent a threatening context for some players to explore cultural issues.	Use of a game environment may provide a safer arena for role-players to confront cross-cultural issues than a face-to-face condition. Game environments provide a safe, nonthreatening environment for individuals to interact and confront cultural differences [15, 16, and 17].
Role-play can produce "high cognitive load" particularly for novices.	Implement intelligent scenario authoring support to guide a scenario author in creating well-structured scenarios to optimize the execution of role-play. Role-player difficulties are minimized under conditions in which there is a well-structured role-play scenario and well-prepared role-play facilitators [7, 8].

**Table 2: Issues related to game-mediated role-play**

Game-Based Issues	Mitigation Hypotheses & Supportive Literature
Conveyance and interpretation of emotional disposition may not be intuitive.	Use natural verbal communication channels via Voice over Internet Protocol (VoIP). Research suggests that information in the human voice encodes information concerning the speaker's emotional state that is driven by involuntary physiological changes to the speaker's speech production as well as interpretation and encoding of cultural information [18].
Nonverbal behaviors may be difficult to map to physical and graphical control inputs.	Use in-game graphical affordances to create a cross-cultural interaction graphical metaphor (e.g., iconography). Evidence suggests that users are creative in establishing norms for encoding and transmitting nonverbal cues in alternative ways [19, 20, and 21].
Fidelity of visual avatars may not be adequate to convey required nonverbal communication in a game-mediated environment.	Establish fidelity requirements experimentally. If required, license premade avatars, as they provide high levels of visual and animation fidelity. Tromp and Snowdon [6] have shown that even visually crude block-forms can be successful in communicating nonverbal social cues.
Cross-cultural user populations have culture-specific interpretation for use of color, user preferences, and icons.	Use a configurable Human-Computer Interface (HCI) design approach that can be regionalized based on the user demographics. Design preferences vary across cultures and are culture-specific [22]. Also, analyses can be used to understand practical design considerations that exist for individual cultures of interest [23]

In addition to the military representative and the two local residents, there was a final role-player who acted as a “director”. The intention of the director was to act as a facilitator so that as the role-play unfolds (in either condition), the director can inject dialog, actions, and events to “move things along.” Each session (live action and game-based) lasted approximately 20 minutes. Each role-player was provided only with the role-player guide handout for their specific role, whereas the director was cognizant of all role-player handout content.

Cognizant of the issues and mitigation approaches synthesized from the role-play training literature, described in Tables 1 and 2, CulturePad used a modified version of a game-based training simulation for use in the game-mediated condition. This system, termed the Virtual Environment Cultural Training for Operational Readiness (VECTOR [24, 25]), was developed by CHI Systems to enable trainees to virtually interact with members of another culture through cultural familiarization scenarios. VECTOR users select dialogue options in order to interact with game characters controlled by Human Behavior Representations (HBRs) which are designed to reflect a Kurdish Arabic culture.



**Figure 2:** Pilot – game role-play.

The modified game-based simulation allows multiple users to control avatars behaviors via emoticons and behavior iconography. In addition, it has the added capabilities to record behavioral data from role-play sessions and support direct verbal communications through integration of Voice over Internet Protocol (VoIP). It also provides a scenario-authoring component to standardize CulturePad scenario development and create scenario-specific role-player guides for participants.

#### 4. Pilot Results

The initial feedback about live-action and game-based simulation confirmed that preparation is key to a successful role-play experience.

Further, results suggest that although it has some limitations, game-based simulation is realistic, engaging, and an appropriate method to observe and practice culturally oriented behaviors.

A summary of the subject responses to the pilot questionnaire Likert-scale ratings and open-ended questions is provided in the table below. The scale ranged from 1 (Strongly Disagree) to 6 (Strongly Agree) with statistical means for each question provided below.

**Table 3:** Sample of participants’ responses to the pilot study questionnaire

Question	Mean
Role-playing scenario was realistic.	5.0
Gestures of the avatars were realistic.	4.8
In game Mosque added to fidelity of role-playing.	4.8
Scenario briefing was adequate.	5.0
I believe the game-based approach could be useful in eliciting culturally oriented behaviors.	4.5
Pressing action buttons is a viable approach.	4.0
I felt as engaged in game-based role-playing mode as I did in the face-to-face role-playing.	3.5

#### Open-Ended Question Responses

Talking seemed natural, but showing emotion was difficult.
It was difficult to express emotion in the game with the included head gestures, body language, and facial expressions.
I could express my anger better in face-to-face.
Increase time for practice.
Increase available body language.
Split screen so you can see yourself and the others.

#### 5. Conclusions

Because of an increased need for cultural familiarization training for U.S. military forces, researchers have begun to test whether distributed multiplayer training environments would be an effective approach to provide cultural training role-play to a greater audience. The work described here investigated whether a game-based solution could reproduce the live role-play cultural training experience effectively.

Based on our observations and the results of the pilot, we found that two key elements are responsible for an effective application of cultural role-play. First, proper preparation and guidance in creating and enacting role-play based training are important, regardless of medium. Second, the type of game-based

interface is significant in enabling the user to convey and appreciate the cultural and interpersonal aspects of the role-play interaction. An important finding from our pilot was that specific cultural behaviors and gestures are transportable from a realistic face-to-face encounter to a game-based environment hosted on a computer.

Current efforts are under way to explore multiple interface configurations to increase the immersion and cultural training effectiveness of simulated role-play. We are also investigating input modality, in addition to the visual human-computer interface. Nontraditional game-based controllers such as the Wii remote are being examined in addition to more established devices such as keyboard, mouse and joystick-button combinations.

Although these various interaction methods are being studied for their ability to improve current training effectiveness and user acceptance, direct-mapping-based technology will be evaluated for its use in future systems. Avatar control that allows a direct mapping of users' intent and action – such as those involving eye-tracking, gesture-tracking, and physiological measurement – could provide a more seamless interaction experience. These technologies, though not currently practical for wide distribution because of issues such as high cost, intrusiveness, and imprecision, provide interesting options for the future, as they show promise for reducing the cognitive load [26] associated with role-play.

Additional research is needed to identify the specific learning principles and cultural competencies that apply to the use of gaming environments for cross-cultural role-play as a training medium. For example, perspective-taking is increasingly acknowledged as an important skill with regard to cross-cultural competency [27, 28]. We hypothesize that a game-based role-play environment may facilitate perspective-taking as learners utilize extraction skills to interpret subtle cultural cues from alternative roles within the game environment.

## 6. Acknowledgments

The work reported in this article was performed under contract to the U.S. Army Research Institute (contract number # W91WAW-08-C-0053) monitored by Don Lampton and Bruce Knerr. Important contributions to the reported work have also been provided by CHI Systems staff members Joanne Barnieu, Amanda Hafich, and Heather Lum. We extend our sincere appreciation to our cultural expert, Muzaffar Shaikh. The judgments and opinions presented

in this paper are those of the authors and do not represent official positions of the U.S. Army.

## REFERENCES

- [1] Barba, C., Deaton, J., Santarelli, T., Kerr, B., Singer, M., & Belanich, J. (2006). November Virtual environment composable training for operational readiness (VECTOR). In *Proceedings of the 25th Army Science Conference*. (pp. 25-32). Norfolk: World Scientific Pub Co Inc.
- [2] Hill, R. W., Belanich, J., Lane, H. C., Core, M. G., Dixon, M., Forbell, E., Kim, J., & Hart, J. (2006, November). Pedagogically structured game-based training: Development of the ELECT BILAT simulation. In *Proceedings of the 25th Army Science Conference*. (pp. 82-95). Norfolk: World Scientific Pub Co Inc. Orlando, FL: U.S. Office of the Under Secretary of Defense Acquisition, Technology and Logistics, Orlando FL.
- [3] Johnson, W. L. (2007). Serious use of a serious game for language learning. In R. Luckin et al. (Eds.), *Artificial intelligence in education* (pp. 67–74). Amsterdam: IOS Press.
- [4] Benton, D., & Santarelli, T. (June, 2008). Modeling cultural behaviors via multi-player gaming environments. Paper presented at the *International Training and Education Conference*, Stockholm, Sweden.
- [5] Mayo, M., Singer, M. J., & Kusumoto, L. (2005, December). Massively multi-player (MMP) environments for asymmetric warfare. In *Proceedings of the 27th Interservice/Industry Training Systems and Education Conference*. (pp. 220-231). New York; ACM Transactions on Modeling and Computer Simulation. Arlington, VA: National Training Systems Association.
- [6] Tromp, J., & Snowdon, D. (1997) Virtual body language: Providing appropriate user interfaces in collaborative virtual environments. *ACM Symposium on Virtual Reality Software and Technology* (pp. 37-44). New York: ACM.
- [7] Van Merriënboer, J., Schuurman, J., De Croock, M., & Paas, F. (2002). Redirecting learners' attention during training: Effects on cognitive load, transfer test performance and training efficiency. *Learning & Instruction*, 12(1), 11–37.
- [8] Sisk, D. A. (1995). Simulation games as training tools. In S. M. Fowler & M. G. Mumford (Eds.), *Intercultural sourcebook: Cross-cultural training methods* (Vol. 1; pp. 81–92). Yarmouth, ME: Intercultural Press.
- [9] Flowers, A. Magerko, B., and Mishra, P. (2006). Gamemasters and Interactive Story: A Categorization of Storytelling Techniques in Live

- Roleplaying. *Futureplay*, 2006. London, Ontario.
- [10] Vogel, D., van Genuchten, M., Lou, D., Verveen, S., van Eekhout, M., & Adams, T. (2001). Exploratory research on the role of national and professional cultures in a distributed learning project. In *IEEE Transactions on Professional Communications*, (pp. 114–125).
- [11] Freeman, M. A., & Capper, J. M. (1999). Exploiting the web for education: An anonymous asynchronous role simulation. *Australian Journal of Educational Technology*, 15(1), 95–116.
- [12] Raybourn, E. M. (2002). Toward cultural representation and identification for all in community-based virtual environments. In N. Carbonell & C. Stephanidis (Eds.), *Lecture Notes in Computer Science* (pp. 219–238) Berlin: Springer Berlin / Heidelberg.
- [13] Rosenberg, A. (2002). Introduction: Before the game. In *Gamemastering secrets* (pp. 3-22, 2nd ed.). Randolph, MA: Grey Ghost Press.
- [14] Bernard, R. M., de Rubalcava, B. R., & St-Pierre, D. (2000). Collaborative online distance learning: Issues for future practice and research. *Distance Education*, 21(2), 260–277.
- [15] Pedersen, P. (1995). Simulations: A safe place to take risks in discussing cultural differences. *Simulation & Gaming*, 26(2), 201–206.
- [16] Kass, A., Burke, R., Blevis, E., & Williamson, M. (1993). Constructing learning environments for complex social skills. *Journal of Learning Sciences*, 3, 387–427.
- [17] Gee, J. (2003). *What video games have to teach us about learning and literacy?* New York: Palgrave Macmillan.
- [18] Scherer, K. R. (1986). Vocal affect expression: A review and a model for future research. *Psychological Bulletin*, 99, 143–165.
- [19] Cherny, L. (1999). *Conversation and community: Chat in a virtual world*. Stanford, California: CSLI Publications.
- [20] Churchill, E. F., & Bly, S. (1999). Virtual environments at work: Ongoing use of MUDs in the workplace. In *Proceedings of the International Conference on Work activities Coordination and Collaboration* (pp. 99-108). New York: ACM.
- [21] David B. H., & Brewer, J. P. (1997). Electronic discourse: Linguistic individuals in virtual space. *Educational Technology*, 32(1), 53–67.
- [22] Evers, V., & Day, D. (1997). The role of culture in interface acceptance. In S. Howard, J. Hammond, & G. Lindegaard (Eds.), *Human Computer Interaction, Interact '97* (pp. 4-12). London: Chapman and Hall.
- [23] Barber, W., & Badre, A. N. (2001, June). Culturability: The merging of culture and usability. In *Proceedings of the 4th Conference on Human Factors and the Web*. (pp. 15-21). New York: Assn for Computing Machinery, Basking Ridge, New Jersey, USA
- [24] McCollum, C., Deaton, J., Barba, C., Santarelli, T., Singer, M., & Kerr, B. (2004, December). Developing an immersive, cultural training system. In *Proceedings of the 27th Interservice/Industry Training Systems and Education Conference*. (pp. 52-61).
- [25] Santarelli, T., McCollum, C., Barba, C., & Deaton, J. (2004). Applying a cognitive architecture to control of virtual non-player characters. In *Proceedings of the 2004 Winter Simulation Conference* (pp. 890).
- [26] Abbe, A., Gulick, L. M., & Herman, J. L. (2008). *Cross-cultural competence in army leaders: A conceptual and empirical foundation* (ARI Study ReportSR 2008-01). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.
- [26] VanLehn, K., Ball, W., & Kowalski, B. (1989). Non-LIFO execution of cognitive procedures. *Cognitive Science*, 13, 415–465.
- [27] Rentsch, J. R., Gunderson, J., Goodwin, G. F., & Abbe, A. (2007). *Conceptualizing multicultural perspective taking skills* (TR 1216). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (ERIC Document Reproduction Service No. ADA475106)

# Modeling Social Influence via Combined Centralized and Distributed Planning Control

James Vaccaro, Staff Research Scientist; Clark Guest, Associate Professor

Lockheed Martin Corporation; University California San Diego

[jim.vaccaro@lmco.com](mailto:jim.vaccaro@lmco.com); [cquest@ucsd.edu](mailto:cquest@ucsd.edu)

**Abstract.** Real world events are driven by a mixture of both centralized and distributed control of individual agents based on their situational context and internal make up. For example, some people have partial allegiances to multiple, contradictory authorities, as well as to their own goals and principles. This can create a cognitive dissonance that can be exploited by an appropriately directed psychological influence operation (PSYOP). An Autonomous Dynamic Planning and Execution (ADP&E) approach is proposed for modeling both the unperturbed context as well as its reaction to various PSYOP interventions. As an illustrative example, the unrest surrounding the Iranian elections in the summer of 2009 is described in terms applicable to an ADP&E modeling approach. Aspects of the ADP&E modeling process are discussed to illustrate its application and advantages for this example.

## Introduction

We propose using an Autonomous Dynamic Planning and Execution (ADP&E) approach that integrates both a centralized and distributed planning control capability to more realistically model complex social group interactions. In our recent survey of implemented models within social science, they do not successfully model future influence operations because they do not integrate enough cognitive realism in each automated-human (agent) to represent real world conditions and events. This makes the current models unsuitable for large-scale, complex problem domains. More specifically, implemented models fail to capture several aspects of human behavior because these models do not include the ability to adjust to very large, partially observable, and uncertain environments, nor use human abilities in dynamic planning to maintain agility in these ever-changing environments.

In addition, many techniques assume a completely distributed (decentralized) approach that uses simplified cognitive agents with common goals to create swarm-like behavior [1]. This leads to emergent events when the cumulative cognitive state reaches a tipping point. In the same context, other techniques rely on completely centralized control of agents to optimize their coordination and lead to more optimal strategies of cooperative event behavior, which can suspend reactions of discontent and generate strong unified positions [2]. Both of these approaches are goal-directed, but the centralized approach relies more on reputational or social utility, while the distributed approach relies more on intrinsic or expressive (i.e., individual or psychological) utility.

Real world events are actually driven by a mixture of both centralized and distributed control of individuals (agents) based on their situational

context and internal makeup. Given the level and type of education, age, interests, experiences, religious affiliation, economic status, etc., individuals have varying degrees of both centralized and distributed behavioral influences that either enhances or detracts from their current environmental status or cross-cuts their current environmental circumstances. For example, some people may have partial allegiances to multiple contradictory authorities (e.g., religious vs. science, dictator vs. democracy, etc.), which could create a cognitive dissonance within these people.

This further could create an opportunity for change, given their uncertainty in their future, and their willingness to seek change from their current conditions. Does this form an opportunity for external forces to intervene and pursue a psychological influence operation (PSYOP) to redirect the event toward a change beneficial to its interests, or does meddling at such a time backfire and strengthen the opposition's claims and perhaps tip the balance in our adversaries favor? An autonomous dynamic planning and execution (ADP&E) framework has been built that includes variability in searching, selecting, and rewarding plans based on both individual and group behavior. Difficult questions such as this PSYOP mentioned above can be addressed in modeling and simulation if centralized and distributed planning are successfully integrated within the model via this ADP&E framework. They will thus better model the balance of using both centralized and distributed planning-influence control and further understand its sensitivity through simulating interactions among similar and differing social groups with differing parameter sets.

## Background

Currently implemented cognitive approaches can be analyzed from a game theory perspective to determine their problem domain footprint. On

the one hand, reactive planning algorithms, such as temporal difference reinforcement learning can learn two-player stochastic games, such as Backgammon [3]. On the other hand, deep search algorithms, such as decision-tree search using alpha-beta pruning can plan many moves ahead for a two-player deterministic game, such as chess [4]. However, note that these games are both two-player and fully observable, while the real-world is many players and partially observable. Further, hybrid solutions have been proposed to handle more complex real world and game problems [5]. We propose using a more powerful hybrid approach that integrates more realistic features of social interaction by extending an ADP&E approach with both a centralized and distributed planning capability.

An illustrative example will be investigated to better model and predict cumulative behavior amongst more cognitively realistic agents based on their interaction. The analyzed example will be akin to the situation in regards to the 2009 Iranian elections, where there was a ruling faction and a dissenting faction in conflict. The ruling faction has some centralized authority for control of individuals and the dissenting faction also has some centralized authority for control of individuals. In addition, the individuals have some intrinsic freedom to choose the centralized control or act more independently among themselves. There are pressures from both sides (rulers or dissenters) and in both directions (centralized and distributed).

We can enhance a current city simulation with some new features to better realize the behavior portrayed by the media. A small city has already been implemented for game playing multi-agent scenarios that includes movement models and line-of-sight. Agents can move based on prescribed waypoints and connections and observe based on proximity and line-of-sight. Communication connectivity can be added to the model for simulating the short-range (e.g., talking, signaling), mid-range (e.g., megaphone, video recording) and long-range (e.g., internet, cell phone) communication channels. The ruling authority can cut some communication as they did in Iran, but the dissenting faction can adapt their behavior by using alternative forms of communication. Also, peaceful and violent behavior can be exhibited from both sides, and scaling of confrontations can be investigated. However, individuals and group behaviors and communications will be limited to both simplify and exemplify the approach.

A design and implementation strategy has been studied on the election defiance scenario in Iran. This paper describes an approach to

implementing such a simulation and describes the benefits of such a system.

### Approach

We describe here a five step approach to designing, implementing, and demonstrating a social science simulation to study the causal precursors that drive the effects in the current situation in Tehran, where protests continue sporadically against the conservative regime.

1. A baseline is necessary to allow interaction among actors. This has been accomplished using technologies that form urban environments into game models [6]. Figure 1 provides a simple viewpoint of a small city model with a variety of connected waypoints (not illustrated).
2. The players of the simulation or game need to be identified. In the case of the Iranian situation, eight player types are identified and described.
3. Each player must have enough planning ability to interact with the other players in a similar environment and illustrate realism in thought processes and ability to reassess and change strategies. This can be accomplished by integrating intrinsic-, extrinsic-, and expressive-utility in each player, and this is described from each player's point of view. These utilities are implemented via a value function that is an integral part of the ADP&E system.
4. The interactions must be identified according to the current power structure and number of agents under each authoritarian player. The interactions are identified in Figure 2 and each interactive link will be described in detail.
5. Each player is identifiable as a planner in an ADP&E system, where their plans and perceptions impact all players involved simultaneously, and where higher order affects are plausible and likely. In other words, within each planner, their parameters dictate their behavior and interaction in an attempt to maximize their own utility, while readjusting their plans to counter other planners' activities. Once implemented, parameters can be tuned to illustrate social behavior on a more complex scale.

### Step 1: Urban Environmental Game Models

In previous work, an automated technique has been developed to: generate an urban terrain movement model for computer gaming from a Compact Terrain DataBase (CTDB), increase the simulation speed of operations to allow much faster than real time operations, and a programming interface for planning algorithms has been defined to integrate multiple planners

into the model. An example city model is shown in Figure 1.

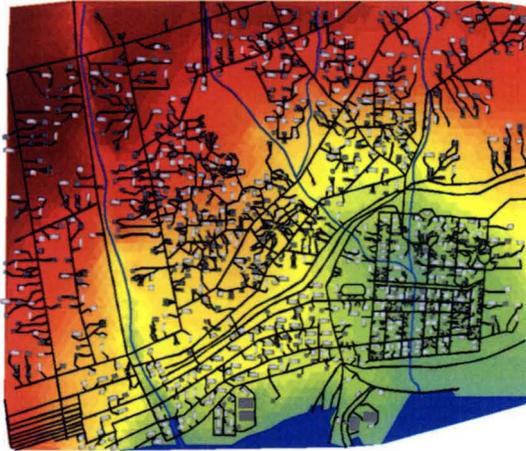


Figure 1. Example City Game Board

To better understand the order of magnitude of this city model, Figure 1 shows a top-down picture of the terrain model used. The model is a small city of approximately 4 km x 5 km. More specifically, there are 3649 buildings with over 12,000 floor locations. There were over 31,000 waypoints generated for this terrain model.

### Step 2: Major Game Players

There are five major players in the election situation in Iran, where the people are protesting against the election results, which appear to be drastically different than prior polls indicate. The five major players in this conflict are: the supreme leader Ayatollah Ali Khamenei who backs the government declared incumbent President Mahmoud Ahmadinejad, the leading challenger Mir Hossein Mousavi, the general in charge of Iran's Revolutionary Guard Mohammad Ali Jafari, the religious hierarchy, and the people.

The supreme leader is a 70-year-old cleric. He reigns over Iran's Islamic system as part pope, part commander in chief and as a one-man supreme court. President Mahmoud Ahmadinejad was the winner of the June 12, 2009 election. He is an ultra-conservative who has isolated Iran from the rest of the world through condemnations of the United States, Israel, and United Nations. The president is backed by the supreme leader and is a puppet, so he is not considered a player here. Mohammad Ali Jafari oversees the 125,000 members of Iran's military. This revolutionary guard (RG) takes direct orders and is considered the strong arm of the supreme leader. The religious hierarchy is under direction of the supreme leader as well, but some clerics are asking for reform and a recount of the election. Thus, we have broken this group into two groups, a clerical reform player and a clerical conservative player. The people are by far the

largest player in this conflict. This group can be divided into three camps: the conservatives that side with the incumbent, the reformists that side with the reform party, and the people that want to remain neutral.

As an assumption, some players are considered as single agent planners, such as the supreme leader, the reform leader, and the religious clerics. The remaining two planners are the revolutionary guard and the people. These planners require many agents in order to show the escalation of the conflict. The proper ratio is not known but there are over 7 million people living in Tehran and only 125 thousand guards in the entire country. However, the guards are well trained and armed. There are more players in the Iranian election situation than the ones described here, but these eight should be enough to sufficiently simulate the conflict.

Players\Metrics	Intrinsic Utility	Expressive Utility	Reputation Utility
Supreme Leader	Suppress Protests	Zero Tolerance/ Block Some Media	Treated As God/ Can Do Little Wrong
Reform Party	Ignite Protests/ Avoid Violence	Keep Reform Movement Alive	Adjust to People's Needs
Revolutionary Guard	Take Orders	Use Force	Never Show Fear
Religious Hierarchy Conservatives	Make People Subservient	Teach Religious Obedience	Back Religious Beliefs
Religious Hierarchy Reformists	Gain Power	Demand Recount/ Reject Violence	Empathize/ Gain People's Favor
People Conservatives	Follow Religion Verbatim	Demand Others to Follow	Hard Working/ Poorer Class
People Neutral	Follow leader and keep low profile	Avoid areas of conflict/ Be Safe	Maintain Respect/ Peace
People Reformists	Believe Reform Will Help Economy	Instigate Protests/ Free Speech	Defend Women/ Debate/ Dialogue

Table 1. Players and Their Utility Metrics

### Step 3: Utility

To appreciate the escalation of the conflict in Iran three measures of utility can be used for each player: intrinsic, expressive, and reputational utility. Intrinsic utility is the measure of what that player thinks is important and wants to accomplish. Expressive utility is the measure of how a player will deliver their message. Reputational utility is how the player perceives other players' opinion of their actions.

These players' metrics are shown in Table 1. This table is a qualitative description of the utility metrics. In an implementation, these metrics must be translated into some quantitative form that is reflected in their agents' actuators and sensors. For instance, the revolutionary guard's reputational utility is not to show fear, so they will never retreat when confronted to maintain fear in the people.

#### Step 4: Interactions

Player interactions are too many to build a real model of the Iranian election conflict. However, a simplified interactive model can be created if assumptions are made. Figure 2 shows such a

simplified representation. The interactions are labeled one to thirteen with interactions six and seven expanded for the multiple religious hierarchy players and people players, respectively.

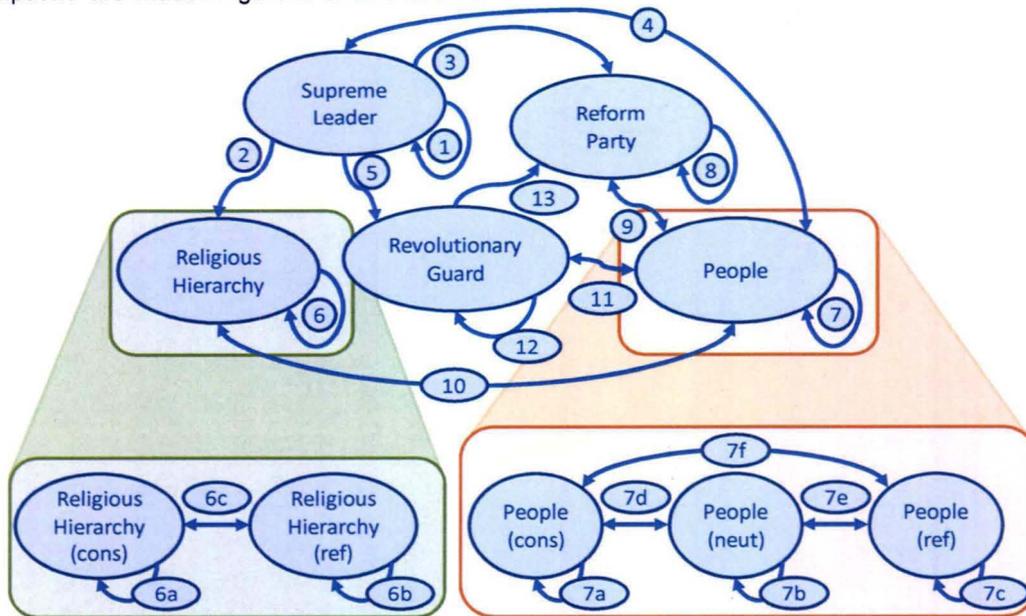


Figure 2. Players' Interactions

Connection 1 in Figure 2 is the supreme commander contemplating plans to suppress the protests, his intrinsic utility goal. Connection 2 is the supreme leader giving direction to the religious hierarchy, especially Ayatollah Ahmad Jannati Massah who heads Iran's 12-member Guardian Council, which certifies election results and is closely allied with Khamenei. Connection 3 is the limitations imposed on the reform party by the supreme leader. Many times these directions are ignored, such as not attending a religious rally to honor the dead. Connection 4 is the interaction between the people and the supreme leader. The supreme leader demands no protests and many people defy him by attending rallies. Connection 5 is the supreme leader's use of the revolutionary guard (RG) to forcibly take to the streets and break up protests. Also, the RG acts as an agent, which attempts to cut communication by confiscating cell phones and detaining people. Connections 6a-c are the religious hierarchy contemplating plans to either gain power (reformist group) or maintain allegiance to the supreme leader (conservative group). Connections 7a-f are the interactions among the people. The conflict among the people escalated into violence in first few days of protests. Connection 8 is the reform party contemplating plans as things unfold. For instance, the reform party decided to have large events centered on honoring the dead, which appealed to many people and created large crowds. Connection 9

was the interaction between the people and the reform party. They worked together to create large peaceful protests that further aggravated the supreme leader. Connection 10 is the mixed messages received from the clerics, some sided with the supreme leader while others demanded a vote recount or void election. Connection 11 exemplifies the conflict between the protesters and the RG. Many people have been killed and arrested in this conflict and is triggered by their unwillingness to back down on both sides. Connection 12 represents the RG contemplating maneuvers to break up protests, raid reformists homes, confiscate communication devices, and detain uncooperative people. Finally, connection 13 is the RG's attempt to subdue the reform party, such as detaining them from going to rallies.

#### Step 5: ADP&E System

The proven approach used here has five tiers, from the inner cycle of dynamic planning, executing, and assessing plans for players and agents, through the highest level, adapting players' strategies using tournament play through multiple games. Figure 3 illustrates this ADP&E implementation framework.

This system concept was built from the ground up to be an efficient and modular approach. This approach has been already applied for two applications, the game RISK [7], and an urban search and rescue operation [8].

- First, the core cycle was developed as an action and response system, where individual action sequences are planned, executed, and assessed in various model environments, with varying projected expectations, over many cycles, and for all agents in the correct time sequence.
- At the second level, agents execute a particular plan, and each agent's action set is stored separately for modularity.
- Third, the player is the conceiver and conductor of a plan that encompasses all agent activities. A player has a set of parameters that determine its choice of planned actions, and how often to re-plan those actions.
- Fourth, a game is the domain where action sequences are executed in the model environments, which will always lead to a final goal state. The final goal state must be achievable, because human intervention is prohibited in this framework and a game only completes when the final goal is achieved.
- Fifth, tournaments of games are arranged, so that players can improve their parameter settings over the course of many tournaments. Through evaluating each player's progress, and modifying the best players' parameters, players can improve their play.



Figure 3. ADP&E Framework

At the heart of this approach is a core planning cycle for each of the eight players of the game. Figure 3 shows an illustration of this cycle. The core cycle has three components: (1) plan-generator (PG); (2) plan-executor (PE); and (3) plan-assessor (PA). The *plan-generator* is considered the search engine for contemplating plans for each player. PG strings together individual actions to form plans for each agent

based on current perception of situation. The utility metrics described above can be used to evaluate plans and choose the better ones. Formulations as to how to generate and choose plans have been examined on two very large planning problems and are described in two previous papers [7] [8]. The *Plan-Executor* executes the plans in time sequential order. The *plan-assessor* estimates how well the remaining plan will execute given new observed information acquired from the environment while executing the plan. This cycle can be run after each executed action.

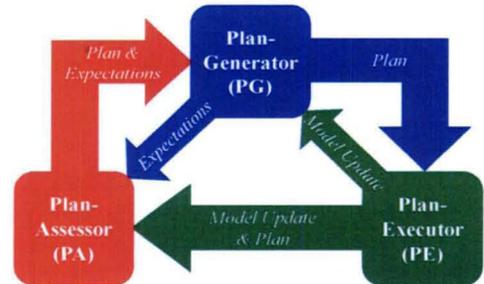


Figure 4. Planning Core Cycle

The three components use three objects that are manipulated and shared among the components. These three objects are the (1) *plans*, (2) *models*, and (3) *expectations*. *Plans* are generated by PG, executed by PE and assessed by PA. All players can be run in separate threads and execute independently. *City Models* are used in PG to predict future states, are used in PE to observe the real states, and are used in PA to observe whether expectations will be met. The models used in PG and PA are virtual-state city models, which are approximate to the real-state model used in PE. The real-state model is a real-world model, where a plan is executed. Virtual-state models do not know the real states until observed and are initialized to reasonable expectations. Thus, there are nine perceptions of the city model based on which planner is under consideration. There is one virtual model for each planner and a real-world model where all planners can execute their actions. *Expectations* are the measure of how well a plan achieves a desired goal (utility metrics), such as breaking up a protest. Expectations are projected both by the generated plan in PG and by the plan used in PA. The two expectations are compared to see if the expectations projected in PA still meet or exceed the originally generated plan expectations projected in PG. Each agent has an expectation for its plan. If expectations are met to a prescribed degree, a plan is retained; otherwise a plan is reformulated in PG.

If implemented, such a simulation tool can provide three major advantages. First, tuning parameters is crucial to matching historical

records. The versatility in choosing alternative actions under uncertainty (e.g., reformist people were younger and more educated, using high tech devices for communications, something the leaders did not consider in initial plans), the timing of actions/ responses (e.g., the government lost credibility when saying the election was true when they did not use any time to investigate), the amount of reassessment and replanning (e.g., people switched to alternative forms of communication when services were cut, such as twitter, and cell phones) of each of the eight players is critical. These are just three instances where agile planning is used in real world social events, and there are many other areas to investigate. Thus, tuning planner parameters in key aspects is essential to matching real world scenarios. The tuning of parameters can be learned via developed techniques already established for two other applications [7] [8].

The second advantage is the use of an ADP&E system to predict how real-time events will unfold. When a model has been developed that accurately predicts the evolution of historical events for a culture as described above, it can be tuned to follow the course of current events and could predict their future development with less uncertainty. These predictions can be further fine tuned to account for shifting alliances and priorities. Once a baseline of activity has been established, the ability to identify underlying causes such as those that lead to unexpected results is valuable information in itself.

The third advantage of such a simulation tool is to inject possible outside influences into the model and see if and how they alter the course of events. Models such as these could self train to produce the most desirable effects with the smallest perturbations. Further, trained models may be examined to determine that observations of the evolving environment are most useful to determine that plan expectations are being met.

### Summary

This paper has proposed the application of ADP&E to modeling social influence in a combined centralized and distributed context. Individual agents have partial allegiances to one or more, potentially conflicting, central authorities, as well as their own internal goals and principles. Agents are not simply reactive, but proactively plan and execute action sequences in these contexts. ADP&E can provide a means of modeling the social forces at work within an individual agent, as well as the shifting allegiances and conflicts among agents. Into this complex, dynamic hierarchy, various PSYOP interventions can be injected, and the

micro and macro reactions of the system observed.

The unrest surrounding the Iranian elections in the summer of 2009 have been used as an illustrative example of ADP&E modeling. The defining elements of that situation have been deconstructed into items and relationships prerequisite for the formation of a model. Application of ADP&E to that model has served to explain the features of ADP&E, and describe its benefits for such social influence models.

### References:

1. Eric Bonabeau, Marco Dorigo and Guy Theraulaz. *Swarm Intelligence: From Natural to Artificial Systems*, 1999
2. Ghallab, Malik; Nau, Dana S.; Traverso, Paolo. *Automated Planning: Theory and Practice*, Morgan Kaufmann, 2004
3. G. Tesauro, "Programming backgammon using self-teaching neural nets," *Artificial Intelligence*, V. 134, 2002, pp. 181-199.
4. R. Levinson, F. H. Hsu, J. Schaeffe, T. A. Marsland, & D. E. Wilkins, "The role of chess in artificial-intelligence research," *ICCA Journal*, V. 14, N. 3, 1991, pp. 153-161.
5. J. Vaccaro, C. Guest, "Planning an Endgame Move Set for the Game RISK: A Comparison of Search Algorithms," *IEEE Transactions on Evolutionary Computation*, Vol. 9, No. 6, December 2005, pp. 641-652.
6. J. Vaccaro, C. Guest, "Modeling Urban Terrain for Simulating Search and Rescue Operations to Train Artificial Planners," *The 17<sup>th</sup> IASTED International Conference on Applied Simulation and Modelling (ASM'08)*, Corfu, Greece, June, 2008.
7. J. Vaccaro, C. Guest, "Learning Multiple Search, Utility and Goal Parameters for the Game RISK," *IEEE World Congress on Computational Intelligence (WCCI'06)*, Vancouver, Canada, July 2006.
8. J. Vaccaro, C. Guest, "Automated Dynamic Planning and Execution for a Partially Observable Game Model: Tsunami City Search and Rescue," *IEEE World Congress on Computational Intelligence (WCCI'06)*, Hong Kong, China, June 2008.

# Individualized Cognitive Modeling for Closed-Loop Task Mitigation

Guangfan Zhang; Roger Xu; and Wei Wang  
*Intelligent Automation, Inc.*  
*gzhang@i-a-i.com*

Jiang Li  
*Old Dominion University*

Tom Schnell and Mike Keller  
*University of Iowa*

**Abstract:** An accurate real-time operator functional state assessment makes it possible to perform task management, minimize risks, and improve mission performance. In this paper, we discuss the development of an individualized operator functional state assessment model that identifies states likely leading to operational errors. To address large individual variations, we use two different approaches to build a model for each individual using its data as well as data from subjects with similar responses. If a subject's response is similar to that of the individual of interest in a specific functional state, all the training data from this subject will be used to build the individual model. The individualization methods have been successfully verified and validated with a driving test data set provided by University of Iowa. With the individualized models, the mean squared error can be significantly decreased (by around 20%).

## 1. INTRODUCTION

In recent years, researchers have been actively performing machinery/electronics diagnostics and prognostics for automated aviation systems. To ensure mission success, the functional states of human operators also need to be monitored since mismatched Operator Functional State (OFS) and workload (either over-load or under-load) conditions can lead to disastrous consequences [1].

According to [2], OFS can be defined as the multidimensional pattern of human psycho-physiological condition that mediates performance in relation to physiological and psychological costs. Different contributing factors, including environmental factors (Altitude, noise, etc.), individual state (circadian rhythms, sleep loss, illness, etc.), and task characteristics (physical load and cognitive load), can affect the OFS and lead to suboptimal performance in human operators. It is challenging to consider all those factors to predict the OFS accurately in real time. Furthermore, current available OFS modeling tools have limited applicability as they do not account for the considerable individual differences due to individual physical fitness and adaptability to external/internal conditions.

In this paper, we introduce a closed-loop Adaptive Task Management System (ATMS) to identify hazardous states that are likely to lead to operational errors and dynamically aid operators to minimize human errors. Key innovations in the framework include 1) a systematic approach to perform OFS assessment considering all the contributing factors, 2) a committee machine-based regression model with advanced feature selection method to accurately build the mapping between input parameters and output functional state, 3) a two-step model individualization technique for individual OFS monitoring, and 4) efficient task management to address both over-load and under-load situations. An accurate OFS assessment is the foundation of the ATMS. Therefore, in this paper, we focus on how to build the real-time individual OFS assessment model.

This paper is organized as follows. In Section 2 we describe the closed-loop ATMS framework. In Section 3, we introduce the enhanced committee machine-based OFS assessment method. The focus on section 4 is to build the individualized OFS assessment model. In Section 5, the individualized OFS assessment model performance is verified and validated with a driving test dataset. Section 6 concludes this paper.

## 2. Closed-Loop ATMS Framework

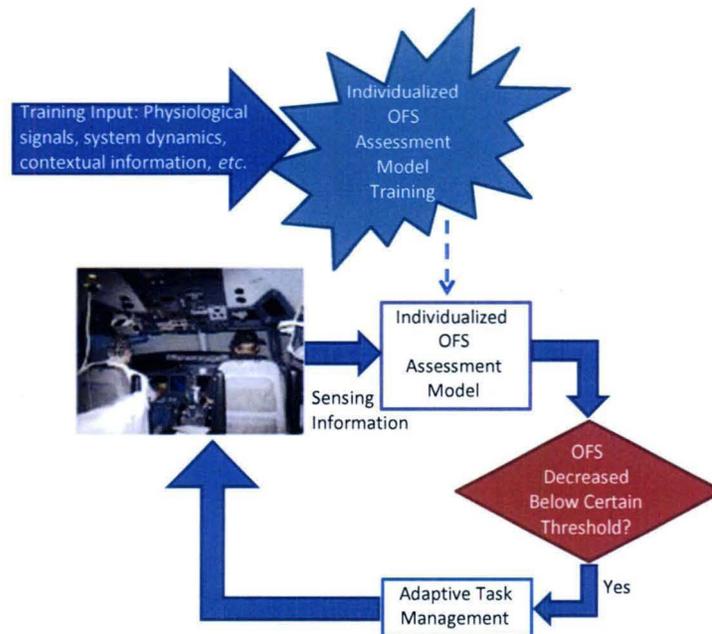


Figure 1: ATMS framework

The framework of the closed-loop Adaptive Task Management System (ATMS) is shown in Figure 1. In the ATMS framework, to accurately monitor the Operator Functional State (OFS) for each individual, the OFS assessment model is trained and individualized using different sources of training inputs (physiological signals, system dynamics measurements, etc.). An enhanced committee machine-based OFS assessment model is employed to map input parameters to individual OFS, in which the responses of multiple neural networks (committee members) are combined into a single response to improve efficiency and accuracy. To further boost the OFS assessment performance, we utilize an advanced feature selection algorithm [3] to select different features for each committee member.

Due to large individual variations, a generalized OFS assessment model trained using data from large number of subjects usually does not yield satisfactory performance when applied to an individual operator. We individualize the generalized OFS assessment model using data from the individual of interest, as well as selected subjects whose data has been used in training the generalized model. The selection is based on a similarity measure: if a subject's training data is similar to the individual's data in specific functional states (for example, they are close to each other in the sense of Euclidian distance in the feature space computed from the data), all the training

data from this subject will be used to individualize the generalized model.

If the OFS of an individual decreases below a certain threshold, a task performance augmentation strategy can be applied to even-out workload and maintain the operator in an optimum cognitive workload level. As a result, the operator can be continuously engaged and able to respond quickly and appropriately to unusual situations.

### 3. Enhanced Committee Machine-based OFS Assessment

The basic procedure for real-time OFS assessment is shown in Figure 2. It includes pre-processing, feature extraction/selection, and regression. We have developed an enhanced committee machine-based regression method for the OFS assessment.

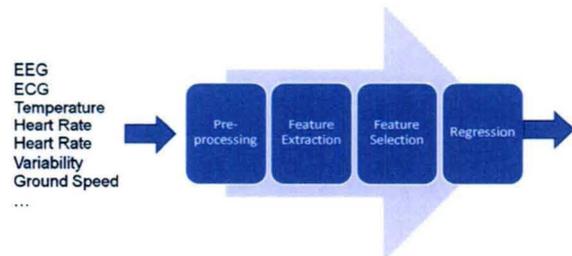
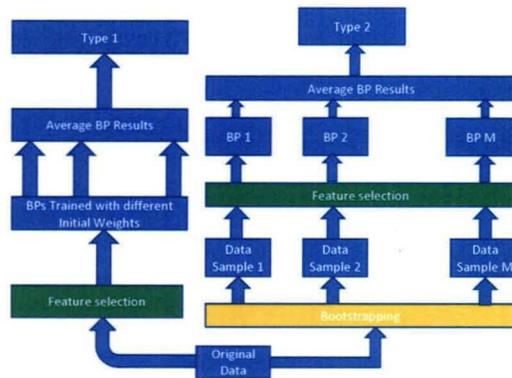


Figure 2: Real-time OFS assessment procedure

### 3.1 Committee Machine

A committee machine is a strategy to improve classification or regression performance by combining responses from multiple diversified committee members (trained perceptrons in neural networks, for example). The performance of the committee machine is often better than that of each committee member [4] based on two main reasons. First, if committee members have the diversity property, i.e. they are unlikely to make errors in the same feature space, the errors from individual committee members will be canceled by each other to some extent. Second, since the committee machine “averages” its individual member’s estimation, the variance of the committee machine can be significantly reduced.

Two types of committee machines, as shown in Figure 3, are implemented using a multilayer perceptron trained by the standard Back Propagation (BP) algorithm as the base regression model for OFS assessment. The base regression model is combined with an adaptive learning factor to make training algorithms converge much faster than the traditional BP [5]. Also, both committee machines are combined with an advanced feature selection algorithm, Piecewise Linear Orthogonal Floating Search (PLOFS) [3].



**Figure 3:** Two types of committee machines

The first type of committee machine was built by training each committee member using different initial weights. It is trained based on an assumption that each member will converge to a different location in the solution space. Given the fact that the error surface of a neural network has multiple local minima, the local minimum that a training algorithm converges to will differ depending on its initial condition. Therefore, we combine the predictions from different committee members as the final solution. Each member with the diversity property covers partially the solution

space, and the final committee can then cover larger solution space.

Type 2 committee machine contains a set of multilayer perceptrons trained by different bootstrapped datasets after feature selection. To make each of the committee member diversified, each member further uses a different set of features, which is different from the traditional “bagging” training technique that simply aggregates bootstrapped individuals and selects the same features for all the committee members.

### 3.2 Feature Extraction and Selection

Many features can be extracted for OFS assessment. For EEG, total spectral power can be calculated in the alpha, theta, beta, and gamma bands as these bands reflect cognitive states. In addition, we can examine signal coherence between inter-hemispheric electrodes such as F1 and F2 (from the 10-20 electrode placement system). Measures of signal coherence focus on the high-alpha bands (9-12 Hz), as increased coherence among these signals is thought to distinguish higher levels of cognitive activity. With the eye tracking data, we can examine blink frequency, percent eye closure (PERCLOS), average eye closure speed (AECS), mean/variation change of pupil size over time and the percentiles of pupil size. We can also extract features related to eye movements. The increased frequency of saccades may indicate an increase in multi-tasking demands, requiring operators to split attention. It may also indicate increased demands on spatial working memory, as operators may need to maintain visual data to integrate it across multiple displays.

With the large amounts of features, we need to carefully evaluate the features and select a subset of features that can best estimate the OFS. A feature selection algorithm usually evaluates the fitness of features first, and then searches for different combinations of features with the goal of maximizing the fitness value [6-11]. Two common types of features selection algorithms are filter approaches and wrapper approaches. A filter type method ranks features according to some predefined criteria such as mutual information, class separability measure without any actual model assumed between outputs and inputs of the data, a feature is then selected or discarded based upon the ranking. A wrapper approach utilizes a model to evaluate the fitness values of features and features are selected using the fitness as a guide. Usually, wrapper approaches give better results than filter approaches but have higher computational complexities [7]. In the OFS assessment model, we utilize a wrapper type algorithm, Piecewise Linear Orthogonal Floating Search (PLOFS), to select features for the



individual's OFS, and then adjust the weights of each of the selected committee members using a linear regression approach to form a final estimation of the OFS. Figure 6 compares the individualization process (right) to a generalized OFS assessment model (left).

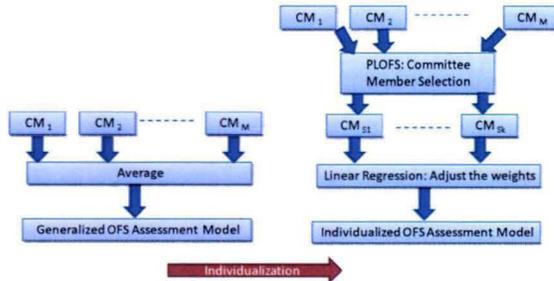


Figure 6: Model individualization

### 5. Experimental Study

We utilized the driving test dataset collected using the Cognitive Avionics Tool Set (CATS) software developed by Operator Performance Laboratory (OPL) [12] in the University of Iowa to verify and validate the individualized OFS assessment approach. The enhanced committee machine and individualization strategies were implemented on the driving test dataset.

CATS [12] is a powerful, database driven data visualization and analysis package. The toolset synchronizes a large number of incoming data streams operating at different update rates into a single, unified file. CATS provides a rich set of visualization tools to inspect physiological data. Figure 7 shows a snapshot of the CATS software.

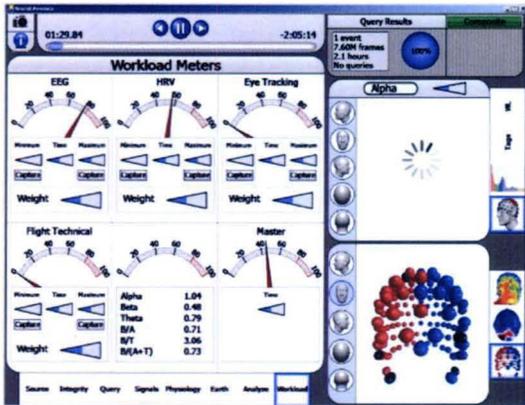


Figure 7 CATS GUI

In the driving test dataset, subjects performed a driving test in two hours. During the test, different types of information were collected, including description of the task, system dynamics related

information (such as ground speed), performance measures, physiological signals (128-channel EEG, ECG, respiration, etc.), and eye tracking data. And the workload was analyzed every second based on the driving scenario (city-driving, stopped, highway passing, etc.).

Five subjects' data were used to verify and validate the individualized OFS assessment methods. We first trained a generalized model for each subject using a leave-one-out method, which trains the model using the other four subjects' data and leaves the data from the subject being modeled for testing (Method 1 in Figure 8). We then implemented the previously described model individualization methods for each subject (Method 2: Individual Model 1; and Method 3: Individual Model 2). For a comparison purpose, we also trained an individual model for each subject by randomly selecting half of the data from that subject, and tested the model with the rest half (Method 4). The performance, Mean Squared Error (MSE), is compared in Figure 8 and Table 1.

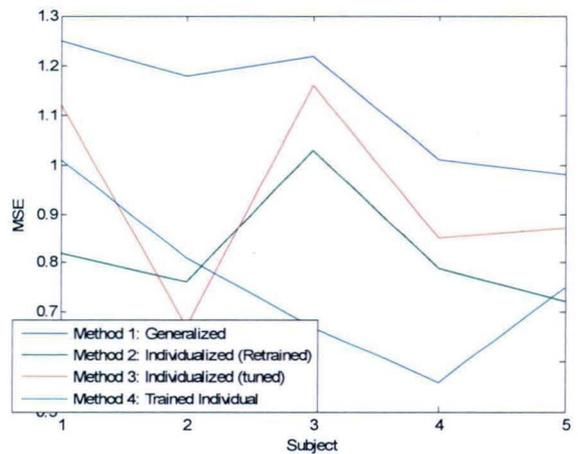


Figure 8: Performance comparison

It is clear that the two model individualization methods (Method 2 and Method 3) significantly reduce the MSE given by generalized models (Method 1). Compared with a generalized OFS assessment model, the mean squared prediction error is about 20% lower. The performance using the two model individualization techniques are comparable to that of the individual model trained with sufficient data from the individual, but only require very limited data for training or individualization (5-minute in our experiment).

**Table 1: Performance comparison**

Method	Subject 1 MSE ± STD	Subject 2 MSE ± STD	Subject 3 MSE ± STD	Subject 4 MSE ± STD	Subject 5 MSE ± STD
1	1.25 ± 0.029	1.18 ± 0.024	1.22 ± 0.023	1.01 ± 0.021	0.98 ± 0.02
2	0.82 ± 0.018	0.76 ± 0.018	1.03 ± 0.021	0.79 ± 0.017	0.72 ± 0.018
3	1.12 ± 0.02	0.67 ± 0.02	1.16 ± 0.026	0.85 ± 0.016	0.87 ± 0.022
4	1.01 ± 0.009	0.81 ± 0.016	0.67 ± 0.017	0.56 ± 0.017	0.75 ± 0.015

## 6. Conclusions

The research effort results in a successful development of an individual OFS assessment model for closed-loop task management. The model incorporates novel committee machine-based OFS assessment with an advanced feature selection method, Piecewise Linear Orthogonal Floating Search (PLOFS), and two different individualization techniques have been developed to improve the OFS assessment performance for each individual. The experimental results show significant improvements of the individualization techniques.

## Acknowledgements

This project was funded by the NASA (Contract No: NNX09CE90P). We thank Dr. Alan T. Pope, our COTR, for his comments and suggestions as we performed this research. We also thank Dr. Kara Latorella for her comments and feedback during various discussions over the course of this project.

## References

1. Alan T. Pope and Edward H. Bogart, "Identification of hazardous awareness states in monitoring environments," SAE Transactions: J. Aerospace, vol. 101, 1992, pp. 449-457.
2. NATO Advanced Research Workshop on Operator Functional State and Impaired Performance in Complex Work Environments (2002 : Il Ciocco), "Operator Functional State: The Assessment and Prediction of Human Performance Degradation in Complex Tasks," (NATO Asi) (Vol 355)
3. Jiang Li et al., "Feature selection using a piecewise linear network", IEEE Transaction on Neural Network, vol. 17, no. 5, pp. 1101-1115, 2006.
4. L. Breiman, Combining artificial neural nets, vol. 31, ch. Combining predictors. Springer, 1999.
5. Jiang Li, Manry, M. T., Liu, L. M., Yu, C., and Wei, J., "Iterative improvement of neural classifiers," in Proceedings of the Seventeenth International Conference of the Florida AI Research Society, 700 - 705 (May 2004).
6. L. Breiman, J. Friedman, R. Olshen, and C. Stone, CART: Classification and Regression Trees. CBelmont, California: Wadsworth, 1983.
7. R. Kohavi and G. John, "Wrappers for feature subset selection," Artificial Intelligence, vol. 97, no. 1-2, pp. 273-324, 1997.
8. R. Kramer, Chemometric Techniques for Quantitative Analysis, (1998) Marcel-Dekker, ISBN 0-8247-0198-4.
9. P. M. Narendra and K. Fukunaga, "A branch and bound algorithm for feature subset selection," IEEE Trans. Comput., vol. 26, pp. 917-922, 1977.
10. P. Pudil, J. Novovicova, and J. Kittler, "Floating search methods in feature selection," Pattern Recognition Letters, vol. 15, pp. 1119-1125, 1994.
11. M. Raymer, W. Punch, E. Goodman, L. Kuhn., and A. Jain, "Dimensionality reduction using genetic algorithms," IEEE Transactions on Evolutionary Computation, vol. 4, pp. 164-171, 2000.
12. Schnell T., Keller, M., Macuda, T., "Application of the Cognitive Avionics Tool Set (CATS) in Airborne Operator State Classification," Best Topic Paper, Augmented Cognition International Conference, Oct. 1-3, 2007, Baltimore Waterfront Marriott Hotel in Baltimore, Maryland.

# Cognitive Modeling Of Video-Game Player User-Experience

Corey J. Bohil; Frank A. Biocca

Michigan State University

bohil@msu.edu

**Abstract.** This paper argues for the use of cognitive modeling to gain a detailed and dynamic look into user experience during game play. Applying cognitive models to game play data can help researchers understand a player's attentional focus, memory status, learning state, and decision strategies (among other things) as these cognitive processes occurred throughout game play. This is a stark contrast to the common approach of trying to assess the long-term impact of games on cognitive functioning after game play has ended. We describe what cognitive models are, what they can be used for and how game researchers could benefit by adopting these methods. We also provide details of a single model – based on decision field theory – that has been successfully applied to data sets from memory, perception, and decision making experiments, and has recently found application in real world scenarios. We examine possibilities for applying this model to game-play data.

## 1. INTRODUCTION

A major goal of video game research is to understand and influence what a player is thinking during game play, and perhaps to effect long term changes in the game player. One major theme in this research includes game impact on personality traits and emotional states – most famously, effects on player aggressiveness. This type of research is typically carried out via survey methods. Players answer a battery of questions before the game begins to assess their personal traits and current emotional state. After game play has completed, the player answers more questions and any changes are attributed to the effects of the intervening game play period. Another prominent theme is the effect of games on some aspect of cognition (e.g., spatial skills [5]). Although this second research theme is concerned with cognitive effects of games, it is often carried out by pretesting the participant using a standard laboratory task (e.g., a speeded search task) to assess the trait of interest (e.g., visual acuity), followed by a game play period, and then posttesting using the same standard laboratory task again to see if performance has changed.

These approaches take a bird's eye view of the cognitive phenomena underlying game play. They treat the mind like an impenetrable black box, observing or manipulating inputs to the cognitive system, and observing the concomitant outcomes. Although these research endeavors are valuable, they take an indirect route to understanding cognition during play. In both approaches outlined above, game play effects are measured after the fact. In the case of questionnaire methods, the data is subjective – participants give some indication of

the game's effects through their answers to various questions. And although the second research approach aims at understanding something about cognition as it pertains to games, it too focuses on effects and measures indirectly by assessing changes after the fact.

What is needed is an approach that allows one to track cognitive effects of games *during* the game play session. As the player progresses through the game, seeking to accomplish various goals, making decisions, all manner of cognitive phenomena come to bear. Learning is required (e.g., what strategies worked before?). Attention has to be allocated. Memories of previous outcomes have to be accessed. Decisions have to be made. The outcome of these cognitive events translate into the player's observable performance in the game, their level of enjoyment or accomplishment, their learning. Understanding these dynamic events as they unfold throughout the course of play, rather than trying to infer something about them subjectively or indirectly after the game is over, would be of great value to the designer of games concerned with changing behavior, communicating messages, or just maximizing engagement and fun. In attempting to maximize game efficacy, the designer would likely benefit from knowing what a player is looking at when making decisions, what dimensions are most salient, which dimensions are routinely ignored, and when options are confusable.

In recent years, there have been efforts to peer into the inner workings of the mind during the game play events that cause them. Brain imaging techniques (e.g., fMRI) have been used to associate brain activity known to occur during aggressive thought with violent game content [8]. Other studies have

tracked psychophysiological events (e.g., EEG) to infer mental states during play [1]. This approach to studying player cognition during game play is a welcome addition to the field. This research gives us valuable data synched in time to game events, and we can learn a lot by trying to interpret it. However, at this point in time, both imaging and psychophysiological data are difficult and expensive to obtain, the measures are still relatively crude, and findings are often difficult to interpret. The degree to which one can relate the observed bodily states to aspects of cognition such as attention or executive function is an issue of lively (sometimes withering) debate [7].

Luckily, there is another approach to understanding cognitive phenomena at our disposal – cognitive modeling. For the past several decades, cognitive psychologists have devised and tested scores of detailed mathematical models that offer precise accounts of the cognitive underpinnings of behavior, and demonstrated their links to theoretical structures like memory and attention. Given the mature state of this field, as well as its widespread representation throughout academia, it is surprising to find that it has very little representation in the game studies literature (although some applications are noted below). The objective of this paper is to provide a basic understanding of what cognitive models can provide researchers, and to advocate their use in studying video games.

## **2. COGNITIVE MODELING**

What, exactly, is a “cognitive model”? A cognitive model is a mathematical interpretation (i.e., specification) of the set of principles embodied in a theory of cognition. Cognitive models make specific assumptions about the information represented in the cognitive system (e.g., words and their meanings), along with the processes acting on this information to produce observable cognitive behavior (e.g., classifying an object). More concretely, a model receives inputs like a person in an experiment (e.g., size of objects on a screen), performs mental operations (e.g., like comparing perceived stimulus information to information stored in memory), and outputs a response (e.g., emits a classification of the object).

Models such as these are valuable for several reasons. First, they require a researcher to move past the initial stages of theorizing – often characterized by vague verbal descriptions of mental entities and their interactions – to taking a detailed, specific stance on these quantities and relationships. Doing this affords the research

community a better opportunity to evaluate and criticize a theory’s quality. Second, making detailed quantitative statements in a cognitive model allows a researcher to make precise, testable predictions. A third benefit is that simulating model behavior on a computer can lead to unexpected observations and insights that the researcher might not otherwise have reached. It is widely agreed in the modeling community that this is an important benefit of modeling.

There currently exists a wide array of cognitive models that have been vetted over the years by many experiments and data sets. These models elucidate a range of topics. Many models are designed to capture steady-state performance in cognitive tasks like recognition memory, discrimination ability, attention allocation, to name a few. These models are intended to account for specific, circumscribed aspects of cognition such as recognition, categorization, attention, etc. Another class of models – known as connectionist models (also called neural net or parallel distributed processing models) – mimic fundamental aspects of brain anatomy (i.e., populations of single processing units or artificial neurons communicating activation levels back and forth) and capture learning over the course of many training trials. A third class of models – known as cognitive architectures (e.g. ACT-R, EPIC, Soar) – attempt to capture several aspects of cognition in a single unified framework (e.g., attentional processes, memory, visual search tendencies), reflecting the fact that all these processes come into play simultaneously in the human cognitive system. Cognitive architectures have found wide application in human-computer interaction research and have even made their way into game research to some extent [4]. Existing applications of cognitive modeling in game research tends to take on a computer-science flavor. These models are valuable tools for making the game respond to the player in interesting ways or to create “smarter” non-player characters [3]. Our aim in this paper is to encourage much more widespread adoption of these techniques for gaining general understanding of the cognitive capacities invoked during video game play.

## **3. APPLYING COGNITIVE MODELS TO GAME STUDIES**

The value of models for game research lies in the fact that models require inputs and produce outputs. In between they offer precise statements about attention, learning, decision strategies and biases, and so on. In doing so, a model often tells

the researcher *why* performance looks as it does. Although a model can't tell the designer exactly how to craft a game environment that teaches or entertains, discovering that current inputs place unrealistic demands on attention might offer guidance by narrowing the range of necessary modifications to gain desired results. An important detail, of course, is how one goes about applying these models.

Within a cognitive model lie parameters that capture the modeled quantities (e.g., attention weights, learning rate, response biases). These values are indicators of the mental underpinnings of observable behavior. In order to make inferences about cognition, these models are often "fit" to a set of data. The computer takes the output of the model (i.e., predicted responses to events), compares it to player data (i.e., actual responses) and adjusts the internal parameter values (i.e., changes assumptions about attention, etc.) until the predicted responses are as close as possible to the data. The resulting adjusted parameter values indicate things like how confusable the stimuli were or which stimulus dimensions garnered the most attention. These parameter values can be used to make predictions for the player in later game sessions or scenarios.

It is also important to verify that what the model tells us is correct. In order to do this, researchers often attempt to fit a model to data using fixed parameter values gleaned from prior knowledge of the research participant. Achieving a good model fit (i.e., a good prediction of player performance) by setting parameter values *a priori* is a powerful demonstration that one understands the player's cognitive processes during play.

One way to obtain fixed parameter values for a *priori* prediction is to fit the model (by adjusting free parameters) to one data set, and then use the best-fitting parameter values to see if the model accounts for additional data sets (without re-adjusting the parameter values). Another way to demonstrate our understanding is to set the model parameters based on something else we already know about the player.

For example, one could take advantage of the kinds of data acquired through the survey methods described above. One recent (non-game) study used results from a survey designed to assess whether a person has an "action" orientation (tendency to accept risks to expedite achieving a goal) or a "state" orientation (tendency to be more deliberative in order to avoid risks). Scores on this questionnaire were converted into parameter values

in a cognitive model and used to predict response probabilities and response time distributions in a sports-related task [6]. Such an approach grounds model parameters in knowledge about the participant even before experimental manipulation begins, and can still enable the model to make interesting predictions about behavior.

Another possibility would be for the researcher to set model parameters to reflect instructions given to the player (either before the game or inside the game). Instructing a player to pay attention only to RED enemies, for example, should be reflected in a model's attention weight parameters (assuming the model has them) and consequently in the model's predicted response probabilities (and hopefully lead to a good model fit). An important long-term goal of modeling is to find parameter values that can lead to valid predictions across several experimental conditions without the need to adjust parameters to account for each data set.

One challenge to applying cognitive models to data from video games is that events of interest must be operationally defined. For example, some agreement might need to be reached about what constitutes "fighting or fleeing" in a game scenario. Another example would be determining what qualifies as a response option. Depending on the question under study, it may be wise to compare performance only in situations with a constant number of response options. Such apples-to-apples comparisons might be necessary when trying to determine response probabilities or response time distributions.

#### 4. DECISION FIELD THEORY

Among the many aspects of cognition that can be modeled and examined in games, perhaps the most natural starting point is to look at decision making. One popular class of models that illuminates decision making is known as "sequential sampling" models. Sequential sampling models simulate the accumulation of information (i.e. sampling) over time in support of each choice alternative, leading to the eventual selection of one option over others. Decisions are triggered by internal choice thresholds – the first accumulation process to reach threshold wins, and the corresponding choice is made. Figure 1 depicts this sampling process for three choice options.



**Figure 1:** Information accumulation for three choice options

In this section, we describe a sequential sampling model based on Decision Field Theory [2]. Variants of this model have been successfully applied to a wide range of phenomena, including decision making, perception, and memory, among others. The model has mostly been applied to data from standard laboratory tasks, but has recently been used to explain decision making in a sports judgment task [6].

On a given experimental trial (in the context of games, an operationally defined recurring event), the model assumes that each set of choice options can be characterized by values along salient dimensions. For example, when trying to choose the best weapon for a fight, the player might consider three weapons along dimensions such as strength, range, and ammunition supply. Each weapon has its own set of values on these dimensions, and the player makes some assessment of these values. Table 1 illustrates some hypothetical values.

**Table 1:** Hypothetical dimension values for weapon choices

	Strength	Range	Ammo
Option 1	1.0	50	0.6
Option 2	0.5	150	0.8
Option 3	0.7	100	0.8

Also, each player is likely to display some difference in preference for the choice dimensions. For example, the player's decision might be most strongly influenced by the strength dimension 70%

of the time. Range might be the most influential dimension 20% of the time, and ammo only 10% of the time. These values are model parameters. The model uses these values on each trial, along with other parameters representing initial biases, memory from trial to trial, and similarity between options, to produce a decision.

Figure 1 displays a characteristic example. The figure shows the (simulated) stochastic accumulation over time of evidence (to the cognitive system) in favor each of the three options. The first option to reach an internal decision threshold "wins" the race, and supplies the response. As the figure shows, not only is a choice determined from this process, but also the time to reach threshold. From trial to trial, responses and termination times will vary, and over trials the model will provide response probabilities and response time distributions that can be compared to a player's data. By adjusting the internal model parameters in order to fit the observed responses, the model tells a tale about the player's attention focus, memory, biases, and the confusability of the response alternatives.

This decision-field theoretic model could potentially answer many interesting questions. For example, how does action orientation predict game play? How does decision strategy change as a result of learning throughout the game? Which dimensions receive the most attention, and which the least?

## 5. DISCUSSION

We've argued in this paper that cognitive modeling provides a detailed and dynamic view into cognition – at the individual player level – as it unfolds during video game play. Currently, this powerful approach is seldom utilized in game research. This is a shame, since cognitive modeling is a mature field, and there are many useful models available that have been affirmed by decades of research in carefully controlled experiments.

Models can offer clues into the inputs required to produce the outputs desired. If a game is to have educational value, (or for communication or even just for fun), then variables that influence model behavior should be manipulated to moderate player behavior. Currently, most game design is guided by heuristics, prior experience, and flashes of insight.

Many of the cognitive models in existence today are ready for extension to new areas. In fact, the field of cognitive psychology is increasingly marked by attempts to extend the reach of cognitive theories to real-world scenarios. The application of these tools

is especially timely considering the recent explosion of research into serious games (games designed to communicate and educate players). Designers of such games would likely benefit from a tool that can help foster a deeper understanding of what players focus on and are affected by during game play.

Finally, cognitive modeling dovetails well with the imaging and psychophysiological research mentioned above. The relatively recent emergence of the field of cognitive neuroscience attests to this. Cognitive models have become so powerful that competition between theories is often difficult to assess on the basis of behavioral data alone. Neuroscience data is now routinely used to place biological plausibility constraints on computational models. In turn, cognitive modeling imparts a deep level of meaning to neuroscience results. Models help neuroscientists understand the cognitive implications of their data.

In conclusion, cognitive modeling presents a powerful method for understanding what a player is thinking about while playing a video game. Research papers that describe cognitive models often report their model derivations in detail so that interested readers can adopt these methods. Our hope is that we've been able to convince readers of the allure of cognitive modeling for their own game research.

## 6. ACKNOWLEDGMENTS

We thank Bradly Alicea, Nick Bowman, Andy Boyan, Allison Eden, & Mayur Mudigonda for their comments on an earlier draft of this article. This work was funded by an AT&T endowment to the second author.

## REFERENCES

1. Baumgartner, T., Valko, L., Esslen, M., & Jancke, L. "Neural correlate of spatial presence in an arousing and noninteractive virtual reality: An EEG and psychophysiology study." *Cyberpsychology & Behavior* 9,1 (2006), 30-45.
2. Busemeyer, J. R., & Townsend, J. T. "Decision field theory: a dynamic-cognitive approach to decision making in an uncertain environment." *Psychological Review*, 100 (1993), 432-459
3. Funge, J. D. *AI for Computer Games and Animation: A cognitive Modeling Approach*. AK Peters Ltd., Wellesley, MA, USA, 1999
4. Gray, W. D. (Ed.) *Integrated Models of Cognitive Systems*. Oxford University Press, New York, 2007
5. Kaveri, S., & Greenfield, P. M. "Effect of video game practice on spatial skills in girls and boys." *Journal of Applied Developmental Psychology* 15,1 (1994), 13-32.
6. Raab, M., & Johnson, J. G. "Individual differences of action-orientation for risk-taking in sports." *Research Quarterly for Exercise and Sport*, 75,3 (2004), 326-336.
7. Uttal, W. R. *The New Phrenology: The Limits of Localizing Cognitive Processes in the Brain*. The MIT Press, MA, USA, 2003
8. Weber, R., Ritterfeld, U., & Mathiak, K. "Does playing violent video games induce aggression? Empirical evidence of a functional magnetic resonance imaging study." *Media Psychology* 8,1 (2006), 39-60.



# Developing Collective Training for Small Unmanned Aerial Systems Employment

Paula J. Durlach<sup>1</sup>; Heather Priest<sup>1</sup>; Glenn A. Martin<sup>2</sup>; and Jay Saffold<sup>3</sup>

1. U.S. Army Research Institute for the Behavioral and Social Sciences

*Paula.Durlach@us.army.mil; Heather.Priest@us.army.mil*

2. Institute for Simulation and Training, University of Central Florida

*martin@ist.ucf.edu*

3. Research Network, Incorporated

*jsaffold@resrchnet.com*

**Abstract:** The projected use of small unmanned aerial systems (SUAS) in military operations will produce training requirements which go beyond current capabilities. The paper describes the development of prototype training procedures and accompanying research simulations to address this need. We initially constructed a testbed to develop simulation-based training for an SUAS operator equipped with a simulated vertical-lift and land SUAS. However, the required training will go beyond merely training an operator how to pilot an SUAS. In addition to tactics, techniques, and procedures for employment of SUASs, collective training methods must be trained. Moreover, the leader of a unit equipped with SUAS will need to learn how to plan missions which incorporate the SUAS, and take into account air space and frequency management considerations. The demands of the task require the leader to allocate personnel to the SUAS mission, communicate and coordinate with those personnel during the mission, and make use of the information provided. To help address these training issues, we expanded our research testbed to include a command and control node (C2 node), to enable communications between a leader and the SUAS operator. In addition, we added a virtual environment in which dismounted infantry missions can be conducted. This virtual environment provides the opportunity for interactions among human-controlled avatars and non-player characters (NPCs), plus authoring tools to construct scenarios. Using these NPCs, a collective exercise involving friendly, enemy, and civilian personnel can be conducted without the need for a human role-player for every entity. We will describe the results of our first experiment, which examined the ability of players to negotiate use of the C2 node and the virtual environment at the same time, in order to see if this is a feasible combination of tools for training development.

## 1. INTRODUCTION

The demonstrated usefulness of unmanned aerial systems (UASs) has led to a steady increase in their employment for reconnaissance and surveillance over the last decade. One area of research and development concerns the employment of small UASs (SUASs). If SUASs can be made light enough to be man-portable and easy enough for almost any Soldier to operate, they could provide unprecedented situation awareness at the small military unit level. Several types of SUASs are already in use by the military, and the U. S. Army is currently evaluating an SUAS with vertical take-off and land, and hover capability. If the evaluation is positive this system will be deployed. This will create a large training demand, which will require both virtual and live simulation. System operators will require training on systems operation and maintenance, and their leaders will require training on system management and a means to conduct team-level mission exercises [1]. Anticipating this training demand, we developed a research simulation testbed to explore how simulation could best be used for these purposes. By analogy with the successful

use of simulation for pilot training, we initially focused on developing simulation-based operator training exercises and evaluating the usefulness of various performance measures for their ability to contribute to a standards-based simulation training curriculum [2], [3], [4]. We have subsequently expanded the testbed to allow for team-level mission exercises. This paper will describe the evolution of the testbed.

## 2. OPERATOR TRAINING SIMULATION

Considering the extensive use of simulation-based training in manned aviation [5], it seems natural to extend the use of simulation training to unmanned aviation systems. As such, in collaboration with the Institute for Simulation and Training at the University of Central Florida, the U.S. Army Research Institute began by developing a research testbed to develop and test simulation-based training exercises as well as performance measures which would be appropriate to use for training standards.

## 2.1 The simulated SUAS

The characteristics of the simulated SUAS (SSUAS) were loosely based on a prototype Micro-Aerial Vehicle (t-MAV) developed under the Defense Analysis Research Project Agency's MAV Technology Demonstration. The t-MAV was a ducted-fan vertical lift vehicle which could hover, rotate in place, and travel at an airspeed of up to six knots under manual control, and over 25 knots under waypoint navigation. We incorporated these characteristics into the SSUAS. We developed a flight model, which, similar to the t-MAV, caused the vehicle to tilt forward one degree for every knot of forward speed, and which gave it some inertial properties (e.g., when in forward movement, it took time to actually stop and assume a hover after the hover command was issued). Like the t-MAV, the SSUAS was equipped with two cameras, one facing forward, and one facing downward. The tilt produced by forward movement of the SUAS tilted camera angles (e.g. while moving forward, the downward camera pointed somewhat behind the vehicle). Some features of the SSUAS were configurable, so that the effect of various aspects on operator performance could be investigated. For example, the cameras could be fixed or have the ability to pan and zoom.

## 2.2 The operator control unit

The operator control unit (OCU) was designed to be reconfigurable, so that the effect of OCU design on operator performance could be investigated. For example, the OCU display could be configured to show one camera view at a time or both camera views simultaneously. Figure 1 shows one particular OCU configuration, and illustrates several of the potential features. In particular, an altimeter on the left edge, the camera view with a heading tape, an overhead map showing the SSUAS position, and flight controls. Icons on the tool bar controlled functions such as switching camera views and taking still photographs. Though not illustrated here, the OCU could also provide the operator the opportunity to program automated flight paths based on preset waypoints, and launch or interrupt these automated missions. In manual mode, the OCU could be controlled by a mouse or by a two-thumb stick game controller.

The OCU is written in Linux using freely available software (Open Scene Graph for rendering and OpenAL for audio) and requires no additional licenses to be purchased. The SSUAS and a base station are transmitted using

the DIS protocol so that both can be displayed in other systems. Any modem PC and video card can satisfactorily run the OCU.



Figure 1: Example OCU Interface

## 2.3 The synthetic environment

The SSUAS could be operated in one of two synthetic environments, each based on an actual Military Operations in Urban Terrain (MOUT) training areas. Both simulated small towns, but differed in their specific features. Any OpenFlight database can be loaded although the overhead map feature requires an additional image file. The map can be an actual map image or an aerial view depending on the need. In addition to features inherent in these environments, other entities could be imported through Distributed Interactive Simulation (DIS) communication protocol (we used OneSAF Testbed Baseline v2.5). This allowed for the display and routing of various types of vehicles and dismounted personnel.

## 2.4 Research findings

We developed operator training missions intended to train manual flight control, concentrating on two types of missions. One focused on flight skill. To conduct these we set up obstacle courses delineated by poles placed in various configurations. Trainees had to learn to manually maneuver the SSUAS along a designated path around the poles. The other type of mission focused on using the SSUAS for reconnaissance. In these missions, trainees maneuvered freely around the environment in order to find and photograph targets (both dismounted personnel and vehicles).

Participants were given an initial introduction to the OCU, and the opportunity to practice simple maneuvers and functions. They were subsequently asked to complete a series of missions during which performance measures were collected. Different trainees were given

different OCU configurations, and performance effects of these configurations were examined in order to investigate the sensitivity of various measures (e.g., number of collisions, number of targets detected, time to complete mission). Our aim was to determine which performance measures were sensitive enough to be useful for future standards-based simulation training. Our results suggest that temporal measures (time to complete mission) is the most sensitive measure we assessed, and therefore likely the most useful for setting standards (e.g., must be able to complete mission within a set time with no collisions). For further details on this research, the reader is referred to [2], [3], [4].

### 3. TEAM-LEVEL MISSION SIMULATION

Like individual pilot training, team training in aviation has benefited from simulation [6], [7]. For SUSAs, the makeup of the team may depend on the specific system, but in the context of a small Army unit, it will likely consist of at least the operator and a robotics noncommissioned officer (NCO), and/or the unit commander. Effective team performance will require team members to coordinate, communicate, and hold a shared understanding of the task, their equipment, and their teammates [7]. Thus, it is not sufficient to merely train an operator how to operate a system. The leaders in a unit equipped with an SUAS will need to learn how to plan missions that integrate the SUAS, and take into account air space and frequency management considerations. The leader will need to allocate personnel to the SUAS mission, communicate and coordinate with those personnel during the mission, and make use of the information provided by those personnel [1]. The unit will need to learn tactics, techniques, and procedures associated with the employment of the SUAS, and collective training methods will be required to accomplish this. To help address these training issues, we expanded our research testbed to include a command and control node (C2 node), to enable communications and information exchange between a leader and the SUAS operator. In addition, we added a virtual environment in which dismounted infantry missions incorporating use of the SSUAS can be conducted.

Specifically, the system was expanded to include three separate elements: 1) GDIS: a virtual immersive environment that replicates one of the synthetic MOUT sites and can be populated with human-controlled avatars and semi-intelligent computer generated forces (non-

player characters or NPCs). 2) C2 node: a command and control node enabling communications between the commander and SSUAS operator, and 3) the OCU: the pre-existing OCU was modified to allow for interaction with the C2 node. As a whole, this system offers a great deal of flexibility in that participants may operate avatars in GDIS, and/or may operate the C2 node or the OCU, thus simulating an entire small unit equipped with an SUAS. The SSUAS is visible to characters in GDIS and can “sense” the GDIS environment and transmit these sensor images to the OCU and/or C2 node.

#### 3.1 C2 Node

The C2 node was created to simulate a nominal command and control station. Like the OCU, the interface is reconfigurable. For example, the experimenter can choose to have blue force tracking displayed on an overhead map or not. Or the experimenter can choose to allow the C2 to receive streaming video from the SSUAS or not. Figure 2 shows one particular C2 node configuration and shows many of the features available, including an interactive map grid that shows the location of the SSUAS, NPCs, and players within the GDIS environment, a window for receiving pictures and/or streaming video from the SSUAS/OCU, text windows for sending and receiving messages, and menus for mission planning. Mission planning includes inserting routes, no fly zones, and flagging entities, as well as sending and receiving information (e.g., mission plans, texts).

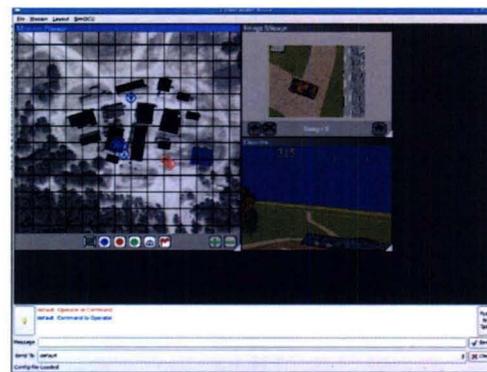


Figure 2: C2 node Interface

#### 3.2 Modified OCU

The OCU was modified to include features and functionality that enable communication and coordination with the C2 node. Figure 3 shows a specific modified OCU configuration, including some of the new features. These include a new

window to display still photos, which can be labeled and sent to the C2 node, and a window for exchanging text messages with the C2 node. The OCU can also receive mission plans from the C2 node. Similar to the C2 node, blue force tracking can be enabled or disabled, so that the effect of having this capability on mission performance can be investigated. All of the information exchanged between the OCU and C2 node is time stamped and saved to a text file for subsequent analysis.

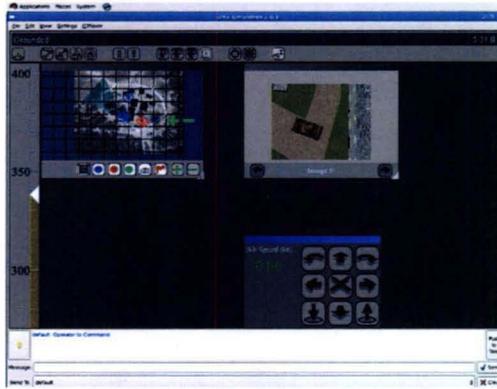


Figure 3: Modified C2 node interface

### 3.3 GDIS Virtual Environment

The GDIS virtual environment (developed by Research Network Inc.) has the ability to function alone, allowing multiple distributed human players to control avatars, which can maneuver, shoot, emote, and communicate with other players. We have integrated GDIS with the OCU and C2 node, such that the SSUAS is an entity that appears in GDIS, but is controlled from the OCU. In addition, we have added substantial artificial intelligence capabilities (AI) to allow for semi-automated NPCs. This allows a multi-person scenario to be conducted without requiring a human role-player for every character. Figure 4 shows a screenshot from GDIS with NPCs and the SSUAS visible. The system is user-friendly with regard to the development of scenarios, having relatively sophisticated AI specified by menu-based authoring. Scenario authors can add NPCs, assign them to teams, and assign individuals or teams to waypoint-based routes. Authors can also add operational vehicles and a range of objects, including improvised explosive devices (IEDs).

NPCs in GDIS have a number of settings, including team membership, weaponry, competency (i.e., novice, expert fighter), and rules of engagement (ROEs). Routes can be

created using waypoints, and specific behaviors can be assigned to waypoints. NPCs can then be assigned to the routes and will act out the behaviors that are associated with each waypoint when they are reached. For example, "patrol" can be assigned to a waypoint, and an NPC arriving there will engage in patrolling behavior according to a selected amount of time and a selected radius of the waypoint. Behavioral characteristics can also be altered at waypoints. For example, ROEs can be changed so that they are different inside vs. outside of a town. Moreover, in order to make scenario branching more sophisticated, contingencies can be set up at waypoints. This allows behavior to change according to context. For example, the waypoint may direct the NPC (or NPC team) to go to the next waypoint only if another NPC team has reached another specific location. These if/then contingencies are specified through the menu-based authoring system in the same manner as the more simple waypoint-associated options.



Figure 4: GDIS system environment with NPCs and the SSUAS

Finally, some team-level behaviors have been constructed (e.g., building search), so that an NPC team will perform the behavior in a coordinated way, without requiring the scenario author to script the behavior of each team member. Using this scenario authoring system, the interactions of NPCs with one another, with human-controlled avatars, and with the environment can be made to appear complex and realistic. Figure 5 shows a screen capture of some of the menus for scenario generation. Specifically, the screen shows the assignment of players and squads to specific routes.

In addition to specifying NPC behavior in pre-constructed scenarios, mission controllers can take over control of an NPC during scenario,

manually manipulate its behavior, and subsequently return it to autonomous mode.

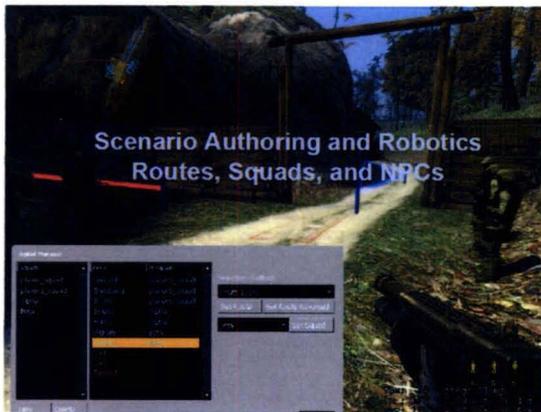


Figure 5: GDIS Route menu

GDIS was also created with an eye towards future compatibility with military systems and software. With the explosion of available game engines available to the Army, this research (and GDIS SimBridge) is being designed to leverage off these available technologies easily and allow for insertion of latest technologies as they become available. However, the military is currently using several different types of game-based applications, so there is no standard game engine being used as the basis of these applications. As a result, GDIS currently interfaces with the HL2 engine (Mod Type) and the GameBryo engine (Source Type).

### 3.4 Potential of the test bed for small unit missions with SSUAS

The integration of the OCU, C2 node and GDIS environment allows for the simulation of small unit level dismounted missions, which incorporate the use of an SSUAS. The operator of the OCU views the GDIS environment through the OCU video imagery, and can exchange information with the unit commander equipped with the C2 node. The unit commander can either be in a notional command center, or actually in the GDIS environment, by providing him or her with a computer running GDIS in addition to the C2 node.

In order to determine whether this latter configuration was feasible for a user, we conducted a pilot experiment to assess the ease or difficulty a person would have if assigned to use the C2 node and control an avatar in GDIS at the same time. We varied the workload demands of the C2 node (low or high) and the

GDIS task (low or high), and each participant completed four missions representing the combination of these conditions. After some practice maneuvering their avatar in GDIS, and basic training on the C2 node, participants were given missions in which they were asked to visit specified buildings (in GDIS) and classify (on paper) the people they discovered as Soldiers, doctors, or refugees. In addition, they had to report (via text messaging using the C2 node) the presence or absence of specified targets in pictures sent to them through the C2 node. This represented the low-low workload condition. For the high C2 node workload condition, another C2 node task was added: on request, reporting the position of the SSUAS using the C2 node map grid. For the high GDIS workload condition, another GDIS task was added: on request, report (by text message) the location of a specific person in GDIS, using the GDIS interactive map. The order in which the missions were conducted was counterbalanced across participants. A metric that considered both accuracy and time to complete each mission was used to evaluate performance.

We found that our manipulation of workload had a far weaker impact on performance than simply the opportunity to practice. Regardless of the workload condition, performance improved over time from mission one to three, with performance on missions three and four roughly equivalent. Individual difference factors (e.g., video game experience and spatial ability) also influenced performance. Specifically, participants with higher spatial ability (as measured by the Cube-Comparison Test [8]) tended to perform better ( $r = .39, p < .05$ ). Game-playing habits also affected performance. The time spent playing video games ( $r = .71, p < .05$ ) and the frequency with which they played ( $r = .68, p < .05$ ) correlated positively with the overall score which considered both accuracy and speed of mission.

## 4. CONCLUSION

Our initial research using all elements of the testbed indicate that it is feasible for a participant to work with the C2 node and control an avatar in GDIS at the same time. This research was conducted before we had the full AI functionality in the NPCs described above. Now that those capabilities are in place and we have established that people can work well with the systems, we can begin to craft more realistic scenarios. These will enable us to examine the coordination and communication

issues that units will have in integrating use of an SUAS into a mission, as well as methods to overcome such issues through training, the use of standard operating procedures, and the development of tactics, techniques, and procedures.

8. ETS (1976). *Kit of Factor-Referenced Cognitive Tests*. Princeton: N.J.

Note. Opinions expressed in this paper are those of the authors and do not represent an official position of the U.S. Army or the Army Research Institute.

## 5. REFERENCES

1. Durlach, P. J. (2007). PACERS: Platoon Aid for Collective Employment of Robotic Systems. ARI Research Report 1876, Arlington, VA.
2. Billings, D.R., & Durlach, P.J. (2008). Effects of input device and latency on performance while training to pilot a simulated micro-unmanned aerial vehicle. (ARI Technical Report 1234). Arlington, VA: U.S. Army Research Institute for the Behavioral & Social Sciences.
3. Durlach, P.J., Neumann, J.L., & Billings, D.R. (2008, May). Training to operate a simulated micro-unmanned vehicle with continuous or discrete control. (ARI Technical Report 1229). Arlington, VA: U.S. Army Research Institute for the Behavioral & Social Sciences.
4. Billings, D. R. & Durlach (submitted for publication). "How input device characteristics contribute to performance during training to operate a simulated micro-unmanned aerial vehicle," *Human Computer Interaction*.
5. Salas, E.; Bowers, C. A.; Rhodenizer, L. (1998). "It is not how much you have but how you use it: Toward a rational use of simulation to support aviation training," *International Journal of Aviation Psychology*, vol. 8, issue 3, pp. 197-209.
6. O'Connor, P.; Campbell, J.; Newon, J.; Melton, J.; Salas, E.; Wilson, K. A. (2008). "Crew resource management training effectiveness: A meta-analysis and some critical needs," *International Journal of Aviation Psychology*, vol. 18, issue 4, pp. 353-368.
7. Salas, E.; Rosen, M. A.; Burke, C. S.; Nicholson, D.; Howse, W. R. (2007). "Markers for enhancing team cognitive in complex environments: The power of team performance diagnosis," *Aviation, Space, and Environmental Medicine*, vol. 78, Special issue: Operational applications of cognitive performance enhancement technologies. pp. B77-B85.

# Methods Used in Game Development to Foster FLOW

Isaac Ben Jeppsen – Producer

*Breakaway Ltd*

*ijeppsen@breakawayltd.com*

**Abstract.** Games designed for entertainment have a rich history of providing compelling experiences. From consoles to PCs, games have managed to present intuitive and effective interfaces for a wide range of game styles to successfully allow users to "walk-up-and-play". Once a user is hooked, successful games artfully present challenging experiences just within reach of a user's ability, weaving each task and achievement into a compelling and engaging experience. In this paper, engagement is discussed in terms of the psychological theory of Flow. I argue that engagement should be one of the primary goals when developing a serious game and I discuss the best practices and techniques that have emerged from traditional video game development which help foster the creation of engaging, high Flow experiences.

## 1. INTRODUCTION

Serious games and simulations are often focused on training or teaching a specific skill or procedure, and consequently there is a tendency to focus on the fidelity of the product at the expense of other salient factors. In contrast, the focus of a game developed purely for entertainment purposes, is to engage the user with fidelity taking a supporting role. An example of this shift in focus is the differences between usability analysis performed on traditional software and user experience analysis performed on games [1]. Usability as it applies to software is usually viewed from the perspective of business software; the goal is to provide a tool or tools to create or modify an artifact-- such as a document or picture. In general, this perspective is concerned with trying to create interfaces which are transparent to the user. Conversely, games provide a tool for users to experience an artifact; such as the experience of being a criminal, hero, doctor, etc. While straight forward, this shift in focus from transparent tool use to compelling experience has important implications for serious game development teams. It is not sufficient to make an easy-to-use tool; it must also engage the user in an optimal manner.

Most of us have experienced the gratification that arises from having sufficient skill to deal with a problem or challenge just outside of our comfort level. In other words, an experience characterized by the feeling of being so

engrossed and focused on the task at hand that everything else is eclipsed. This is referred to in sports as "being in the zone". This state of optimal experience is what Dr. Csizenthmihalyi refers to as a state of Flow [2]. Creating optimal experiences is what games, both for serious and for entertainment purposes, are all about. Developers strive to create this optimal experience by immersing the user not only in an accurate environment, but also in one where the player is compelled to explore and to experiment. Fidelity alone is not enough to create immersion. Immersion arises out of a combination of both salient details and a unique perspective or role experienced by the user.

In what follows, I give a brief summary of the theory of Flow and how it relates to games. Then, I discuss how the practices and structures in game development have evolved to support the creation of optimal experience. Though originating from entertainment game development, the practices discussed should be generalizable to any immersive application.

## 2. FLOW AND USER EXPERIENCE

The theory of Flow was motivated by Dr. Csizenthmihalyi's desire to identify what happiness is, and the elements that contribute to positive emotions in general. From the experiments he conducted to investigate this psychological state, he found that naïve notions of happiness (e.g. care free, stress free, etc.) do not explain why people

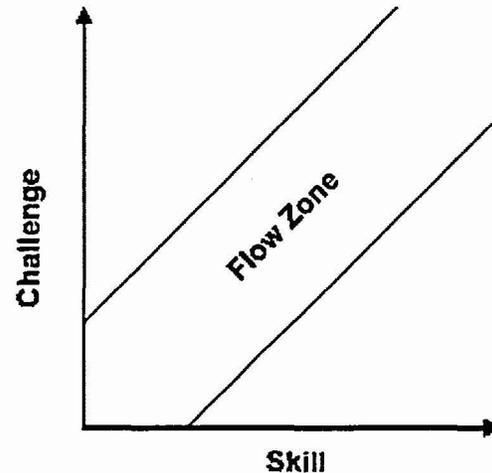
performing demanding and complicated tasks report a sense of well being and enjoyment. To explain the phenomena of optimal experience, he developed a theory that he coined Flow. Flow is a set of eight heuristics that describe common features that contribute to an optimal experience. The eight components are [2]:

- Clearly defined goals
- Concentration on task at hand
- Merging of action and awareness
- An altered sense of time
- Clear and responsive feedback
- Balanced level of challenge and difficulty
- A sense of control over the task at hand
- A challenging task requiring skill to execute

These eight components serve as a good definition for engagement as it pertains to games. Through a process of evolution facilitated by the fiercely competitive commercial game market, successful game development teams have been deliberately leveraging some or all of these heuristics to deliver the compelling and immersive experience gamers have come to expect.

The most obvious example of the use of Flow in games is the game 'fIOW' [3] created by Jenova Chen. While this game started as a thesis project, it has since been released commercially to both critical and public acclaim. In this game Chen uses the two dimensions of challenge and skill to design a game that dynamically adjusts the difficulty of game play through the user's own choices. This is in contrast to indirectly controlled methods that analyze game play and adjust difficulty by changing variables such as enemy spawn rates and weapon damage. An indirect method has several problems, including multiplayer balancing, feedback to the user and a tendency to create a less than optimal experience.

As can be seen in figure 1, the goal of the game 'fIOW', and indeed any game, is to keep the player "in the zone".



**Figure 1** Showing the zone between challenge and skill where the user has an optimal experience

### **3. GAME DEVELOPMENT TEAMS – ANATOMY, STRUCTURE, PROCESS**

Due to the fundamentally subjective experience of games and the wide range of people that play them, simply knowing the components of Flow does not imply this knowledge will translate to the implementation. As such, creating Flow in games remains more an art than science. Even efforts that go into much more detail than Flow, such as the heuristics of PLAY [4], acknowledge the fundamentally artistic nature of game development.

The artistic nature of games requires a different approach to the development process compared to other commercial software development. These differences can be seen in both the team structure and in the development process in general. For example most software development teams don't include visual artists throughout the processes. Having this resource throughout development allows for more sophisticated mockups to be created giving quicker and more accurate representations before significant resources are devoted. This is just one of many examples of the subtle to gross

differences in game development team dynamics compared to more traditional efforts.

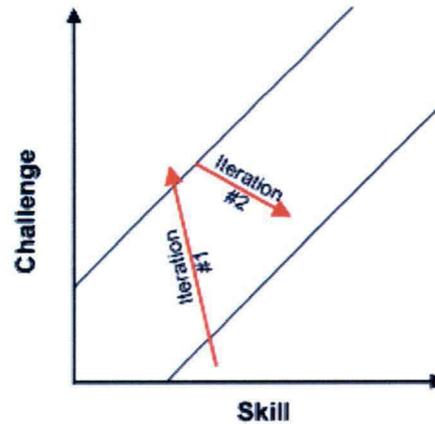
Due to the competitive nature of commercial game development it is often hard to get a detailed picture of a particular development studio's structure and practices; however there has been a growing body of work by both academic researchers and game development professionals that allows for some broad generalizations to be made. The main high level components that directly contribute to the development of games that deliver a flow experience are:

- Rapid iteration
- Motivated Multidisciplinary teams
- Vision holder(s)
- Play testing

### 3.1 Rapid Iteration

As any movie or game executive can attest, ideas are abundant and playable products are scarce. Only the details that emerge through the production process allow for a given idea to be evaluated as being "good" or "bad". These details are even more important in game development due a game's interactive nature. For a game to be successful it has to be "fun," and at present, there is no way to determine the "fun" of a game without playing it. Game development takes an initial idea such as, "let's manage things," and through a series of successive iterations discovers "where the fun is". This processes starts by using mockups, concept documents, and playable prototypes and grows into the final polished product that (hopefully) gets shipped.

While rapid iterative development is beginning to get a lot of traction across the software development industry, successful game developers have been early and often trail-blazing practitioners. One of the reasons for the rapid adoption of Agile production methodologies such as Scrum [5] by game development studios is the emphasis on a "playable" product at the end of each short milestone or "sprint" (usually between 2 and 6 weeks). As illustrated in Figure 2, by taking small incremental steps, the team is able to make corrections to keep the game fun before it becomes infeasible to change.



**Figure 2** Rapid iteration used to evaluate "fun" Iteration #1 starts with the game being too easy and ends being a little too hard.

The importance of rapid iteration should not be underestimated. As games become more complex the ability to rapidly create, test and experience a game will only become more important.

### 3.2 Motivated Multidisciplinary teams

Modern day AAA games (an AAA game refers to a high budget, high production value game,) are incredibly complex endeavors often involving core team sizes in the hundreds. Smaller independent game titles can have teams ranging anywhere from one person to forty or fifty, the former being very rare for a game of any significant complexity. At a high level production, teams are composed of designers, software engineers, artists, sound engineers, quality assurance, subject matter experts and production staff. In contrast to development teams in other industries, most if not all of these positions are involved from the conception through the completion of the project. Involvement of all disciplines early in the project is due to the user centered design focus inherent in video game development. At the beginning, artists will create concept mockups that illustrate not only the functional elements of the user interface, but the aesthetics as well. In tandem, designers will be working out the logical flow of the interface from the user perspective, and engineers will be focusing on the logical flow as well as the supporting infrastructure from the software perspective. This parallel development helps to give a more complete picture of the product

early and at every stage of the production processes.

A maxim from usability experts states "know thy user, for they are not you" and while this largely holds true for game usability, game development teams tend to be more diverse and enthusiastic than developers in other markets of software development. For example, developers are almost universally consumers of the products they develop. If asked, a typical game developer, in any discipline, will likely categorize him/herself as a "gamer" (i.e., someone who plays games regularly if not obsessively). Game development teams in this respect, more closely reflect the movie industry in which professionals that create movies got into the career due to a passion for the medium as a consumer. Having developers that understand the "user" viewpoint helps to internalize a user-centered design approach. And while this doesn't guarantee that the end result will be widely appreciated, it does allow for larger leaps forward in creating and maintaining Flow-inducing games than a strictly formal approach could achieve.

### 3.3 Vision Holder

While the previous sections have focused on the team as a whole, a single authoritative vision holder is essential to insuring that the end product does not suffer from the "design by committee" effect. A vision holder also ensures that the overall game experience does not become confused and diluted due to a lack of cohesion. Again the film industry provides an excellent comparison, in the role of a director. While there is not always a title on a game development team that directly corresponds to that of a movie director, the role is often filled by either the producer or lead game designer.

The role of the vision holder is to internalize the details of the game and to integrate the contributions of all the team members into a cohesive whole. The complexity of modern games involves many different disciplines and specialties and requires that someone bring it all together to create something greater than the sum of its parts.

### 3.4 Play testing

As stated earlier, game development teams typically have a lot of "users" embedded in the form of developers; and while this helps to guide development, there is an inevitable feedback loop or echo chamber effect that can cause developers to lose touch with their larger audience. This is where user experience analysis or "play testing" [1] comes in. Play testing is very similar to usability testing in business software with a few key differences. In play testing, the goal is to measure not only the ease with which the user is able to interact with the game, but also the level of engagement of the user. Play testing in recent years has become more sophisticated with larger studios and publishers using a broad array of measurements and analysis tools to refine the user's experience. These tools include traditional methods such as surveys and focus groups, and more technologically driven approaches such as gaze tracking and EEG analysis.

While the costs of both the traditional and technological methods are continuing to decrease, they are currently prohibitive for smaller studios. Smaller studios may not be able to bring the same resources to bear; they can and do employ informal versions of the above with varying degrees of success.

The choice of play testing method should be driven by the development methods mentioned previously. Of these, the ability to integrate the play testing into a demanding and rapidly changing production schedule is of utmost importance. In order for play testing to be useful it must be done in a way that encourages rapid iteration. Formal play testing studies can take months if not years to organize and gather the data required, representing an unacceptable risk in terms of both time and resources required.

The emergence of brain imaging technologies combined with advances in neuroscience and cognitive models offer promising directions for a more automated and efficient approach to play testing. By combining cognitive models of various player profiles with psychophysiological data collected as people play

games, it may become possible to create cheaper and more reliable methods to gauge the flow of the user. The automation and consequent decrease in time and resources that these approaches promise may make them excellent candidates for inclusion in future game development efforts.

#### 4. SUMMARY

The number of problems that games can be applied to has just begun to be explored. A few projects have begun to explore using games to harness human's spatial and pattern matching abilities. For example "Folding@home" [6] uses a game to help determine how proteins fold. As the number, difficulty and complexity of problems that gaming is used to solve increases, it is imperative that development teams and those who manage them understand what motivates people to play games, and how to foster an environment that nurtures the creation of games people want to play.

#### 5. REFERENCES

1. Thompson, Clive. "Halo 3: How Microsoft Labs Invented a New Science of Play". *Wired Magazine*. 2007, 15.09.
2. Csikszentmihalyi, Mihaly. *The Psychology of Optimal experience*. New York : Harper & Row, 1990.
3. Chen, Jenova. *fiOw. Welcome to Flow in Games*. [Online] [Cited: August 27, 2009.] <http://intihuatani.usc.edu/cloud/flowing/>.
4. *Game Usability Heuristics (PLAY) For Evaluating and Designing Better Games: The Next Iteration*. Desurvire, Heather and Wiberg, Charlotte. San Diego : Springer, 2009. HCII 09.
5. Schawber, Ken and Beedle, Mike. *Agile Software Development with Scrum*. Upper Saddle River : Prentice Hall, 2001. 0-13-067634-9.
6. *Folding@home - Main*. Stanford University. [Online] [Cited: August 27, 2009.] <http://folding.stanford.edu/English/Main>.
7. *Flow in Games (and Everything Else)*. Chen, Jenova. April 2007, *Communications of the ACM*, pp. 31-34.



# Making Games not Work: Paradoxes Embedded in Game-Based Training and Concepts for Overcoming Them

Phillip N. Jones; Taryn Cuper

MYMIC LLC

*phillip.jones@mymic.net; taryn.cuper@mymic.net*

**Abstract:** An interest in game-based training solutions is natural. All one has to do is watch someone fully engaged in a modern game to see the potential of harnessing that attention for training. However, the reality of game-based training has not fully satisfied these expectations. This paper explains two paradoxes that must be overcome for games to support training. These paradoxes are a result of the realities of the basic human condition clashing with the requirements of learning theory. Both paradoxes arise from the concept of “engagement” that is central to games. The first comes from a more robust definition of engagement, which is the condition of Flow or Optimal Experience. Flow is the state game developers want to see in users. One aspect of Flow is loss of sense of self as the individual becomes immersed in the experience. The paradox arises because this loss of self directly contradicts the learning requirement of self-reflection. The second paradox comes from theories of play, which state in part that play requires a level of individual freedom. The contradiction arises when game-based play must be harnessed to an organizational training program or regimen. The paper will discuss these paradoxes in the context of an effort to design a game-based training modality to train combat medics and will close with a review of compensating strategies identified by the designers. The paper will provide information important to anyone interested in conceptualizing and designing game-based training.

## 1. INTRODUCTION

This paper describes the identification of and strategies to overcome two key paradoxes inherent in leveraging “games” to support organized learning or training, or the creation of “serious games.” The genesis of this work was a six month effort to conceptualize and design a computer-based modality for training Combat Medics and Corpsmen to react to an Improvised Explosive Device (IED), specifically in terms of site management. As the intended training audience is assumed competent in specific medical diagnostics and treatment, the training solution focuses on higher level cognitive requirements, or decisions on the management of casualty care in a complex and hostile environment. A key requirement for the computer-based training modality was that it be “sim-game based”, or more specifically, of the “simulation / strategy game genre.”

In order to maximize value to the eventual customer, the research team approached the conceptualization and design of this system grounded research into operational and training environment requirements and constraints. For this project, that grounded research included, among other efforts, a review of both operational and theoretical literature as well as input from subject matter experts. The review of theoretical literature, which will be described in detail here,

included examining the concept of games in general and serious games in particular.

The intent of the research into games and serious games was to provide cognitive rigor to the advantages of pursuing effective learning through a serious game strategy such that the eventual conceptualization and design would incorporate these advantages. The investigation, however, identified issues in a rigorous definition of the game construct that culminated in the identification of two true paradoxes. Having identified the two paradoxes, the team built strategies to mitigate their impact.

## 2. DEFINING THE GAME CONSTRUCT

Anyone who has observed another immersed in a computer-based game can appreciate the attraction of serious game-based training. The lead author has many times observed the intense commitment of his teenage children towards advancing their fantasy character or building a better virtual community, wishing he could transfer that commitment to far more critical but less exciting pursuits, such as math or English. A search of Google Scholar for “game based” returns over 17,400 hits; 778 of those hits are for the first eight months of 2009. Thus, it is no surprise that stakeholders in organizational and individual performance are

pursuing serious game based learning approaches more and more. Despite their increasing efforts, the author still does not have to tell his teenagers to get off the math game and get to bed.

The first question when asked to develop a game-based solution is: What is a game? The answer is not obvious. There are multiple characterizations of a game, seemingly based upon individual perspectives [1]. Webster's New World Dictionary defines game as, "any form of play or way of playing; amusement; recreations; sport; frolic; play" [2]. Clark Abt, who penned the term serious game in his 1970 book, pre-artificial intelligence (AI), *Serious Games* [3], defined a game as:

Reduced to its formal essence, a game is an activity among two or more independent decision-makers seeking to achieve their objectives in some limiting context. A more conventional definition would say that a game is a context with rules among adversaries trying to win objectives. We are concerned with serious games in the sense that these games have an explicit and carefully thought-out educational purpose and are not intended to be played primarily for amusement.

In 2005, Mike Zyda, Director of USC's GamePipe Laboratory, provided a post-AI definition of a serious game as [4]:

"a mental contest, played with a computer in accordance with specific rules that uses entertainment to further government or corporate training, education, health, public policy, and strategic communication objectives"

Both of these characterizations can encompass any number of training modalities, including modalities that the serious games community probably would not accept within their portfolio. In recognition of the untidiness present when defining a game, the authors changed strategy, opting away from a game as a discrete phenomenon and towards defining a game construct through a set of game traits, traits that would exist in various numbers and depth in the variety of training modalities.

Following this new strategy, the literature review identified twenty separate game traits, certainly only a partial list of total set. Several of these

traits were clearly classifiable under two meta-traits: Play and Optimal Experience or Flow. The exploration into how to incorporate these meta-traits within the serious game-based solution demonstrated a paradox associated with each meta-trait. These paradoxes stand in the way of development of effective serious game-based training solutions.

### 3. GAME META-TRAITS

#### 3.1 Play

As seen from the above and other definitions, play is an inherent feature of games. Games could be said to be the application of play. This raises the question of what is play, a question as difficult to answer as the earlier question of what is a game.

Again, the answer is not readily apparent. Brian Sutton-Smith in his book, *The Ambiguity of Play* [5], takes the perspective that play is so intertwined in human experience and that there is such diversity in the forms and manners of play that it cannot be separately defined. Rather, it can only be approached through discussion, or "rhetoric" of varied approaches towards the study of play.

French philosopher Roger Caillois studied play and identified six essential elements [6]. These are:

1. Free: in which playing is not obligatory; if it were, it would at once lose its attractiveness and joyous quality as diversion;
2. Separate: circumscribed within limits of space and time, defined and fixed in advance;
3. Uncertain: the course of which cannot be determined, nor the result attained beforehand, and some latitude for innovations being left to the player's initiative;
4. Unproductive: creating neither goods, nor wealth, nor new elements of any kind; and, except for the exchange of property among the players, ending in a situation identical to that prevailing at the beginning of the game;
5. Governed by rules: under conventions that suspend ordinary laws, and for the moment establish new legislation, which alone counts;

6. Make-believe: accompanied by a special awareness of a second reality or of a free unreality, as against real life.

The paradox that is serious play arises from the conflict between two of these essential elements and training. Effective play requires freedom, uncertainty, and lack of productivity. Yet, training is an organizational imperative and training programs are evaluated based upon their support to organizational objectives [7]. Effective and focused training is conducted within a well defined program with designated performance objectives and time and other resource constraints and synchronized with other training events. The conflict between the necessary freedom of play and the requirements and constraints of training is the first decisive paradox to be overcome.

### 3.2 Optimal Experience

A second necessary meta-trait of games is engagement [8] [9]. Again, there is the question of definitions. What is engagement and where does it come from?

Psychologist Mihaly Csikszentmihalyi provided an answer with his theory on "Flow" or optimal experience [10]. Through empirical research, Dr. Csikszentmihalyi identified a condition where human happiness was obtained in a balance between individual capability and challenge, or when information coming into individual awareness is harmonious with individual goals. This harmonization creates optimal experience, or the opposite of psychic entropy.

Dr. Csikszentmihalyi identified eight elements for establishing flow. These are:

1. The individual confronts tasks that he/she has a chance of completing or challenging tasks that require skills
2. The individual must be able to concentrate on the activity
3. The activity has clear goals
4. The activity provides feedback to the individual.
5. The activity provides a deep but effortless involvement by the individual
6. The activity allows the individual to exercise a sense of control
7. During the activity, the sense of self disappears, but emerges stronger afterwards
8. The sense of time is altered

The fifth and seventh elements, the requirement for deep but effortless work and the loss of the sense of self, generate the second serious game paradox. Effective adult learning requires recognition by the learner that they possess behaviors that, to reach individual objectives, necessitate change [11]. The learner has to perceive their actual abilities against desired abilities as well as the path to close the gap.

In addition, learning requires cognitive effort. For example, in the theory of self-regulated learning, self-reflection is a significant part of the learning process [12] and has been linked to success in learning [13]. Similarly the American Psychological Association identifies cognitive and meta-cognitive learner-centered principles, to include [14]:

- The learning of complex subject matter is most effective when it is an intentional process of constructing meaning from information and experience.
- The successful learner, over time and with support and instructional guidance, can create meaningful, coherent representations of knowledge
- The successful learner can link new information with existing knowledge in meaningful ways.
- The successful learner can create and use a repertoire of thinking and reasoning strategies to achieve complex learning goals.
- Higher order strategies for selecting and monitoring mental operations facilitate creative and critical thinking.
- Learning is influenced by environmental factors, including culture, technology, and instructional practices.

Thus, in a serious game-based training event, the more perfect the game, the less perfect the learning, as Flow indicates effortless involvement and learning requires effort. In addition, Flow requires loss of the sense of self, while learning requires self-evaluation. Last, Flow requires cognitive focus towards the immediate tasks, while learning requires focus on tasks over time and experience.

### 4. MITIGATING META-TRAIT PARADOXES

As previously stated, this investigation into the nature of games was initiated by a project to conceptualize and plan a game-based modality to train Combat Medics on IED-ambush site

management. Along with identifying desired game traits for the eventual solution, the study team also conducted literature reviews and interviews of subject matter experts to identify performance competencies as well as operational and training conditions and constraints. From this effort, the team developed an extensive set of training modality requirements and then compared over 100 models and simulations to identify candidate systems. Concurrently, the team developed the desired user's experience to maximize learning effectiveness and efficiency.

A central part of this conceptualization and planning effort was the structuring of the training experience to overcome or mitigate the effects of the above paradoxes.

#### 4.1 Mitigating the Play Paradox

The play paradox arises from the conflict between permitting the trainee a sense of freedom and supporting organizational learning requirements through directing him/her along an organizationally directed learning path. The team followed two strategies.

As part of the training solution design, the team included a training matrix that would guide the learner from novice towards expert performance. The training matrix incorporated two axes, cognitive load and environmental conditions. As the trainee progresses through the matrix, he or she would experience greater cognitive load and more complex execution conditions.

Mitigating the play paradox required the inclusion of freedom within this directive matrix. This was accomplished through the provision of multiple missions at each step within the designated training matrix, giving the trainee a choice over which mission he would play in order to fulfill the current training requirement. Giving learners control over the manner in which they learn can be an effective approach to encouraging engagement [15] [16].

The team proposed that the combined coordinates of the two axes provide a small set of similarly challenging scenarios or missions. Each square in the 2D matrix would be a separate scenario or mission. Thus, in a training matrix defined by four cognitive load and environmental condition levels, the trainee would start at the score of two, or at the first load and first environment level ( $1+1=2$ ). Advancing to a

score of three would provide two scenarios ( $1+2=3$  or  $2+1=3$ ) and so forth. One advantage of this solution was that as the trainee reached the end of the training matrix, options reduced until there was only one scenario remaining. This scenario would serve as a gate for movement to a higher matrix.

#### 4.2 Mitigating the Flow Paradox

Mitigating the flow paradox was more challenging and required some potential sacrifice in the training design. This started with the intent to create an environment conducive to the trainee entering a personal flow experience. The first route to this objective was to prioritize environmental and experiential fidelity. The chosen game engine should be able to maximize the realism of the IED-ambush scene as well as accurately model IED-ambush related medical trauma. The team assessed that this dual realism would quickly inject the user into a flow state where they could lose their sense of self.

The second route taken to mitigate the flow paradox was a willful decision to delay self-reflection until after the scenario was complete. This is an acceptance of risk as learning theory would lean towards self-reflection immediately following a learning experience. In the context of Combat Medic training, this would mean a pause from the simulated experience immediately after reaction to the IED. This pause, however, would break the flow and so the team opted to forego it. Instead, the Medic would "stay in character", completing the scenario prior to an opportunity for the necessary self-reflection. At higher levels within the training matrix, this might mean multiple IED-ambushes within the same scenario.

The team compensated for this delay in reflection by emphasizing self-reflection in a post scenario review phase. The review would be guided by an intelligent tutoring capability, but would emphasize review of actions by the trainee, including the recording of those actions. This in turn was intended to support long term learning strategies necessary to learner-centered development.

### 5. CONCLUSION

Serious games do possess a promise of more efficient and more effective learning. However, they are not a panacea. Preparatory work must

be accomplished in order to maximize their potential. This preparatory work includes understanding the real interplay between serious games and humans. This paper has addressed two aspects of that interplay, identified during research into the conceptualization and design of a serious game meant to train Combat Medics. These aspects, Play and Flow, generate serious game paradoxes, which must be mitigated for a successful serious game application.

## 6. REFERENCES

1. Akilli, G (2007) Games and Simulations: A new Approach in Education? In D. Gibson, C. Aldrich, and M. Prensky *Games and Simulations in Online Learning* (pp. 1-20). Hershey, PA Information Science Publishing.
2. *Webster's New World Dictionary* (4<sup>th</sup> ed). (2007). Springfield MA: Merriam-Webster.
3. Abt, C. (1970). *Serious Games*. New York: The Viking Press.
4. Zyda, M. (2005). From Visual Simulation to Virtual Reality to Games. *IEEE Computer*. Vol. 38 (9), pp. 25 - 32.
5. Sutton-Smith, B (1997). *The Ambiguity of Play*, Cambridge, MA, Harvard University Press.
6. Caillois, R. (1958). *Man, Play and Games*. (Meyer Baraxh, Trans.) Urbana IL, University of Illinois Press
7. Kirkpatrick, D (1994) *Evaluating Training Programs: The Four Levels*. San Francisco, CA, Berrett-Kohler Publishers.
8. Quinn, Clark N (2005) *Engaging Learning: Designing e-Learning Simulation Games*, San Francisco, CA, Pfeiffer
9. Becker, K. (2007) Pedagogy in Commercial Video Games, In D. Gibson, C. Aldrich, and M. Prensky *Games and Simulations in Online Learning* (pp. 21-47). Hershey, PA Information Science Publishing.
10. Csikszentmihaly, M. (1990). *Flow: The Psychology of Optimal Experience*. New York, NY. Harper Perennial Modern Classics.
11. Knowles, M, Holton, E. and Swanson, R (2005). *The Adult Learner*, 6<sup>th</sup> Ed. San Diego, CA. Elsevier
12. Zimmerman BJ. (1990). Self-regulated learning and academic achievement: An overview. *Educational Psychologist*. 25(1):3-17.
13. Zimmerman BJ. (2002). Becoming a self-regulated learner: An overview. *Theory into Practice*. 41(2): 64-70.
14. Alexander PA & Murphy PK. (1998). *The research base for APA's learner-centered psychological principles*. In *How Students Learn: Reforming Schools through Learner-Centered Education*. American Psychological Association: Washington, DC. 1-22.
15. Morrison, G.R., Ross, S.R., & Baldwin, W. (1992). Learner control of context and instructional support in learning elementary school mathematics. *Educational Technology Research and Development*. 40:5-13.
16. Cordova DI & Lepper MR. (1996). Intrinsic motivation and the process of learning: Beneficial effects of contextualization, personalization, and choice. *Journal of Educational Psychology*. 88(4): 715-730.



# QuickStrike ASOC Battlefield Simulation: Preparing the War fighter to Win

Richard L. Jones, Lt Col, USAF (ret)  
VT MÄK  
rjones@mak.com

**ABSTRACT.** The QuickStrike ASOC (Air Support Operations Center) Battlefield Simulation fills a crucial gap in USAF and United Kingdom Close Air Support (CAS) and airspace manager training. The system now provides six squadrons with the capability to conduct total-mission training events whenever the personnel and time are available.

When the 111<sup>th</sup> ASOC returned from their first deployment to Afghanistan they realized the training available prior to deployment was inadequate. They sought an organic training capability focused on the ASOC mission that was low cost, simple to use, adaptable, and available now. Using a commercial off-the-shelf simulation, they developed a complete training system by adapting the simulation to their training needs. Through more than two years of spiral development, incorporating lessons learned, the system has matured, and can now realistically replicate the Tactical Operations Center (TOC) in Kabul, Afghanistan, the TOC supporting the mission in Iraq, or can expand to support a major conflict scenario. The training system provides a collaborative workspace for the training audience and exercise control group via integrated software and workstations that can easily adapt to new mission requirements and TOC configurations. The system continues to mature. Based on inputs from the war fighter, new capabilities have been incorporated to add realism and simplify the scenario development process. The QuickStrike simulation can now import TBMCS Air Tasking Order air mission data and can provide air and ground tracks to a common operating picture; presented through either C2PC or JADOCs.

This organic capability to practice team processes and tasks and to conduct mission rehearsals proved its value in the 111<sup>th</sup> ASOC's next deployment. The ease of scenario development and the simple to learn and intuitive game-like interface enables the squadrons to develop and share scenarios incorporating lessons learned from every deployment. These war fighters have now filled the training gap and have the capability they need to train to win.

## INTRODUCTION

The ASOC, a subordinate element of the Air Operations Center (AOC), is the principle command and control (C2) node in the close air-ground battle. It is the senior air C2 node aligned with the Army, (typically co-located with the highest Army echelon in theater) responsible for managing air assets in support of ground maneuver and in the execution of Close Air Support (CAS). As such it plays a critical role in ensuring the Army and the Air Force operate effectively together. Recent experiences in Iraq and Afghanistan underscore the crucial nature of this role.

The ASOC environment is uniquely challenging and demanding. Fighter Duty Technicians (FDTs) and Fighter Duty Officers (FDOs) work in a dynamic and fast-paced setting where communications expertise, tactical knowledge,

planning ability, weapons systems knowledge, coordination skills, and teamwork all play a significant role. Situational awareness (SA) of the ground battle and the air picture is essential. FDO/FDT performance is critical in getting bombs on target quickly and safely, providing direct support for Joint Terminal Attack Controllers (JTACs) or ground forces in contact with the enemy.

## 1. BASELINE EXPERIENCES

Following a combat deployment, the 111<sup>th</sup> ASOC completed a thorough internal debrief of all aspects of the deployment from preparation through execution and re-deployment. One of the most important debriefing items was the inadequate operations spin-up training. From the perspective of the FDOs and FDTs operating in the Joint Operations Center (JOC), pre-deployment preparation had significant flaws that

were summarized in two broad categories: Focus, and C2 Tools.

### **1.1 Focus**

Spin-up training was based on operational assumptions that were not valid in the current conflict. The conflict was non-linear and asymmetric. From an Air Force perspective this model of warfare combined with the theater's unique geography and infrastructure and friendly/enemy Order of Battle meant that CAS was the primary means of support to troops-in-contact situations. That is, there were often no other supporting arms options (artillery, naval gun fire, or army aviation) for troops that needed support. Due to the nature of the conflict, CAS coverage was required across the entire theater practically 24 hours-a-day.

While 111<sup>th</sup> ASOC spin-up training and standard combat mission ready training had been intense and rigorous, it had focused largely on a more traditional linear, symmetric fight. Over a period of time, the focus had slid into a familiar and comfortable rhythm of force-on-force war fighter exercises. This was the wrong focus for Afghanistan.

In hindsight, and assessing a more generic view of ASOC training, the major flaw was not that the 111th were focused on training for one model of warfare rather than the other; it was that the training curriculum and capabilities did not accommodate multiple models of warfare.

### **1.2 C2 Tools**

In theater, the C2 systems and tracking mechanisms were significantly different to those the 111th had trained with during spin-up. The training had typically used large scale, customized C2 systems that covered all aspects of C2 from planning through execution (for example Theater Battle Management Core System – TBMCS). In theater, the tools were the product of networking technology and a blend of simple and flexible software applications. This patch-work C2 architecture was in place due to expediency, operational need, and some very talented individuals who found ways to add functionality and capability piece by piece to a baseline communications network.

Most notably, Internet Relay Chat (IRC) was the primary method of communication in theater. In addition, Excel spreadsheets and common mapping tools such as FalconView added vital

elements. The unit was familiar with the Excel and mapping tools but IRC posed new challenges: learning a new shorthand 'chat' language and a need for transmitting time critical C2 data in (real-time) text format rather than through voice.

Also, operators had to develop softer skills such as an ability to lead and execute the CAS fight appropriately through a text interface. This meant interacting with and coordinating with a wide range of C2 agencies in a distributed, stressful and sometimes confused C2 environment using text alone. This required a unique combination of tactical and doctrinal knowledge, coordination capability, leadership, motivational ability, and of all things, typing and prose capability! The 111th learned to use IRC the hard way – on the job training while in theater.

Following redeployment and debriefing, Squadron leadership began formulating a solution to the training environment so that follow-on deployments had the right training focus and exposure to the correct C2 tools.

It is also significant to point out that during debriefing positive aspects of training were also identified. Defining what went well provided squadron leadership with valuable reference points on areas of training that needed overhaul. Most notably, operations personnel pointed out that training in CAS doctrine and the overall CAS process was sufficient and useful even in the asymmetric, non-linear arena.

## **2. REQUIREMENTS DEFINITION AND PRELIMINARY DESIGN**

Squadron leadership made a simple but far-reaching request following the debriefing sessions: "Find a way to capture the experiences we learned in theater so we can prepare better for the next time." In addition to this top-level challenge, the squadron was charged with developing a way to improve training focus specifically for pre-deployment spin-up and to include training on IRC and the associated applications. The challenge was issued with one caveat: ensure that the resultant approach was flexible enough to allow for a variety of warfare models, not just reactionary effort to the Afghanistan experience.

Squadron personnel looked at a variety of responses to these challenges. Briefings, computer based training and specific (stand-alone) applications training were considered and

assessed to be a valid way to prepare for a combat deployment. However, the operations section began to look at live constructive simulation and quickly assessed them as the most effective method to create the desired training environment. Through simulation, an accurate, holistic, immersive training environment could be created, presenting a realistic series of decision points to be resolved by the ASOC crews. Squadron personnel anticipated that this simulation environment, if properly constructed, would be broadly capable in terms of the warfare models used and would be extensible and flexible in terms of C2 tools that may be used in the future.

### 2.1 Preliminary Design

The preliminary simulation design was outlined in a PowerPoint briefing and endorsed by squadron leadership. The initial design included a simulation engine (yet to be defined), mIRC (a shareware IRC program), and Excel spreadsheets for Air Tasking Order (ATO) breakouts.

A preliminary design review presented a variety of simulation frameworks that could be applied, ranging from commercial-off-the shelf (COTS) simulations to research and development efforts that involved intelligent agents. Due to squadron cost and schedule constraints the decision was made to pursue a COTS solution.

Based on this framework, the next decision was the choice of a suitable simulation engine. Because this was an unfunded Air Force effort being undertaken at the squadron level, there were three characteristics to be optimized: cost, usability, and classification level. The ideal simulation engine would be cheap to purchase, would require no ongoing contractor support, could be operated by an average squadron operator, would be unclassified and, as a result, easy to field, maintain, and store.

### 2.2 Screen Available Simulations

Available simulations were down-selected to three potential products. These three simulation engines were functionally tested by squadron staff. The three simulations were:

**Decisive Action**, a simulation by Jim Lunsford, published by HPL. A Division and Corps level simulation that depicts combat with maneuver brigades and battalions along with supporting artillery and air strikes. This game was used by US Army Command and General Staff College as a training tool for officers.

**Joint Conflict and Tactical Simulation (JCATS)**, developed by the Lawrence Livermore National Laboratory. An interactive simulation tool sponsored by U.S. Joint Forces Command and managed from the command's Joint War fighting Center. The military uses JCATS for training, analysis, and mission planning and rehearsal.

**Battle Command** and its follow-on USAF version, **Quickstrike**, were developed by MÄK Technologies, now VT MÄK. Battle Command was designed with a game-like feel to support Army battalion and brigade commands and their staff officers in preparing and executing operation orders. The simulation is a military tactical trainer that allows commanders and their staff officers to practice planning and execution skills in a simulated environment.

## 3. ASOC BATTLE SIMULATION DESCRIPTION

Following functional evaluation of the alternatives, Battle Command was chosen as the best initial fit for the simulation engine. The system was already owned by the Air Force and so was freely available. It was unclassified, and it was very easily operated with little instruction required. With a minimum of familiarization training, squadron staff could "drive" the simulation as well as design and create new scenarios. Based on discussions with the contractor, the 111th anticipated swapping out Battle Command with a more air-orientated follow-on (QuickStrike), once that product had finished development. QuickStrike was based on Battle Command but specifically designed for the US Air Force. It accurately portrays a variety of tactical situations, stimulating staff interactions and the rapid decision making needed for successfully conducting mission-essential tasks. QuickStrike supports individual student training and more dynamic multi-player team training events. It is HLA compliant and has the capability to integrate with larger distributed exercises.



**Figure 1**, Combined USAF and UK ASOC Training using the ASOC Battle Simulation

Between late 2004 and late 2006 the 111<sup>th</sup> ASOC assembled and embellished the ASOC Battle Simulation. Radios, Joint Automated Deep Operations Coordination System (JADOCS), Tactical Air Control Party Close Air Support System (TACP CASS) were added to provide more fidelity to real operations. Staff also experimented with comprehensive training programs and developed scenarios that provided operators with near-real-world experience, using unclassified real-world events as simulation events.

A variety of training frameworks were tried, ranging from zero preparation 15-20 minute fights to force-on-force scripted events lasting 8 hours. A variety of asymmetric, non-linear, symmetric and linear scenarios were developed and used. In addition, training objectives and debriefing templates were developed to help focus training events and capture individual training progress. Finally the unit experimented by using the simulation in conjunction with simulated JTAC training events in the field. Although there was no direct connectivity between the simulations, the scenarios were synchronized and information was passed via radio and data link from one agency to the other.

A significant benefit of the simulation was that it enabled the unit to train autonomously. Previously, the best way to train was as a supporting unit in large Army exercises. While these opportunities are still vital in terms of interaction and reciprocal learning and training with Army JOC staff, they were now no longer the only avenue to effective training. In addition, the ASOC Battle Simulation offers flexibility in terms of time, location, and pace of learning for ASOC staff.

In 2007, the ASOC Battle Simulation was adopted US Air Force wide. In 2008, the Royal Air Force (RAF) adopted the simulation to train its Air Operations Co-Ordination Centre (Land) AOCC(L) unit (ASOC equivalent organization).

#### **4. CURRENT ASOC BATTLE SIMULATION CONFIGURATION**

The ASOC Battle Simulation comes to the unit as a complete package of hardware and software, preconfigured and ready to operate. Network setup assistance as well as training in system operation, scenario development, and exercise development and conduct is also provided. The system is comprised of networked workstations and peripherals for the FDO, FDT, and INTEL (intelligence) seats in the training audience and additional workstations for the "Control Group"; those role-playing high control, low control, and lateral organizations. It is easily reconfigured for multiple versions of training audience command and control variations based on the mission. Many of the squadrons with the ASOC mission split up the FDO and FDT tasks due to increased mission complexity. Many now have a Joint Air Request Net (JARN) operator, an Airspace Manager (ASM), an ATO Manager (ATOM), and a Procedural Controller (PC) setup. A VOIP phone system is also provided to emulate telephone communications between the ASOC and outside organizations. The collaborative workspace MiRC chat is provided as well. The ASOCs provide their own radio communications equipment to enhance the realism of the event. A projector and high reflection screen for the display of the Common Operating Picture (COP) are also provided. All of the workstations are laptops making the entire system easily deployable. Currently, the COP is provided by Command and Control Personal Computer (C2PC) by way of an interface with QuickStrike. The investigation of integrating QuickStrike with JADOCS is currently underway. This setup was easily modified to meet the specific training needs of ASOC organizations with theater-specific requirements that drove unique configurations of workstations. As each ASOC has their own training philosophies and facility limitations, the system is flexible enough to be set up in a single room or in multiple locations, separating the training audience from the Control Group.

The "Control Group" runs the exercise and is organized and managed by a chief facilitator, or "Pit Boss". This facilitator is the individual

responsible for organizing control group personnel, designing and running the exercise, and facilitating the after-action review. Due to the intuitive game-like operation and interface of QuickStrike, a single operator can be trained in a matter of hours to manipulate every aspect of the simulation. Scenarios are easily developed from scratch or existing scenarios provided as part of the system can be modified to adjust to changes in procedures, processes, or theater operations. QuickStrike has an event editor built in to develop and manage the input of training injects into the training audience. This gives the Control Group the capability to develop a Master Scenario Events List (MSEL) imbedded in the scenario to provide timed or dynamic inputs at crucial decision-making times in the exercise to reach desired training objectives. As an integral part of the ASOC team, intelligence is crucial to overall situational awareness. To assist, the intelligence workstation has the capability to emulate the Predator or any other UAV view using the VT MAK Stealth 3D viewer, which is also provided in the training suite. The latest version of QuickStrike is Standard Desktop Configuration compliant allowing the software on any Air Force computer and is undergoing the formal USAF accreditation process.

As mentioned briefly earlier, VT MAK provides training on manipulating QuickStrike as well as training for Control Group personnel. Using a tried and true training approach, the Control Group quickly develops the skills required to conduct a training event and develop or modify scenarios. This capability enables each ASOC to design and run comprehensive training events in just a matter of days. Each ASOC is provided "turnkey" scenarios and terrain for both the Iraq and Afghanistan operations. VT MAK collects feedback from the ASOCs and has released several new versions addressing inputs from the field. Requested enhancements include developing a method to populate the QuickStrike air mission data base with all the air mission data from a USMTF TBMCS ATO and displaying all the graphics from the supporting Airspace Control Order, intended to greatly streamline the scenario development task and enable complete mission integration with other exercises such as Virtual Flag.

## **5. FOLLOW-ON DEPLOYMENT AND SIMULATION VALIDATION**

The 111<sup>th</sup> ASOC deployed again in 2007. This deployment was in support of the International Security Assistance Force (ISAF) at Headquarters level. The ASOC was embedded in the Joint Operations Center (JOC), HQ ISAF, Kabul, Afghanistan. The initial advance team communicated operational conditions and operations tempo in theater to the home squadron. Training managers at the squadron easily updated QuickStrike scenarios to reflect current conditions, enabling realistic ASOC Tactics, Techniques and Procedures (TTPs) to be practiced. Operations personnel trained regularly using the simulation and were deemed combat deployable based on their simulation check-ride performance.

Based on this approach to manning the ASOC ISAF JOC, spin-up time in theater was minimal. New arrivals were already familiar with the toolsets, the operations tempo, and the processes needed to perform their duties quickly and efficiently. This resulted in minimum handover overlap and minimum unit changeover impact to the theater C2 network. In addition, as a whole, the response to Troops-in-Contact (TIC) situations was more standardized. In comparison to the unit's 2004 experience, the on-the-job learning curve was significantly reduced and the level of preparedness of the unit was vastly improved. This resulted in maximizing the impact of the ASOC in down-range operations.

One short-coming was the overly simplistic modeling of the complex NATO C2 structure in theater. Although the correct C2 nodes were used in our ASOC Battle Simulation, our interpretation of the C2 structure was based on a US doctrinal interpretation rather than the reality of the NATO C2 structure in Afghanistan. There are significant differences and understanding the finer elements of the NATO C2 structure and becoming effective coordinating between them took some on-the-job training in theater. In hindsight, the unit was constrained by role-playing in a stand-alone simulation rather than having the benefit of participating in an interactive exercise with other NATO agencies. A more accurate representation of the real-world C2 structure (either via distributed mission operations or role playing) would have been extremely beneficial to the FDOs and FDTs during spin-up.

## 6. CONCLUSION

Overall, the ASOC Battle Simulation has opened a new era in ASOC training. Its primary benefits are its flexibility, extensibility and game-like simplicity. The 111<sup>th</sup> has learned valuable lessons overseas that can be easily recreated with the simulation for training purposes. The future success of the simulation depends on two factors: how well the simulation is integrated into a broader simulation training environment, and how effectively unit training staffs employ the system. The 111<sup>th</sup> has identified five attainable objectives that will help guide the progression of the simulation over the coming years:

**6.1 Match real world operations** with simulation scenarios. This objective includes accurate modeling/recreation of the wider C2 infrastructure surrounding the ASOC as it exists rather than modeling an ideal doctrinal state.

**6.2 Integrate the ASOC simulation** in broader simulation-based exercises that incorporate the full range of C2 nodes. For example, linking up with the Air Operations Center (AOC), JTACS, various Army echelons, Control and Reporting Centers (CRCs), AWACS, fighter/bomber aircraft, and Wing Operations Centers (WOCs). With QuickStrike's HLA framework, this distributed simulation operation is technically attainable and will directly support Objective 1 by forcing staff integration of multiple agencies rather than relying solely on role-playing.

**6.3 Use the ASOC Battle Simulation as a tool** to expose non-ASOC personnel, such as Army JOC staff, and the AOC staff, to the complexities of applying CAS in the ground fight.

**6.4 Incorporate both pre-planning and execution** ASOC roles in the simulation. This may require minor software augmentation in order to stimulate a simulated Army Air Support Request (ASR) structure or in a distributed simulation scenario as suggested in Objective 2, the Army could provide actual ASR inputs.

**6.5 Maintain a flexible outlook on ASOC operations**, enabling training preparation for multiple warfare models rather than focusing on a single type of fight. Simulations are inherently flexible in this regard; the responsibility in attaining this objective rests with those charged with designing the training syllabus and the simulation scenarios.

# Virtual VMASC: A 3D Game Environment

**Suchitra Manepalli, Project Scientist; Yuzhong Shen, Assistant Professor; Hector M. Garcia, Senior Project Scientist; Kaleen Lawsure, Project Scientist**

*Virginia Modeling, Analysis, and Simulation Center*  
*smanepal@odu.edu, yshen@odu.edu, hgarcia@odu.edu, klawsure@odu.edu*

**Abstract** The advantages of creating interactive 3D simulations that allow viewing, exploring, and interacting with land improvements, such as buildings, in digital form are manifold and range from allowing individuals from anywhere in the world to explore those virtual land improvements online, to training military personnel in dealing with war-time environments, and to making those land improvements available in virtual worlds such as Second Life. While we haven't fully explored the true potential of such simulations, we have identified a requirement within our organization to use simulations like those to replace our front-desk personnel and allow visitors to query, navigate, and communicate virtually with various entities within the building. We implemented the Virtual VMASC 3D simulation of the Virginia Modeling Analysis and Simulation Center (VMASC) office building to not only meet our front-desk requirement but also to evaluate the effort required in designing such a simulation and, thereby, leverage the experience we gained in future projects of this kind. This paper describes the goals we set for our implementation, the software approach taken, the modeling contribution made, and the technologies used such as XNA Game Studio, .NET framework, Autodesk software packages, and, finally, the applicability of our implementation on a variety of architectures including Xbox 360 and PC. This paper also summarizes the result of our evaluation and the lessons learned from our effort.

## 1. INTRODUCTION

Interactive 3D virtual environments present a unique scope allowing both individuals and organizations to analyze and practice methods that are otherwise difficult. Those methods may range from the military studying geographical regions represented virtually in 3D to launch or defend attacks, to realtors advertising real estates in 3D virtual environments to their clients. Further, the success of online versions of 3D environments, such as Second Life and OpenSimulator, attest to the applicability and power of the virtual environments. While the use cases are many, the process of designing and developing such environments is the same at many levels. To evaluate the effort required and to study the problems that might arise when designing such environments, we have designed and implemented "Virtual VMASC", a 3D game, with the goal to replace front-desk personnel with an Xbox console assisting the guests arriving at the VMASC facility. The implementation provides a visual interface to search and browse for various pieces of information including faculty and staff directory, navigational maps to individual offices, and presents ongoing research information. Guests, who are represented as avatars in the virtual world, are free to walk within the building, perhaps following a .map to a specific room, interact with various entities on the way, and

eventually be able to talk to individuals through video conferences from the console.

The higher-level goals of this effort are to study and evaluate the level of effort needed to model, design, and implement 3D games, and also to study the effectiveness of the various software and platforms chosen for this implementation. We believe our study provides useful information to the Modeling and Simulation community in dealing with similar efforts.

## 2. MODELING

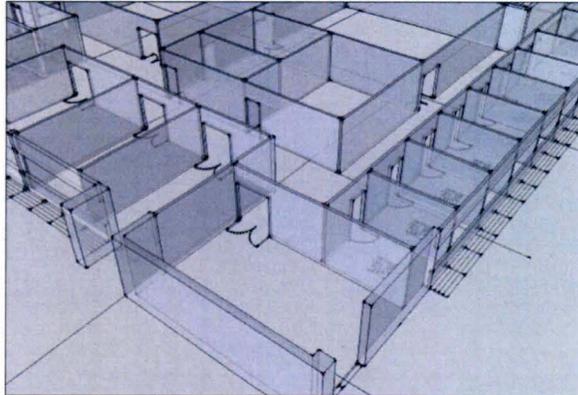
In this section, we discuss our modeling approach and the software and toolset we used for creating the 3D model for the gaming environment. We also highlight the solutions implemented to deal with the problems encountered while constructing, texturing, and prepping the model. Finally, we discuss the areas to improve in the model, which we tabled for the future.

## 3. VMASC 3D MODEL

The VMASC facility, located in Suffolk, Virginia, is a two-floor building, divided into east and west wings by a large atrium, with 120 rooms including office, lab, conference, supply, and utility spaces along with 5 restrooms and 3 kitchens. In order to construct a 3D virtual model of the facility, a variety of 3D digital content creation software technologies were used such as Autodesk 3ds MAX [2], Autodesk Maya [3], and Google

SketchUp Pro [5] as well as 2D image editing software products such as Adobe Photoshop CS4 [1] and Luxology's imageSynth [7].

The process of creating a 3D model of the VMASC building began by importing 2D AutoCAD drawings of the floor plan into 3ds Max. After importing the floor plan, walls were generated by using the extrusion tool which generates 3D extruded objects from 2D lines, in this case the 2D lines from the floor plan as illustrated in Figure 1.



**Figure 1:** 3D objects extruded from 2D floor plan

Wall height, window placements, doors, and other elevation features of the building that are required for performing extrusion were extracted from digital elevation drawings, pictures, and actual measurements of the building. Additionally, certain geometric and floor plan layout features had to be corrected due to changes in the real world building that had not been reflected in the available 2D AutoCAD drawings. In order to make those corrections, the latest paper based floor plans were used along with a physical walk through of the building. In addition to the geometric corrections, certain other features were added such as the glass for the windows enclosing the atrium, east stairs, west stairs, interior windows, benches along the hallways, a roll up door at the loading dock, and a revolving door at the front entrance.

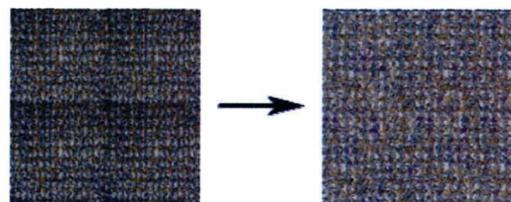
In order to make these corrections and additions, some objects were edited, while others were created from scratch. Some objects were converted into editable meshes for the purpose of modifying their geometry, and a variety of 3ds Max's tools were used for editing including tools that allow the objects to move, rotate, scale, extrude, bevel, clone, align, and attach. For creating missing features, 3ds Max was used as it allows creating simple geometric objects such as boxes and cylinders, complex objects such as knots and spindles, and architectural objects such

as doors, windows, stairs, and handrails. Those standard primitives, complex objects and architectural objects were used to add features that would replicate the VMASC building in the model.

### 3.1 Textures

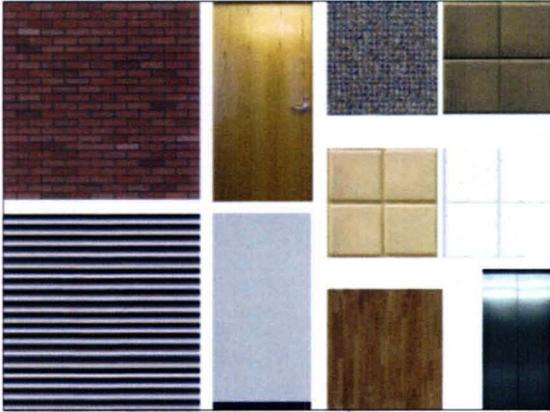
After the initial 3D building structure was created, we added textures and other materials to the model to capture the interior and exterior design of the building. Some existing textures from the 3ds Max texture library were used. However, given the high level of detail desired, it was necessary to use as many realistic textures as possible for the building's exterior, interior, and contents. In order to facilitate this, a Samsung SL310W 13.6 megapixel digital camera was used to photograph the building (in the real world). The images were processed using Adobe Photoshop CS4 and imageSynth before applying it to the model.

Photoshop CS4 offers a variety of tools for adjusting hue, saturation, contrast, brightness, exposure, and color of the images. It also provides tools for cropping, rotating, erasing, color sampling, and layering. Those image manipulation tools were used to crop and properly align the images, remove any undesirable blemishes or shadowing, and correct problems with colors or shading. However, after improving the images in Photoshop, we still faced problems tiling the textures. This problem was eliminated by imageSynth software, which creates seamless textures from the processed images, as illustrated in Figure 2.



**Figure 2:** Uncorrected carpet tile (left) and Corrected carpet tile (right)

After correcting the textures, we created a palette of materials for the entire building, some of which are illustrated in Figure 3. The UVW texture mapping tool from 3ds Max was used to place the textures correctly in the 3D model.



**Figure 3:** Sample material palette used for the VMASC 3D building

The next step in designing the model was to export it to a FBX file format, which is compatible with XNA – our gaming platform. Google SketchUp Pro was used to import the model from 3ds Max and then export to a FBX file as it provided a better FBX file that XNA is compatible with than the one generated from 3ds Max.

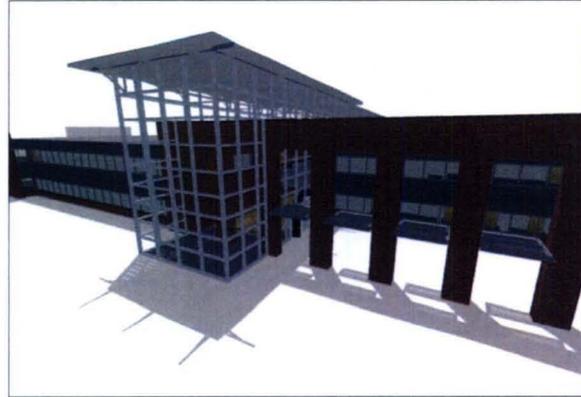
### 3.2 Hierarchical 3D Model

After designing the model, we transformed it into a hierarchical one using Autodesk Maya's hypergraph hierarchy tool. This tool allows the user to arrange the objects hierarchically one within the other, as illustrated in Figure 4 that optimizes the collision detection mechanism as described in the Technical Approach section.



**Figure 4:** Generated hierarchy using the hypergraph tool (partial hierarchy shown)

Snapshots of the final model are illustrated in Figures 5 and 6.



**Figure 5:** Front view of the VMASC 3D model



**Figure 6:** VMASC 3D building with textures

### 3.3 Model Refinement

There are still some additional details that we would like to add in the future. Those details include adding sinks, commodes, and stalls for the restrooms along with cabinetry and appliances for the kitchens.

Additionally, furniture including chairs, tables, desks, bookshelves, cabinets, and other office needs are to be added to the model. Those will have to be created using digital photographs from the real world.

## 4. IMPLEMENTATION FRAMEWORK

We have implemented the 3D game using Microsoft XNA Game Studio [8], which is a software library and toolkit targeted at independent and small game studios, academics, and hobbyists. XNA supports cross-platform game development for personal computers, Microsoft Xbox 360, and Zune media player.

At the core of XNA is the XNA Framework, a set of C# libraries for game development based on the Microsoft .NET Framework. C# is an object-oriented programming language drafted by

Microsoft and approved by ISO as a standard. The XNA Framework encapsulates low-level details involved in developing games and allows game developers to focus more on the content and high-level gaming experience. Game developers can more rapidly learn the truly important and difficult parts of game development without dealing with the low-level details such as lighting, shadow effects, etc. Developers can use both the XNA Framework and the .NET Framework in a game with the former for game-specific tasks such as graphics rendering and managing inputs and the latter for more general programming tasks.

The XNA Device Center lets developers manage and connect to multiple XNA devices, including Xbox 360s and Zune devices. The XNA Game Studio supports features including avatars, animations, and embedded videos. XNA can compress and decompress content such as meshes and textures automatically to reduce storage space usage and deployment time. XNA also supports content access from a device's media library such as songs, pictures, and playlists. The ClickOnce deployment technology can create self-updating Windows based applications that can be installed and run with minimal user interaction.

## 5. TECHNICAL APPROACH

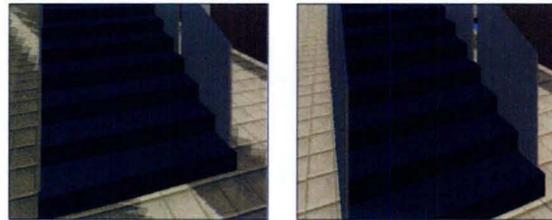
The Virtual VMASC 3D simulation, designed as an Xbox game, involved designing a 3D model, developing game heuristics, and designing an interface to meet our front desk requirements. Transforming the VMASC 3D model, the design of which is discussed in the Modeling section, into a game using XNA while still coherently presenting a real world experience to the player through the avatar resulted in many challenges including presenting a responsive and realistic 3D world, detecting and handling collisions, and providing various viewing (aka camera) modes. Additionally, optimization of the model and texture rendering and collision detection techniques proved to be pivotal for designing a responsive game. The following sub-sections highlight some of those technical challenges and the solutions we adopted for resolving them.

### 5.1 Rendering

A model is usually comprised of a composite of multiple sub-models. While different mechanisms may be used to render those sub-models, it is important that any model rendering mechanism

employed should eliminate lag, jitter, flickers, and other un-real artifacts.

The Virtual VMASC 3D model uses a variety of textures for realistically representing the VMASC building, as discussed in the Modeling section. XNA renders those textures automatically if referenced in the FBX model [8]. However, the building includes a large number of meshes and textures, the rendering of which is process intensive resulting in unrealistic lag and artifacts during model representation as part of the game. In order to mitigate this issue, a known solution based on the *octree* mechanism is implemented [6]. The crux of this solution is to recursively divide a model into eight equally sized cubes until the leaf cubes contain a specified number of spatial objects. Once a model is thus split, thereby resulting in a hierarchical graph (or tree), only those sub-cubes that are in the viewable area (based on the field of view) are processed for rendering, resulting in a cleaner, faster and responsive game. Figure 7 illustrates the visual clarity we achieved after incorporating the octree solution.



**Figure 7:** Unwanted artifacts resulting from the default XNA rendering (left) and Clean display resulting from the integration of the octree technique (right)

### 5.2 Collision Detection and Handling

XNA is in-built with content pipeline architecture for importing *art assets* from the model as binary objects (aka mesh-parts) that may, then, be processed and controlled as required by the game. The content pipeline converts art assets into binary objects using four components [4]:

- a. Importer: XNA supports and provides a number of importers. One such importer for Autodesk is the *FBX* importer, which is used in our implementation. Importers convert a model into managed objects conforming to the Content Document Object Model that is processed further by the content processor.
- b. Content Processor: Content processors process the managed objects generated

by importer and creates custom managed objects, if required for special gaming requirements.

- c. Content Compiler: The compiler bundles together the managed objects generated by the content processor into a compact binary asset for faster run-time loading.
- d. Content Loader: The loader is responsible for locating and loading the compiled asset into memory.

Related to the content pipeline architecture is the process of collision detection. XNA associates bounding spheres with meshes to deal with the location and arrangement of those meshes [6]. It is easy to see, however, that not all meshes are spherical in shape (walls, furniture blocks, doors, avatars, etc.); as such, enclosing those meshes in spheres especially to process the location and detect collisions is unrealistic, although simpler. In our model, we identified that most of the spatial features are box-like structures and enclosing those features within bounding boxes is ideal.

The FBX model we designed also associated bounding boxes with each of those meshes. To derive the bounding boxes information from the model into the gaming runtime, instead of using the default bounding spheres created by XNA, a custom content pipeline is implemented. At the time of rendering, the retrieved information is used to load bounding boxes for the meshes. Those bounding boxes are then used for detecting and handling collisions.

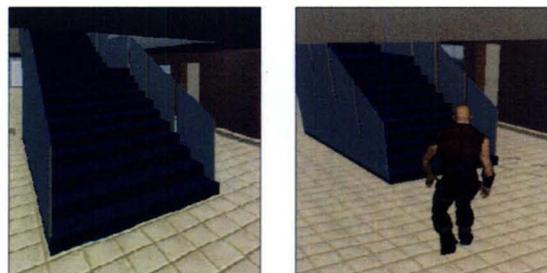
Avatars, which can move around the VMASC building model, are the reason to perform collision detection and handle those collisions. Normally, collisions occur when the avatar hits blocked surfaces like a wall or a closed door. However, handling collisions between an avatar and stairs is complicated. Instead of not allowing the avatar to proceed further on its path during such collisions, it needs to climb up or down depending on the direction of the stair mesh. This requires identifying the direction of the avatar and the orientation of the stair meshes before handling those types of collisions.

Handling collisions may be simply done by verifying if the avatar's bounding box is colliding with any of the building's bounding boxes [6]. However, performing detection in a brute force fashion by checking one bounding box after the other from the building is process intensive, and, given the number of meshes in the VMASC building, has resulted in a substantial lag after every move the avatar makes. To deal with this

issue, we redesigned the FBX model by creating a hierarchy of meshes. That is, as per this re-design, the entire building is a mesh; each of the floors in the buildings is represented as a separate mesh within that building mesh. The rooms are sub-meshes within those floor meshes, and this process was continued until every object is modeled. The advantage of this hierarchical representation is that the number of collision detections is reduced logarithmically compared to the brute-force approach. This is because, in the hierarchy of meshes (logically represented as an n-tree), the avatar may collide only with a particular path leading from the root to the leaf node, and as such all other computations are not necessary to detect the actual source of collision. Although, we designed and implemented this novel way of collision detection recently, early results seem to have corrected the lag problem dramatically.

### 5.3 Camera Modes

We implemented three camera modes giving multiple views of the model as the avatar walks through the building. A first person camera mode, in which the world is viewed through the avatar's eyes, results in a very realistic experience. We also implemented a chase camera mode in which it appears as if someone with a camera is following the avatar. An additional elastic effect where the camera slowly comes to a stop although the avatar had abruptly stopped gives a realistic experience for the viewing user. Finally, a static camera mode that just displays a constant field of view is also implemented. This camera mode is best used when multiple cameras are fixed, thereby, allowing multiple fields of view. Figure 8 illustrates the distinction between the first person and chase camera modes.



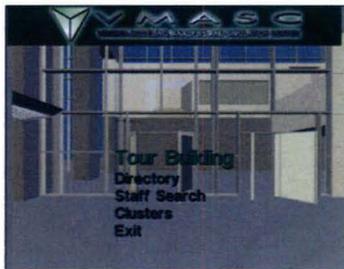
**Figure 8:** First person camera mode (left) and Chase camera mode (right)

### 5.4 Audio and Visual Interface

Since the goal of the game is to achieve virtualized front-desk features vis-à-vis providing

faculty and staff directory, ongoing research and projects, and navigational maps to individual rooms within the facility, a variety of menus and screen flows are implemented. Initially, the game is loaded with a welcome screen, as illustrated in Figure 9, which provides the following menu options:

- a. Tour of the building
- b. Personnel Directory
- c. Staff Search
- d. Cluster Information



**Figure 9:** Virtual VMASC visual interface

Users interacting with the game may tour the building as an avatar, browse through the floors by going up and down the stairs, and visit the offices. Users have the choice of various camera modes to flip through during the tour. The second option offers the users a personnel directory to browse and select any office personnel from the system to get additional information, which includes the selected person's office address, email address, and directions to his/her office. The third option lets the users query a personnel database based on a name. The fourth option provides information about the different research areas that VMASC focuses on.

The game also incorporates audio capability, implemented using the Microsoft Cross Platform Audio Creation Tool [8], which enables users to hear any recorded messages or music configured while interacting with the system.

## 6. CONCLUSION

We found that the XNA game studio coupled with the Autodesk and other software technologies we employed provides a good environment for modeling and developing interactive 3D simulation environments. Autodesk software provides a variety of tools to deal with many modeling issues, and XNA provides functionality to integrate typical gaming scenarios easily. The obvious advantage is the integration of the game developed using

XNA into Xbox 360 consoles. However, there are many challenges that we faced as discussed and those challenges required custom implementations and techniques to be employed.

We believe our study, which resulted in valuable lessons that we learned, would also benefit the Modeling and Simulation community.

## 7. REFERENCES

1. Adobe Photoshop CS4.  
<http://www.adobe.com/products/photoshop/photoshop/>
2. Autodesk 3ds Max.  
<http://usa.autodesk.com/adsk/servlet/index?siteID=123112&id=5659302>
3. Autodesk Maya.  
<http://usa.autodesk.com/adsk/servlet/index?siteID=7635018&siteID=123112>
4. Carter, Chad. "Microsoft XNA Unleashed: Graphics and Game Programming for Xbox 360 and Windows," *Sams*. August 5, 2007, pp. 113-119.
5. Google SketchUp Pro.  
<http://sketchup.google.com/>
6. Grootjans, R. (2008). "XNA 2.0 Game Programming Recipes: A Problem-Solution Approach," *Apress*. July 11, 2008, pp. 63-66, 88-110.
7. imageSynth.  
<http://www.luxology.com/whatismodo/imageSynth/>
8. Microsoft, "XNA Developer Center," 2009.  
<http://msdn.microsoft.com/en-us/xna/default.aspx>

# TEACH (Training To Enable/Achieve Culturally Sensitive Healthcare)

Russell Maulitz, M.D., Ph.D.; Thomas Santarelli; Joanne Barnieu; Larry Rosenzweig; AhN Na Yi; Wayne Zachary, Ph.D.

*CHI Systems, Inc.*

{rmaulitz, tsantarelli, jbarneau, lrosenzweig, ahnnayi, wzachary}@chisystems.com

Bonnie O'Connor, Ph.D.

*Consultant*

BOConnor@atLifespan.org

**Abstract.** Personnel from diverse ethnic and demographic backgrounds come together in both civilian and military healthcare systems, facing diagnoses that at one level are equalizers: coronary disease is coronary disease, breast cancer is breast cancer. Yet the expression of disease in individuals from different backgrounds, individual patient experience of disease as a particular illness, and interactions between patients and providers occurring in any given disease scenario, all vary enormously depending on the fortuity of the equation of "which patient happens to arrive in whose exam room." Previously, providers' absorption of lessons-learned depended on learning as an apprentice would when exposed over time to multiple populations. As a result, and because providers are often thrown into situations where communications falter through inadequate direct patient experience, diversity in medicine remains a training challenge. The questions then become: Can simulation and virtual training environments (VTEs) be deployed to short-track and standardize this sort of random-walk problem? Can we overcome the unevenness of training caused by some providers obtaining the valuable exposure to diverse populations, whereas others are left to "sink or swim"? This paper summarizes developing a computer-based VTE called TEACH (Training to Enable/Achieve Culturally Sensitive Healthcare). TEACH was developed to enhance healthcare providers' skills in delivering culturally sensitive care to African-American women with breast cancer. With an authoring system under development to ensure extensibility, TEACH allows users to role-play in clinical oncology settings with virtual characters who interact on the basis of different combinations of African American sub-cultural beliefs regarding breast cancer. The paper reports on the roll-out and evaluation of the degree to which these interactions allow providers to acquire, practice, and refine culturally appropriate communication skills and to achieve cultural and individual personalization of healthcare in their clinical practices.

## 1. INTRODUCTION

There is a mismatch between the needed grasp of cultural expectations growing out of highly diverse patient populations in 21<sup>st</sup> century America, and the sensitivity that medical providers currently bring to bear to interact with those expectations. Through no fault of their own, health-providers often emerge from a training system that gives little attention to communication skills in general, not to mention those particular "soft people skills" requisite to effective clinical management of the many diverse (ethnically and gender) subpopulations that make up American society.

Virtually every medical specialty has called for specific, measurable improvements in providers' ability to confront diversity, partly to improve the care of specific patients and partly to reduce larger socio-cultural disparities [1].

Thus, a team of physicians, cognitive scientists, and usability engineers for this project approached the problem of cultural diversity

training with the notion that there is minimal evidence to prove existing interventions provide lasting change in cultural understanding of patient needs, or that it modifies trainee attitudes or behaviors toward diverse patient-populations. With funding from the National Institutes of Health's (NIH) National Center on Minority Health and Health Disparities (NCMHHD), the project team created and performed preliminary evaluations of an extensible, malleable, case-authorable, tutoring and assessment system entitled TEACH (Training to Enable/Achieve Culturally Sensitive Healthcare), with an initial panel of interactive clinical cases that explores all aspects of the appropriate care of female African-American breast cancer patients (AABCPs). In the following content and in our conclusion, we characterize each of these desirable feature-sets, e.g., extensibility, as they have been engineered into the TEACH system.

To satisfy the "all aspects" requirement, we conducted exhaustive research on several spectrums of AABCPs' needs along several

relevant continua, including (a) the clinical spectrum from screening to advanced chemotherapy and prognosis, (b) the sociological spectrum from lower socioeconomic to higher socioeconomic personae; and, (c) the cultural spectrum from patient expectations about bodily appearance to spiritual belief-systems that may impact patient autonomy. Further below, we present several findings that address these continua, in terms of system technical features and the implications of such systems for their future expansion.

The appropriate target audience for TEACH is anyone on the training continuum, ranging from medical students to continuing medical education. In our pilot case-panel of the spectrum of AABCPs, however, we have concentrated thus far on house staff trainees—that is, interns and residents—with particular emphasis on surgical subspecialties most often responsible for coaching breast cancer patients. Our focus groups have, therefore, included predominantly surgeons, but also have representation from internal medicine and social work. Future instances of TEACH, as it is extended to other problems in cultural competency-training, will broaden this equation considerably.

As TEACH has developed, focus group members' views (those of healthcare providers and patients) have unanimously reflected the medical education literature's prevailing view: that there continues to be a mismatch between the gender and—especially—ethnic diversity of the caregivers and that of the populations they treat.

## 2. TRAINING APPROACH

Training for culturally competent communication requires not only knowledge of relevant cultural beliefs, barriers, and coping strategies, but also developing communication skills. To identify communication training objectives, we examined and synthesized literature on models of culturally competent communication and provider communication training in cancer care. A number of organizations published reports containing models or frameworks for training cultural communication (e.g., U.S. Department of Health and Human Services Office of Minority Health and Agency for Healthcare Research and Quality). We reviewed models presented in these reports and pared those down to include those likely to support communication with female AABCPs. An example of one model is Stuart and Leibermann's BATHE model [2].

Furthermore, we identified communication skills relevant to physicians who are working with patients who have chronic and sometimes

terminal conditions [3]. This literature helped us identify specific communication skills during patient consultation that lead to positive outcomes. We were then able to derive specific learning objectives that are pervasive across any relevant cultural beliefs, barriers, and coping strategies. Additionally, we conducted an extensive literature review on stories of various female AABCPs to try to elicit the patient's viewpoint on dealing specifically with a breast cancer diagnosis. Finally, through related research, we identified clinical progression stages associated with a cancer diagnosis during which the needs of the patient may change, thus leading to changes in the relevant dialogue and adjustments to required communication skills.

### 2.1 TEACH Filter Concept

Stephen Krashen wrote extensively about his hypotheses on Second Language Acquisition. He identified his fifth hypothesis as the "affective filter" or a type of blockage. The three variables he identified were motivation, self-confidence, and anxiety. He stated that people whose affective filter was high, meaning, they possessed a negative attitude toward language learning with low levels of motivation and self-confidence and high levels of anxiety, would acquire less language and achieve less than those with a low affective filter [4]. His theory, although not completely proven, has gone mostly unchallenged.

For the purpose of developing a "patient profile" for the TEACH initiative, a parallel can be made with Krashen's "affective filter". We will call this the "patient-doctor filter effect". A patient, in this case a female AABCP, is associated with several filters. The higher these filters, the more "blockage" exists. When a blockage of any proportion exists, it is more difficult for a physician to achieve a successful interaction. TEACH is aimed at instructing the physicians (not the patients) the communication skills (which are strategic in nature); therefore, it is in relation to identifying the possible filter types that have come into play and to demonstrating verbal and non-verbal actions that appropriately work within those filters, and, thereby lead to a successful interaction.

After reading and summarizing several journal articles on the topic of female AABCPs and collaborating with Dr. Bonnie O'Connor, Subject Matter Expert on African American culture and folklore, seven filters were identified, which could come into play singularly or in combination with one or more other filters. These filters drive the patient's expectations of the conversation,

her behaviors, and her decoding of messages from the physician. These TEACH filters are:

- Healthcare Coverage equates with Quality of Care
- Breast Cancer Equates with a Death Sentence
- Breast Cancer = Treatment Causing Hair Loss or Loss of Sexuality
- God Works Through Doctors
- Discrimination
- Culturally Indispensable Roles as Caregivers
- In tragedy, preferred coping strategy method is “Positive Reappraisal” (positivism and spiritualism)

## 2.2 Training Interpersonal Skills

Oncologists are facing difficult conversations due to the patient’s perspective of: fears about physical illness, psychological affects, death, treatment, friends and family, and finances, social status, and job [3]. Seemingly, when both the patient’s and the doctor’s perspectives are combined, behaviors are complex. In the case of TEACH, these perspectives are also joined with the cultural beliefs or filters that may be present.

Discrepancies and conflicts emerge by general rule across the entire filter concept with the typical behaviors, such as asking closed questions to gather information quickly, using judgmental responses, reassuring the patient before knowing her needs or concerns, forging ahead with the physician’s agenda without consideration to the patient’s agenda, and not exhibiting effective listening skills.

These above-mentioned behaviors reveal that the physician is overlooking who the patient is and where the patient is coming from. The interpersonal and communication skills learning objectives for TEACH are based on the learner exhibiting the opposing behaviors after the training is completed. In other words, the physician asks open questions in order hear the patient’s concerns.

TEACH scenarios are designed such that the physician is given several opportunities to react in different ways to a patient’s question or statement; one of those ways leads to the most successful interaction, whereas the other ways lead to either continued worry or even anger on the part of the patient. An example would be that the patient with the Positive Reappraisal Coping Strategy states, in a calm voice, that she can handle whatever the physician has told her. The physician has a choice to either comment on her strength or to ask if she really understands the gravity of her situation. The

former will lead to a successful interaction, whereas the latter may even anger the patient. This example demonstrates how TEACH combines both the filter concept and the effective physician communication skills. Additionally, when the filter concept and communication skills are joined by the clinical stages of progression, a comprehensive instructional model emerges as seen in the figure below.

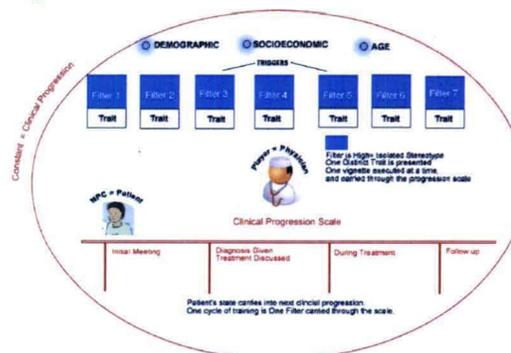


Figure 1: TEACH Instructional Model

Considering that the approach to learning objective derivation for TEACH was based on the filter concept and on identified relevant communication skills, TEACH lends itself to transfer the communication skills to a different set of filters. In other words, other minority populations and clinical contexts could also be analyzed in the same way, through literature review and subject matter expertise, in order to develop a set of filters related to that particular minority group.

## 3. TEACH VIRTUAL TRAINING ENVIRONMENT

TEACH includes a game-engine-based “player” that can execute interactive avatar-based scenarios for cultural skills training. The TEACH virtual training environment incorporates a library of cognitive-model-controlled Non Player Characters (NPCs) that facilitate the delivery of cultural-familiarization training. Through the use of a canonical cognitive model of NPC behaviors using a cognitive architecture and a generic scripting language, TEACH scenarios can be encoded and mapped to scenario-specific NPC dialog and behaviors, thereby providing a set of virtual NPCs with which the trainee can interact.

An important training requirement for TEACH was the use of avatars that possess adequate levels of visual and behavioral fidelity. This fidelity includes interactive avatars capable of a range of affect and expressivity. We utilized a pre-existing cultural training system, termed VECTOR [5], because it provided a great

degree of flexibility while requiring only incremental modifications.

Of critical importance was the issue of avatar veracity for modeling the target population of female AABCPs. The requirements for TEACH training dictated that the scenarios should include voice-acted speech coupled with avatars capable of a range of paralinguistic expressivity. Because these features were not required in the creation of the VECTOR system, this need presented a technology gap. To address this gap, we integrated a high-fidelity character-animation and lip-syncing tool, FaceFX [6], in order to provide highly interactive avatars capable of conveying subtle, non-verbal cues. Using FaceFX provides a smooth pipeline for processing voice-acted .wav files against avatar speech (i.e., dialogue) and produces character asset files which are then used to drive high-realistic game avatars.

The TEACH VTE avatars serve as a form of “virtual” standardized patient and provide for simulated physician-patient interactions with female AABCPs. The trainee is expected to maintain trust with each avatar by communicating in ways that show deference for the patient’s cultural norms (i.e., “filters”) and communication expectations. The simulated patient speaks via a voice-actor while the physician-trainee selects responses from text presented on the screen. One of the forms of performance feedback provided dynamically to the trainee is a “trust bar” based on trainee responses (in the top left), and which is an aggregate measure of patient trust. Additional measures of performance are calculated and stored into the TEACH trainee database for off-line use by an instructor or training administrator.

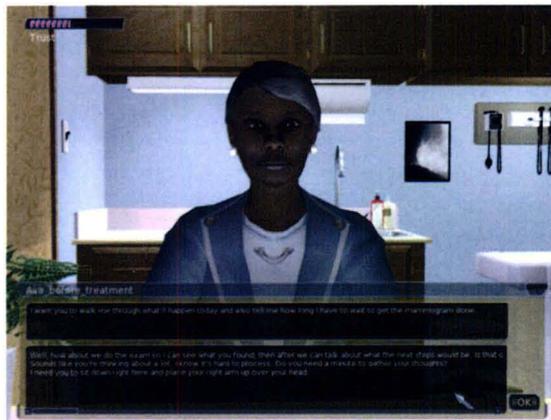


Figure 2: TEACH Patient-Trainee Interaction

### 3.1 Avatar Requirements

A range of dimensions were identified when considering the range of characteristics which would need to be accounted for in the design and development of the AABCP avatars.. Ideally, we wanted to be able to vary the avatars along this range in order to develop a group of representative prototypical avatars. Example dimensions are: visual appearances, behaviors and mannerisms, language and vernacular, and socio-demographic information.

### 3.2 Patient Interviews and Avatar Refinement

On May 6 and May 8, 2009, ten patients, who were all female AABCPs, participated in 45-minute to one hour interviews during which time the patients viewed two videos from the TEACH system, reviewed slides showing the avatars and physical setting, and participated in a lengthy discussion with the interviewer using a series of structured questions.

The length of each interview varied slightly as some patients provided more data than others, and some provided more feedback on the physical avatars. With patients who provided less information than others, the interviewer did not pressure the patient but rather let the interview move more quickly. With those patients who spoke more freely about personal feelings, the interviewer allowed the interview to last longer and extrapolated data for the study.

Overall, the patients had very positive feedback about the TEACH avatars in that their answers to the questions seemed to parallel with what we had been exploring and even, at times, had struggled with. With respect to the latter, the patients were able to shed some light on these aspects. No patient remarked that TEACH was totally unbelievable, and all patients remarked that the patients in the videos “reminded them of someone”, whether this was physical and attitudinal or whether it was an actual conversation that either they have had with a physician or someone they know has had. Finally, based on responses, it was evident that several of the seven filters, representing various cultural beliefs, were a match to the beliefs held by the interviewees.

The categories of feedback from the patients relate to themes, such as visual additions/changes to the avatars, changes in avatar gestures, expressions, or movement, and environment/setting changes. Suggested changes and enhancements to the avatars included adding jewelry, changing body features, changing clothes, changing facial

expressions, and placing the patient in a doctor's office instead of an exam room.

### 3.3 TEACH Authoring Facility

Despite the successes in applying serious-games to soft-skills training, scenario content generation is often an obstacle in developing game-based training systems, particularly in terms of cost. Furthermore, a common criticism of serious games has been the lack of a systematic approach to linking learning objectives to scenario content. This is a noteworthy discrepancy as research has been suggested that if a scenario is linked with training objectives, trainees are more likely to learn the underlying content [7].

To this end, an important design and development challenge was including an authoring capability within TEACH. Such a facility provides two advantages: 1) it allows for systematic and repeatable manipulation of existing scenarios in order to support experimentation within this virtual training environment; and, 2) it provides the ability for end-users of the system to add content in a way that positions scenario creation in the larger context of training-objective articulation, performance measurement, and feedback/assessment.

Previously, a VECTOR scenario editor component was developed to allow for the efficient creation of new game-based scenarios and to integrate instructional design principles into the authoring process to promote more effective training scenarios [8]. To address this concern, the VECTOR scenario authoring tools were augmented to support the unique requirements of TEACH, such as voice-acting and Face-FX processing.

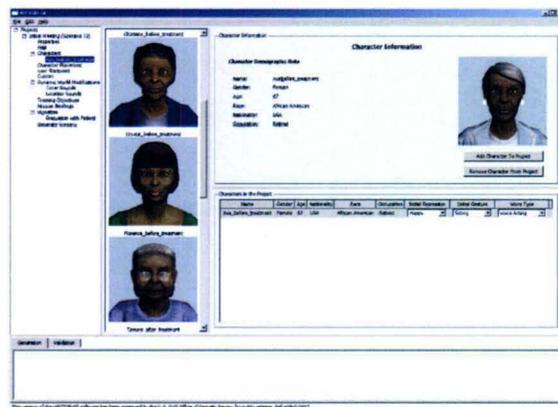


Figure 3: TEACH Scenario Editor

To facilitate consistent scenario creation, a process-model for scenario authoring is included within the TEACH scenario authoring tool. To make scenario authoring more

accessible to a wider audience (i.e., beyond "game" engineers), a cinematic metaphor was used to create the design of the authoring tool interface. The use of cinematic metaphors has been successfully used in similar VTE paradigms [9],[10]. Scenario authoring within TEACH encompasses a number of training aspects, including training objective specification, scenario information, environment specification, plot organization, vignette creation, and scenario generation.

### 4. CONCLUSIONS

Next steps for the TEACH environment will include comparative-efficacy studies of the VE as essentially a "clinical intervention" for the training health of providers. The project team will deploy varying "doses" of TEACH, alongside traditional didactic or "paper" exposition of cultural-competency norms, to trainees in randomized controlled fashion. We will also seek to define the extent to which any effect, if observed, is sustained over time, probably by exposing our sub-populations of trainees, after their differential exposure to traditional or experimental training-interventions, to the same metrics at points in time both immediately following then at an appropriate point (six months or more) remote from the interventions. Hypotheses formed as a result of the TEACH experience are two-fold. From a clinical-effectiveness point of view, we hypothesize that clinician-trainees experiencing TEACH-like VE's will demonstrate a degree of sustained cultural awareness that is dependent on intensity and recency of VE exposure. From the standpoint of the expansion of audiences for TEACH and similar VE's, we hypothesize that the following four metrics will predict success:

1. *Scalability* allows for an application created for an initial small group training framework to be enlarged to a much greater 'N' of trainees without concomitant increase in cost.
2. *Extensibility* provides for addition of new types of virtual cases within previously-designed VE's, allowing new and different forms of norms and filters to be imparted and assessed.
3. *Evaluability* allows for the direct application of comparative-effectiveness metrics to a system. Such a feature prevents the simplistic aspect of some training programs' "show-it-and-trust-it" approach to any domain knowledge.
4. *Authorability* gives tools to non-technical domain experts, such as clinicians, permitting them to populate cases without having to supply code or

otherwise contend with excessively technically-constrained requirements:

## 5. ACKNOWLEDGMENTS

The authors wish to express their gratitude to the National Center on Minority Health and Health Disparities (NCMHHD) for support under contract 1R43MS001642-02. We also thank Priscilla Grant and the extramural programs staff at the NCMHHD, Roy Gay, M.D., for help recruiting patient focus group evaluations; Wally Smith, M.D., for clinical research methodology and assistance with minority health issues; and Ari Brooks M.D., and Andres Castellanos, M.D., both for assistance with provider focus group recruitment and comparative efficacy evaluation; all of the brave women who participated in the cancer-patient interviews; Katonya Mosley for assistance with cultural appropriateness measures and technical voice recording of content; and Jim Stokes and Inez Nelson for technical assistance in content management, database design, and authoring system development.

## 6. REFERENCES

1. Betancourt, J.R.; Carrillo, J. E.; & Green, A. R. (1999). "Hypertension in multicultural and minority populations: linking communication to compliance," *Curr Hypertens Rep*, 1(6), 482-488.
2. Stuart, M.R. & Leibermann, J.R. (1993). *The fifteen minute hour: applied psychotherapy for the primary care physician*. New York: Praeger.
3. Buckman, Robert, M.D. (1992). *How to Break Bad News*. Johns Hopkins University Press: Baltimore, MD.
4. Krashen, S. (1981). *Second Language Acquisition and Second Language Learning*. Oxford: Pergamon.
5. Deaton, J.; Barba, C.; Santarelli, T.; Rosenzweig, L.; Souders, V.; McCollum, C.; Seip, J.; Knerr, B. & Singer, M. (2005). "Virtual Environment Cultural Training for Operational Readiness (VECTOR)," *The Journal of Virtual Reality*, vol. 8, pp. 156-167.
6. King, S. (2008). "Animating Speech in Games." Lecture Notes in *Computer Science*, pp. 234-245, SpringerLink.
7. Belanich, J.; Sibley, D. & Orvis, K. L. (April, 2004). "Instructional characteristics and motivational features of a PC-based game." ARI Research Report 1822. U.S. Army Research Institute for the Behavioral and Social Sciences: Alexandria, VA.
8. Barba, C.; Santarelli T.; Glenn F.A.; Bogert, D. & Belanich, J. (2006b). "An Approach to Scenario Authoring for Virtual Environment Training," *Proceedings of the 2006 Conference on Behavior Representation in Modeling and Simulation (BRIMS)*, May, 2006, Baltimore, MD.
9. Seif El-Nasr, M. (2005). "Desktop 3-D Interactive Drama – Applying Design Principles from the Performance Arts," *Human Computer Interaction International*. NV, July 21-27.
10. Cavazza, M. & Charles, F. (2002). "Character-based interactive storytelling," *ACM Joint Conference on Autonomous Agents and Multi-Agent Systems*, Intary.

# Serious Games That Improve Performance

**Clement McGowan III, Senior Principal**

*Noblis*

*mcgowan@noblis.org*

**Benjamin Pecheux, Principal**

*Noblis*

*benjamin.pecheux@noblis.org*

**Abstract.** Serious games can help people function more effectively in complex settings, facilitate their role as team members, and provide insight into their team's mission. In such games, coordination and cooperation among team members are foundational to the mission's success and provide a preview of what individuals and the team as a whole could choose to do in a real scenario. Serious games often model events requiring life-or-death choices, such as civilian rescue during chemical warfare. How the players communicate and what actions they take can determine the number of lives lost or saved. However, merely playing a game is not enough to realize its most practical value, which is in learning what actions and communication methods are closest to what the mission requires. Teams often play serious games in isolation, so when the game is complete, an analytical stage is needed to extract the strategies used and examine each strategy's success relative to the others chosen. Recognizing the importance of this next stage, Noblis has been developing Game Analysis, software that parses individual game play into meaningful units and generates a strategic analysis. Trainers create a custom game-specific grammar that reflects the objects and range of actions allowable in a particular game, which Game Analysis then uses to parse the data and generate a practical analysis. Trainers have then enough information to represent strategies in tools, such as Gantt and heat map charts. First-responder trainees in North Carolina have already partnered Hot-Zone and Game Analysis with great success.

## 1. INTRODUCTION

Multiplayer, interactive serious games can help people function more effectively in complex settings, facilitate their role as team members, and provide insight into their team's mission. In such games, coordination and cooperation among team members are foundational to the mission's success and provide a preview of what individuals and the team as a whole could choose to do in a real scenario. Serious games often model events requiring life-or-death choices, such as civilian rescue during chemical warfare. How the players communicate and what actions they take can determine the number of lives lost or saved.

However, merely playing a game is not enough to realize its most practical value, which is in learning what actions and communication methods are closest to what the mission requires. Teams often play serious games in isolation, so when the game is complete, an analytical stage is needed to extract the strategies used and examine each strategy's success relative to the others chosen.

Recognizing the importance of this next stage, Noblis has been developing Game Analysis, software that parses individual game play into meaningful units and generates a strategic

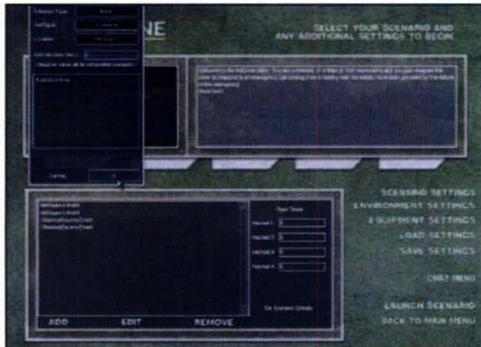
analysis. Game Analysis uses the Extensible Markup Language (XML) to import every game-play action and communication. Trainers create a custom game-specific grammar (ontology) that reflects the objects and range of actions allowable in a particular game, which Game Analysis then uses to parse the data and generate a practical analysis. Trainers have enough information to represent strategies in Gantt and heat map charts, for example.

First-responder trainees in North Carolina have already partnered Hot-Zone and Game Analysis with great success. Trainees found the game easy to play and trainers deemed Game Analysis valuable in defining what strategies worked best. Game Analysis also helped identify individuals who were particularly well-suited for certain aspects of their functional responsibilities. Trainers could see "at a glance" who was the best at communicating status or performing triage, for example. This larger view changed the training exercise from a local trainee perception to the bigger picture of the entire mission.

## 2. HOT-ZONE GAME PLAY

In Hot-Zone, players act as either hazardous materials (hazmat) technicians or the incident

commander in responding to the release of chlorine gas in a shopping mall. With only limited equipment, they must decide how they will rescue and decontaminate as many civilians as possible in the shortest time. Their specific tasks are to neutralize the gas source; set up a security perimeter and decontamination tent; triage, tag, and evacuate people; hand over evacuated people for decontamination and decontaminate themselves.



(a)



(b)

**Figure 1:** Playing Hot-Zone. During game play, the user must choose a strategy for evacuating and treating victims from a chemical gas explosion at a mall. (a) The interface lets the scenario manager tailor the overall difficulty of the game scenario by adjusting parameters such as the number of victims, chemical gas potency, and the decontamination tent efficiency. Such tailoring also allows for various skill levels. (b) The user's avatar is a hazardous materials technician evacuating a victim. Bars (left foreground) display the user's degree of exhaustion (cross) and pace (running man icon), while a smart-strip-like indicator (right) shows the degree of gas potency.

Figure 1 shows screen shots from Hot-Zone, which typically takes between 15 to 30 minutes to play, depending on the difficulty level the trainer chooses for that responder trainee. To accommodate variations in particular scenarios, trainers can adjust

parameters such as the number of participants and victims, amount of equipment, first responder's skill level; and effectiveness of the measures taken. With this customization flexibility, trainers can conduct what-if exercises, such as examining anticipated situations or ideal responses to hypothetical events.

### 3. ANALYZING HOT-ZONE DATA

Given that Hot-Zone's main goal is to improve team performance, trainers need enough information to correct team behavior, which requires more insight than just numbers of lives saved and lost. Many analysis programs and game engines allow only a high level data capture, which trainers then use as a restart point, essentially a second chance for the player to face a game situation and make different choices. Such programs rarely provide enough insight for trainers to alter team behavior.

Thus, for Game Analysis, a key goal was to enable the capture and analysis of all the data generated during game play. In this way, trainers could see the entire range of a player's actions and communications, not just select blocks. Game Analysis stores these actions and communications in a database, along with relevant metadata such as beginning and end times, event triggers, and who is sending and receiving the messages.

As Figure 2 part 1 shows, the game-specific grammar expresses the game's actions and events at an atomic level and groups these atoms into composite actions and happenings, which Game Analysis uses to discover game events. By filtering the composites, it can then obtain player, team, or environmental views and can parse an event sequence into a hierarchical task-oriented description of strategy, as in Figure 2 part 2.

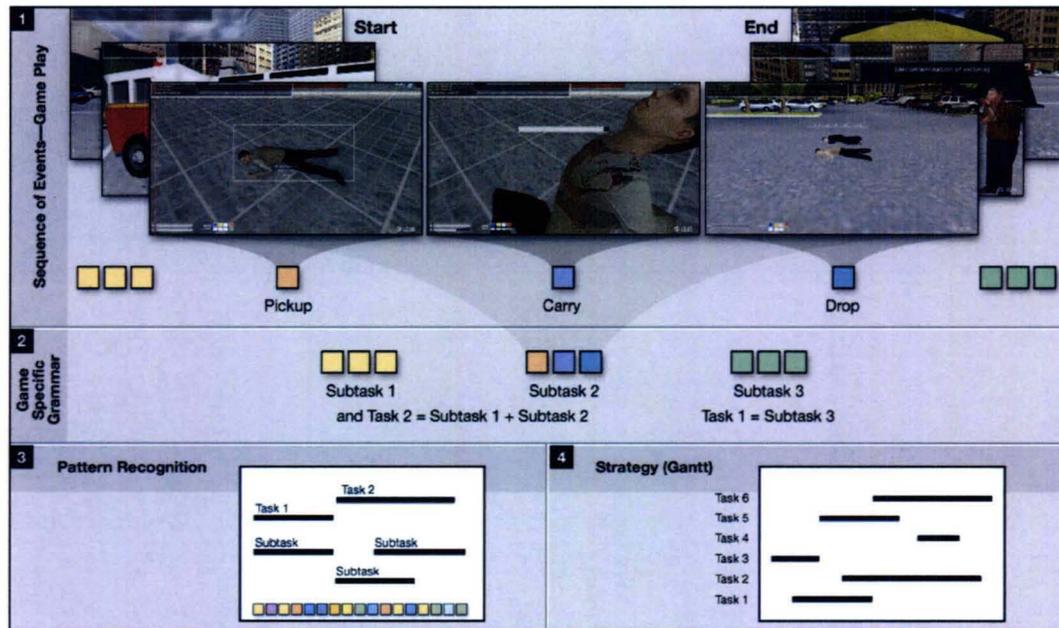
Figure 2 part 3 shows how Game Analysis can aggregate individual strategies to produce an overall team strategy. The game-specific grammar naturally expresses any task hierarchy, and because each task has a start and end time (time-stamping), Game Analysis can capture the task hierarchy and express it in analysis tools such as the Gantt chart in Figure 2 part 4.

Obviously, creating a game-specific grammar requires a detailed analysis of the particular game, which at present an expert must do, but at some point a software novice might be able to do through a user interface. The long term vision is to provide enough flexibility for trainers to use commercial off-the-shelf

(COTS) software to assemble a custom training program.

Another benefit of Game Analysis is that it captures the performance and style of play for individual trainees, which trainers can then relate to that trainee's professional background, skills, and training. Such factors have direct bearing on strategy development. Changing the skill mix and the way players communicate among themselves and with their incident commander can greatly

influence the outcome of scenarios in game play. For example, trainees frequently delayed reporting to the incident commander that decontamination showers were ready. This communication delay often meant a longer total time to complete the evacuation successfully. Such decisions are often overlooked in games that rely almost exclusively on data about victims saved or lost.



**Figure 2:** Using a game-specific grammar to parse atomic events into subtasks. (1) The row of colored squares represents atomic events that make sense in a particular game essentially all actions taken. (2) Applying the game-specific grammar, Game Analysis can sort these actions into specific subtasks. For example, in Hot-Zone, Subtask 2 could be defined as consisting of the sequence pick up victim (orange square), carry victim (purple square), and drop victim (blue square). (3) Game Analysis provides pattern recognition by breaking subtasks into their start and end times and can then generate useful analysis tools, such as (4) a Gantt chart to show the tasks performed and how long it took to perform them.

#### 4. APPLICATION RESULTS

In July 2007, Noblis tested the partnership of Hot-Zone and Game Analysis on trainees at Wake Technical Community College (WTCC) in Raleigh, North Carolina, which trains more than a thousand law enforcement and first responders yearly. The tests took place for one week at 15 fire stations in the Raleigh area, typically two play sessions daily. The first responders did not operate Game Analysis but were shown the analysis results at the end of the second game. Overall 25 different teams played the game.

Overall, first responders were enthusiastic about playing Hot-Zone. In about 15 minutes, all players in all departments—no matter how

old or young or even injured, which is common among first responders—could master game play using the keyboard and mouse controls, in part because even the older players had some experience with gaming for entertainment. Only one team member had difficulty, which he attributed to limited computing experience. Players were very professional during game play, and their professionalism intensified as the scenarios became more challenging. Figure 3 shows a typical level of concentration and interest. At the end of the game, each player saw a quick team performance summary, and a week later, many viewed the results of analyzing their games with Game Analysis (some

players had corrupt data sets that could not be analyzed).



**Figure 3:** First responder trainees from Wake Technical Community College intently playing Hot-Zone. WTCC trains more than a thousand law enforcement and first responders yearly.

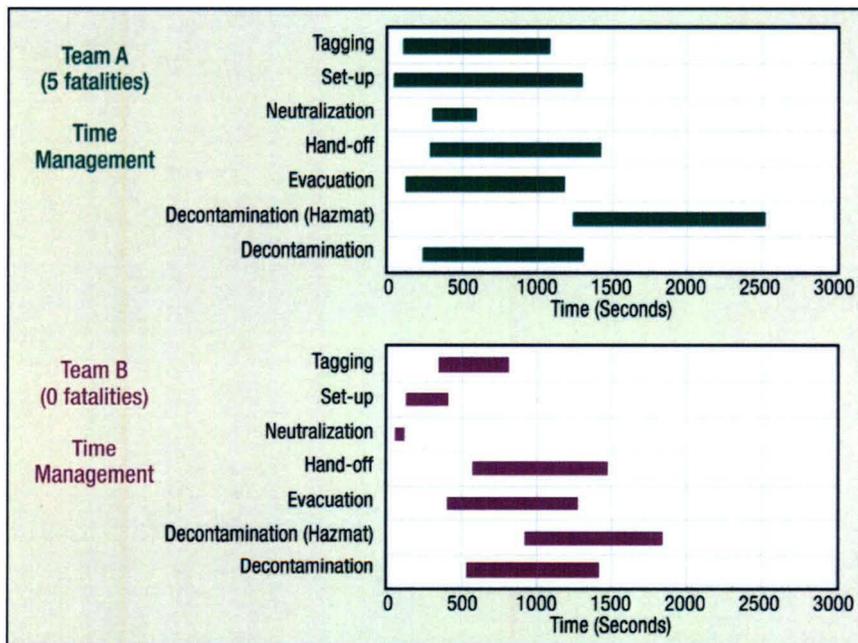
Each of the 25 teams had the same four scenarios in the same order. Beginner: introduces teams to game controls and communication mechanisms. Easy: mall scenario with only a few people and so only a few victims need rescuing. Medium: same as Easy but with a population more

representative of an actual mall. Hard: Same setting as Medium but with a secondary explosion and a heavy dose of chlorine.

Figure 4 shows the results of two WTCC teams. Both teams were experienced first responders, yet Team A had five fatalities and Team B had none. In addition, Team A took 33 percent longer than Team B to complete all the essential tasks (40 minutes vs. 30 minutes).

The charts in the figure reveal some interesting insights into why Team B's performance was superior. Team A took the seemingly natural step of starting to evacuate as soon as possible. In contrast, Team B's initial focus was on finding and neutralizing the source of the chlorine gas. Team A also deferred decontaminating team members until the end, while Team B opted to perform at least some team decontamination in parallel with decontaminating the evacuees.

These differences have important implications for strategy. The more successful strategy was to recognize and attend to the most significant obstacle to overall success—in this case, the source of the chlorine gas—rather than to rush immediately to improve a score, such as evacuating victims right away in the mistaken notion that this will save more lives.

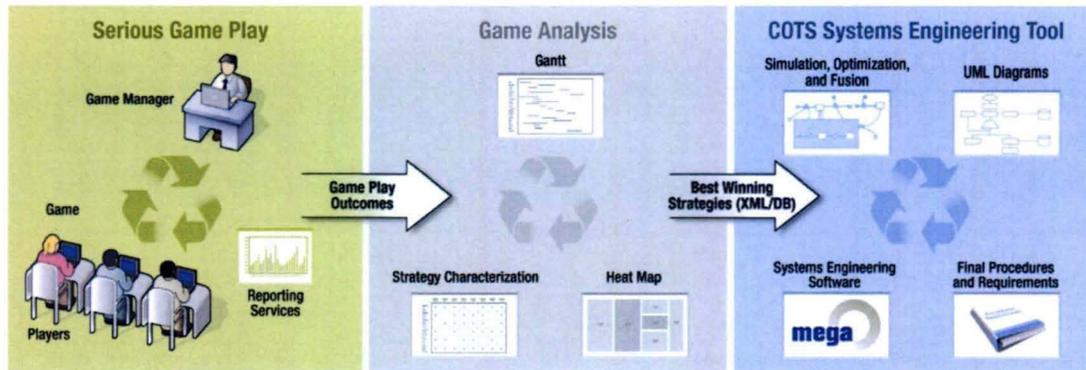


**Figure 4:** Gantt charts from Hot-Zone game play. Given the same initial situation, Teams A and B chose different strategies, leading to different outcomes. Examining game outcomes in terms of choices made leads to better performance, since the focus is on mission success, not on a particular score.

## 5. TOWARD A GAMING PLATFORM

Numerous commercially available systems engineering tools provide disciplined approaches to examining user and system requirements and operational and system performance. These systems engineering tools can combine with a serious game and Game Analysis to form an integrated framework. A framework, such as that in Figure 5, makes it possible to develop and test operational concepts and system designs. The serious game play stage of the framework is the game environment, which

consists of the game to be played, a game scenario editor that the trainer operates, and a quick reporting service that displays limited results after each game play session. The reporting service generates XML logs that can serve as input to Game Analysis. The pattern recognition algorithm in Game Analysis then analyzes the raw data and generates a strategy hierarchy, as described earlier. The strategy hierarchy becomes the basis for generating visualizations of the captured strategy and an XML representation of the strategy that is compatible with the chosen systems engineering tool.



**Figure 5:** An integrated framework for playing serious games and analyzing results. The framework consists of the game play environment, Game Analysis for characterizing strategies, and any XML-compatible system engineering tool. The best strategies identified in Game Analysis become the basis for recommended processes, which are straightforward to represent in systems engineering tools.

Noblis has already used the framework with Hot-Zone and Mega International's Mega modeling suite systems engineering tool ([www.mega.com](http://www.mega.com)). Using Mega Designer, Noblis analysts generated a swim-lane representation of how the team operated. In this representation, each first responder has his own "lane" with the time axis going from the top of the page to the bottom. Individual lanes show the actions that a particular player takes and the time they were taken. The swim-lane representation also shows the communications among players as well as the actions that several players took jointly.

When supplemented with strategy extraction and an analysis capability, serious games are an effective way to improve operational performance, providing feedback on strategies chosen and empirically finding what works best.

The use of serious games for training individuals and teams will continue to grow, as more organizations realize their value in letting trainees experiment with different

strategies and visualize their results. By working through a variety of scenarios, players begin to develop a feel for what will work and for how to act more strategically in emergency situations. Ultimately, they will begin to view a mission more systematically and adapt strategies that lead to the overall successful resolution of an emergency.

A more advanced and practical use of the game analysis framework is in the design of a dynamic and adaptive operation management tool. By collecting and analyzing hundreds of plays and scenarios from many first responder teams, it might be possible to generate an artificial-intelligence-like algorithm that could provide recommendations on resource allocations and task priority to an incident commander. The incident commander could then use the tool by plugging in the current state of an ongoing operation and receive actionable recommendations to optimize its current strategy. This application could lead to more immediate correct actions in a crisis.



# A Chess-Like Game For Teaching Engineering Students To Solve Large System Of Simultaneous Linear Equations

Duc T. Nguyen<sup>1</sup>; Ahmed Ali Mohammed<sup>2</sup>; and Subhash Kadiam<sup>3</sup>

<sup>1</sup>Professor, ODU, <sup>2</sup>Structural Engineer, <sup>3</sup>Ph.D Student, ODU  
<sup>1</sup>dnguyen@odu.edu, <sup>2</sup>ahmedali484@yahoo.com, <sup>3</sup>skadi002@odu.edu

**Abstract.** Solving large (and sparse) system of simultaneous linear equations has been (and continues to be) a major challenging problem for many real-world engineering/science applications [1-2]. For many practical/large-scale problems, the sparse, Symmetrical and Positive Definite (SPD) system of linear equations can be conveniently represented in matrix notation as  $[A] \{x\} = \{b\}$ , where the square coefficient matrix  $[A]$  and the Right-Hand-Side (RHS) vector  $\{b\}$  are known. The unknown solution vector  $\{x\}$  can be efficiently solved by the following step-by-step procedures [1-2]: Reordering phase, Matrix Factorization phase, Forward solution phase, and Backward solution phase.

In this research work, a Game-Based Learning (GBL) approach has been developed to help engineering students to understand crucial details about matrix reordering and factorization phases. A "chess-like" game has been developed and can be played by either a single player, or two players. Through this "chess-like" open-ended game, the players/learners will not only understand the key concepts involved in reordering algorithms (based on existing algorithms), but also have the opportunities to "discover new algorithms" which are better than existing algorithms. Implementing the proposed "chess-like" game for matrix reordering and factorization phases can be enhanced by FLASH [3] computer environments, where computer simulation with animated human voice, sound effects, visual/graphical/colorful displays of matrix tables, score (or monetary) awards for the best game players, etc. can all be exploited. Preliminary demonstrations of the developed GBL approach can be viewed by anyone who has access to the internet web-site [4]!

## 1. INTRODUCTION

Solving large (and sparse) system of simultaneous linear equations (SLE) has been (and continues to be) a major challenging problem for many real-world engineering/science applications [1-2]. In matrix notation, the SLE can be represented as:

$$[A] \{x\} = \{b\} \quad (1)$$

where  $[A]$  = known coefficient matrix, with dimension  $N \times N$

$\{b\}$  = known right-hand-side (RHS)  $N \times 1$  vector

$\{x\}$  = unknown  $N \times 1$  vector.

## 2. SYMMETRICAL POSITIVE DEFINITE (SPD) SLE

For many practical SLE, the coefficient matrix  $[A]$  (see Eq.1) is SPD. In this case, efficient 3-step Cholesky algorithms [1-2] can be used.

### Step 1: Matrix Factorization phase

In this step, the coefficient matrix  $[A]$  can be decomposed into

$$[A] = [U]^T [U] \quad (2)$$

where  $[U]$  is an  $N \times N$  upper triangular matrix.

The following simple example will illustrate how to find the matrix  $[U]$ .

Various terms of the factorized matrix  $[U]$  can be computed/derived as follows (see Eq. 2):

$$\begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{bmatrix} = \begin{bmatrix} u_{11} & 0 & 0 \\ u_{12} & u_{22} & 0 \\ u_{13} & u_{23} & u_{33} \end{bmatrix} \begin{bmatrix} u_{11} & u_{12} & u_{13} \\ 0 & u_{22} & u_{23} \\ 0 & 0 & u_{33} \end{bmatrix} \quad (3)$$

Multiplying 2 matrices on the right-hand-side (RHS) of Eq. (3), then equating each upper-triangular RHS terms to the corresponding ones on the upper-triangular left-hand-side (LHS), one gets the following 6 equations for the 6 unknowns in the factorized matrix  $[U]$ .

$$u_{11} = \sqrt{A_{11}}; u_{12} = \frac{A_{12}}{u_{11}}; u_{13} = \frac{A_{13}}{u_{11}} \quad (4)$$

$$u_{22} = (A_{22} - u_{12}^2)^{\frac{1}{2}}; u_{23} = \frac{A_{23} - u_{12}u_{13}}{u_{22}};$$

$$u_{33} = (A_{33} - u_{13}^2 - u_{23}^2)^{\frac{1}{2}} \quad (5)$$

In general, for a general NxN matrix, the diagonal and off-diagonal terms of the factorized matrix [U] can be computed from the following formulas:

$$u_{ii} = \left( A_{ii} - \sum_{k=1}^{i-1} (u_{ki})^2 \right)^{\frac{1}{2}} \quad (6)$$

$$u_{ij} = \frac{A_{ij} - \sum_{k=1}^{i-1} u_{ki} u_{kj}}{u_{ii}} \quad (7)$$

As a quick example, one computes:

$$u_{57} = \frac{A_{57} - u_{15}u_{17} - u_{25}u_{27} - u_{35}u_{37} - u_{45}u_{47}}{u_{55}} \quad (8)$$

Thus, for computing  $u(i=5, j=7)$ , one only needs to use the (already computed) data in columns #  $i(=5)$ , and #  $j(=7)$  of [U], respectively.

### Step 2: Forward Solution phase

Substituting Eq. (2) into Eq. (1), one gets:

$$[U]^T [U] \{x\} = \{b\} \quad (9)$$

Let's define:

$$[U] \{x\} \equiv \{y\} \quad (10)$$

Then, Eq. (9) becomes:

$$[U]^T \{y\} = \{b\} \quad (11)$$

Since  $[U]^T$  is a lower triangular matrix, Eq. (11) can be efficiently solved for the intermediate unknown vector  $\{y\}$ , according to the order

$$\begin{Bmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ y_N \end{Bmatrix}, \text{ hence the name "forward solution".}$$

As a quick example, one has:

$$\begin{bmatrix} u_{11} & 0 & 0 \\ u_{12} & u_{22} & 0 \\ u_{13} & u_{23} & u_{33} \end{bmatrix} \begin{Bmatrix} y_1 \\ y_2 \\ y_3 \end{Bmatrix} = \begin{Bmatrix} b_1 \\ b_2 \\ b_3 \end{Bmatrix} \quad (12)$$

$$u_{11}y_1 = b_1 \rightarrow y_1 = \frac{b_1}{u_{11}} \quad (13)$$

$$u_{12}y_1 + u_{22}y_2 = b_2 \rightarrow y_2 = \frac{b_2 - u_{12}y_1}{u_{22}} \quad (14)$$

Similarly

$$y_3 = \frac{b_3 - u_{13}y_1 - u_{23}y_2}{u_{33}} \quad (15)$$

In general, one has

$$y_j = \frac{b_j - \sum_{i=1}^{j-1} u_{ji}y_i}{u_{jj}} \quad (16)$$

### Step 3: Backward Solution phase

Since [U] is an upper triangular matrix, Eq. (10) can be efficiently solved for the original unknown

$$\text{vector } \{x\}, \text{ according to the order } \begin{Bmatrix} x_N \\ x_{N-1} \\ x_{N-2} \\ \cdot \\ x_1 \end{Bmatrix}, \text{ hence}$$

the name "backward solution".

As a quick example, one has:

$$\begin{bmatrix} u_{11} & u_{12} & u_{13} & u_{14} \\ 0 & u_{22} & u_{23} & u_{24} \\ 0 & 0 & u_{33} & u_{34} \\ 0 & 0 & 0 & u_{44} \end{bmatrix} \begin{Bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{Bmatrix} = \begin{Bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{Bmatrix} \quad (17)$$

$$u_{44}x_4 = y_4, \text{ hence } x_4 = \frac{y_4}{u_{44}} \quad (18)$$

$$u_{33}x_3 + u_{34}x_4 = y_3, \text{ hence } x_3 = \frac{y_3 - u_{34}x_4}{u_{33}} \quad (19)$$

$$\text{Similarly: } x_2 = \frac{y_2 - u_{23}x_3 - u_{24}x_4}{u_{22}} \quad (20)$$

$$x_1 = \frac{y_1 - u_{12}x_2 - u_{13}x_3 - u_{14}x_4}{u_{11}} \quad (21)$$

In general, one has:

$$x_j = \frac{y_j - \sum_{i=j+1}^N u_{ji}x_i}{u_{jj}} \quad (22)$$

**Remarks**

(a) Amongst the above 3-step Cholesky algorithms, factorization phase in step 1 consumes about 95% of the total SLE solution time.

(b) If the coefficient matrix [A] is symmetrical but not necessary positive definite, then the above Cholesky algorithms will not be valid. In this case, the following  $LDL^T$  algorithms can be employed:

$$[A] = [L][D][L]^T \tag{23}$$

For example,

$$\begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ L_{21} & 1 & 0 \\ L_{31} & L_{32} & 1 \end{bmatrix} \begin{bmatrix} D_{11} & 0 & 0 \\ 0 & D_{22} & 0 \\ 0 & 0 & D_{33} \end{bmatrix} \begin{bmatrix} 1 & L_{21} & L_{31} \\ 0 & 1 & L_{32} \\ 0 & 0 & 1 \end{bmatrix} \tag{24}$$

Multiplying the 3 matrices on the RHS of Eq. (24), then equating the resulting upper-triangular RHS terms of Eq. (24) to the corresponding ones on the LHS, one obtains the following formulas for the "diagonal" [D], and "lower-triangular" [L] matrices:

$$D_{ii} = A_{ii} - \sum_{k=1}^{i-1} L_{ik}^2 D_{kk} \tag{25}$$

$$L_{ij} = \left( A_{ij} - \sum_{k=1}^{j-1} L_{ik} D_{kk} L_{jk} \right) * \left( \frac{1}{D_{jj}} \right) \tag{26}$$

Thus, the  $LDL^T$  algorithms can be summarized by the following step-by-step procedures

**Step1: Factorization phase**

$$[A] = [L][D][L]^T \tag{23, repeated}$$

**Step 2: Forward solution and diagonal scaling phase**

Substituting Eq. (23) into Eq.(1), one gets:

$$[L][D][L]^T \{x\} = \{b\} \tag{27}$$

Let's define:

$$[L]^T \{x\} = \{y\} \tag{28}$$

$$[D]\{y\} = \{z\} \tag{29}$$

Then Eq. (27) becomes:

$$[L]\{z\} = \{b\} \tag{30}$$

Eq. (30) can be efficiently solved for the vector  $\{z\}$ , then Eq. (29) can be conveniently (and trivially) solved for the vector  $\{y\}$ .

**Step 3: Backward solution phase**

In this step, Eq. (28) can be efficiently solved for the original unknown vector  $\{x\}$ .

**3. RE-ORDERING ALGORITHMS FOR MINIMIZING FILL-IN TERMS [1,2].**

During the factorization phase (of Cholesky, or  $LDL^T$  algorithms), many "zero" terms in the original/given matrix [A] will become "non-zero" terms in the factored matrix [U]. These new non-zero terms are often called as "fill-in" terms (indicated by the symbol F). It is, therefore, highly desirable to minimize these fill-in terms, so that both computational time/effort and computer memory requirements can be substantially reduced. For example, the following matrix [A] and vector  $\{b\}$  are given:

$$[A] = \begin{bmatrix} 112 & 7 & 0 & 0 & 0 & 2 \\ 7 & 110 & 5 & 4 & 3 & 0 \\ 0 & 5 & 88 & 0 & 0 & 1 \\ 0 & 4 & 0 & 66 & 0 & 0 \\ 0 & 3 & 0 & 0 & 44 & 0 \\ 2 & 0 & 1 & 0 & 0 & 11 \end{bmatrix} \tag{31}$$

$$\{b\} = \begin{Bmatrix} 121 \\ 129 \\ 94 \\ 70 \\ 47 \\ 14 \end{Bmatrix} \tag{32}$$

The Cholesky factorization matrix [U], based on the original matrix [A] (see Eq. 31) and Eqs. (6-7), can be symbolically computed as:

$$[U] = \begin{bmatrix} \times & \times & 0 & 0 & 0 & \times \\ 0 & \times & \times & \times & \times & F \\ 0 & 0 & \times & F & F & \times \\ 0 & 0 & 0 & \times & F & F \\ 0 & 0 & 0 & 0 & \times & F \\ 0 & 0 & 0 & 0 & 0 & \times \end{bmatrix} \tag{33}$$

In Eq. (33), the symbols "x", and "F" represents the "non-zero" and "Fill-in" terms, respectively.

In practical applications, however, it is always a necessary step to send the original matrix [A] through re-ordering algorithms (or subroutines) [Refs 1-2] and produce the following integer mapping array

$$IPERM (\text{new equation \#}) = \{\text{old equation \#}\} \quad (34)$$

such as, for this example:

$$IPERM \begin{Bmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{Bmatrix} = \begin{Bmatrix} 6 \\ 5 \\ 4 \\ 3 \\ 2 \\ 1 \end{Bmatrix} \quad (35)$$

Using the above results (see Eq. 35), one will be able to construct the following re-arranged matrices:

$$[A^*] = \begin{bmatrix} 11 & 0 & 0 & 1 & 0 & 2 \\ 7 & 44 & 0 & 0 & 3 & 0 \\ 0 & 0 & 66 & 0 & 4 & 0 \\ 1 & 0 & 0 & 88 & 5 & 0 \\ 0 & 3 & 4 & 5 & 110 & 7 \\ 2 & 0 & 0 & 0 & 7 & 112 \end{bmatrix} \quad (36)$$

and

$$\{b^*\} = \begin{Bmatrix} 14 \\ 47 \\ 70 \\ 94 \\ 129 \\ 121 \end{Bmatrix} \quad (37)$$

Now, one would like to solve the following modified system of linear equations (SLE) for  $\{x^*\}$ ,

$$[A^*]\{x^*\} = \{b^*\} \quad (38)$$

rather than to solve the original SLE (see Eq.1). The original unknown vector  $\{x\}$  can be easily recovered from  $\{x^*\}$  and  $\{IPERM\}$ , shown in Eq. (35).

The factorized matrix  $[U^*]$  can be "symbolically" computed from  $[A^*]$  as:

$$[U^*] = \begin{bmatrix} \times & 0 & 0 & \times & 0 & \times \\ 0 & \times & 0 & 0 & \times & 0 \\ 0 & 0 & \times & 0 & \times & 0 \\ 0 & 0 & 0 & \times & \times & F \\ 0 & 0 & 0 & 0 & \times & \times \\ 0 & 0 & 0 & 0 & 0 & \times \end{bmatrix} \quad (39)$$

You can clearly see the big benefits of solving the SLE shown in Eq. (38), instead of solving the original Eq. (1), since the factorized matrix  $[U^*]$  has only 1 fill-in term (see the symbol "F" in Eq. 39), as compared to 6 fill-in-terms occurred in the factorized matrix [U] (shown in Eq. 33)!

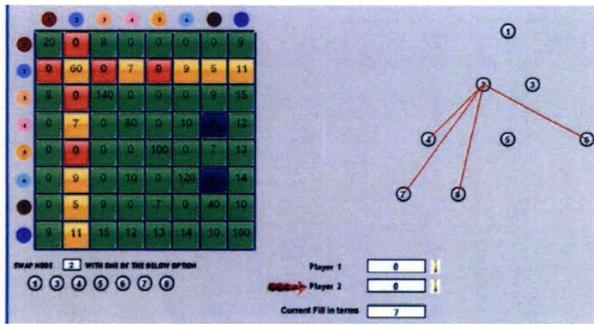
#### 4. ON-LINE CHESS-LIKE GAME FOR REORDERING/FACTORIZED PHASE [4].

Based on the discussions presented in the previous section 2 (about factorization phase), and section 3 (about reordering phase), one can easily see the similar operations between the symbolic, numerical factorization and reordering phases of sparse SLE.

In practical computer implementation for the solution of SLE, the reordering phase is usually conducted first (to produce the mapping between "old $\leftrightarrow$ new" equation numbers, as indicated in the integer array IPERM(-) in Eqs. 34-35).

Then, the sparse "symbolic" factorization phase is followed by using either Cholesky Eqs. 6-7, or the  $LDL^T$  Eqs. 25-26 (without requiring the actual/numerical values to be computed). The reason is because during the "symbolic factorization" phase, one only wishes to find the number (and the location) of non-zero "fill-in terms". This "symbolic" factorization process is necessary for allocating the "computer memory" requirement for the "numerical factorization" phase which will actually compute the exact numerical values of  $[U^*]$ , based on the same Cholesky Eqs. (6-7) (or the  $LDL^T$  Eqs. (25-26)).

In this work, a chess-like game (shown in Figure 1 [4]) has been designed with the following objectives:



**Figure 1:** A Chess-Like Game For Learning to Solve SLE.

(A) Teaching undergraduate/HS students the process how to use the reordering output

IPERM(-), see Eqs. (34-35) for converting the original/given matrix  $[A]$ , see Eq. (31), into the new/modified matrix  $[A^*]$ , see Eq. (36). This step is reflected in Figure 1, when the “Game Player” decides to swap node (or equation) “ $i$ ” (say  $i = 2$ ) with another node (or equation) “ $j$ ”, and click the “CONFIRM” icon!

Since node “ $i = 2$ ” is currently connected to nodes  $j = 4, 6, 7, 8$ ; hence swapping node  $i = 2$  with the above nodes  $j$  will “NOT” change the number/pattern of “Fill-in” terms. However, if node  $i = 2$  is swapped with node  $j = 1, 3, 5$ , then the fill-in terms pattern may change (for better or worse)!

(B) Helping undergraduate/HS students to understand the “symbolic” factorization” phase, by symbolically utilizing the Cholesky factorized Eqs. (6-7). This step is illustrated in Figure 1, for which the “game player” will see (and also hear the computer animated sound, and human voice), the non-zero terms (including fill-in terms) of the original matrix  $[A]$  to move to the new locations in the new/modified matrix  $[A^*]$ .

(C) Helping undergraduate/HS students to understand the “numerical factorization” phase, by numerically utilizing the same Cholesky factorized Eqs. (6-7).

(D) Teaching undergraduate engineering/science students and even high-school (HS) students to “understand existing reordering concepts”, or even to “discover new reordering algorithms”

## 5. FURTHER EXPLANATION ON THE DEVELOPED GAME

1. In the above Chess-Like Game, which is available on-line [4], powerful features of FLASH

computer environments [3], such as animated sound, human voice, motions, graphical colors etc... have all been incorporated and programmed into the developed game-software to be appealing to game players/learners.

2. In the developed “Chess-Like Game”, fictitious monetary (or any kind of ‘scoring system”) is rewarded (and broadcasted by computer animated human voice) to game players based on how he/she swaps the node (or equation) numbers, and consequently based on how many fill-in “F” terms occurred.

3. Based on the original/given matrix  $[A]$ , and existing re-ordering algorithms (such as the Reverse Cuthill-Mckee, or RCM algorithms [1-2]) the number of fill-in (“F”) terms can be computed (using RCM algorithms). This internally generated information will be used to judge how good the players/learners are, and/or broadcast “congratulations message” to a particular player who discovers new (swapping node) strategies which are even better than RCM algorithms!

4. Initially, the player(s) will select the matrix size (8x8, or larger is recommended), and the percentage (50%, or larger is suggested) of zero-terms (or sparsity of the matrix). Then, “START Game” icon will be clicked by the player.

5. The player will then CLICK one of the selected node “ $i$ ” (or equation) numbers appearing on the computer screen. The player will see those nodes “ $j$ ” which are connected to node “ $i$ ” (based on the given/generated matrix  $[A]$ ). The player then has to decide to swap node “ $i$ ” with one of the possible node “ $j$ ”. After confirming the player’s decision, the outcomes/results will be announced by the computer animated human voice, and the money-award will (or will NOT) be given to the players/learners, accordingly. In this software, a maximum of \$1,000,000 can be earned by the player, and the “exact dollar amount” will be INVERSELY proportional to the number of fill-in terms occurred (as a consequence of the player’s decision on how to swap node “ $i$ ” with another node “ $j$ ”).

6. The next player will continue to play, with his/her move (meaning to swap the  $i$ th node with the  $j$ th node) based on the current best non-zero terms pattern of the matrix.

**Note:**

In order to evaluate the students' performance on our developed "Chess-Like Game" for solving SLE, a detailed survey will be conducted for the "Numerical Methods" (CEE-305) course (for 3-rd year undergraduate engineering students) at the end of the Fall'2009 semester. Results of this survey should give us information about success, the level of player engagement, the average scores, etc..., and will be reported in the near future.

## **6. ACKNOWLEDGMENTS**

The authors would like to acknowledge the partial supports, provided in this work through the National Science Foundation (NSF Grant #0836916). Useful comments suggested by Jennifer McNamara (Track Chairperson) are appreciated and have been incorporated in this paper.

## **7. REFERENCES**

- [1] Duc T. Nguyen, "Finite Element Methods: Parallel-Sparse Statics and Eigen-Solutions", Springer Publisher (2006).
- [2] Duc T. Nguyen, "Parallel-vector Equation Solvers for Finite Element Engineering Applications", Kluwer Academic/Plenum Publishers (2002).
- [3] [www.brothersoft.com/downloads/flash-animation-software.html](http://www.brothersoft.com/downloads/flash-animation-software.html).
- [4] <http://www.lions.odu.edu/~amoha006/Fillinterms/FILLINTERMS.html>

# The Emergence of Agent-Based Technology as an Architectural Component of Serious Games

**Mark Phillips**

*VP Business Development, MASA Group*  
[mark.phillips@masagroup.us](mailto:mark.phillips@masagroup.us)

**Jackie Scolaro**

*Senior Modeler, MASA Group*  
[jackie.scolaro@masagroup.us](mailto:jackie.scolaro@masagroup.us)

**Daniel Scolaro**

*Senior Engineer MASA Group*  
[daniel.scolaro@masagroup.us](mailto:daniel.scolaro@masagroup.us)

**Abstract.** The evolution of games as an alternative to traditional simulations in the military context has been gathering momentum over the past five years, even though the exploration of their use in the serious sense has been ongoing since the mid-nineties. Much of the focus has been on the aesthetics of the visuals provided by the core game engine as well as the artistry provided by talented development teams to produce not only breathtaking artwork, but highly immersive game play. Consideration of game technology is now so much a part of the modeling and simulation landscape that it is becoming difficult to distinguish traditional simulation solutions from game-based approaches. But games have yet to provide the much needed interactive free play that has been the domain of semi-autonomous forces (SAF). The component-based middleware architecture that game engines provide promises a great deal in terms of options for the integration of agent solutions to support the development of non-player characters that engage the human player without the deterministic nature of scripted behaviors. However, there are a number of hard-learned lessons on the modeling and simulation side of the equation that game developers have yet to learn, such as: correlation of heterogeneous systems, scalability of both terrain and numbers of non-player entities, and the bi-directional nature of simulation to game interaction provided by Distributed Interactive Simulation (DIS) and High Level Architecture (HLA).

## 1. INTRODUCTION

The worlds of traditional simulation and serious games are converging in many ways. Each of these worlds is seeking traction in the military training market and each discipline brings its own strengths to the challenge of military training. Traditional simulation has always focused on high fidelity, academic accreditation, and validation – building systems that had lofty goals, but often unwieldy solutions. These ambitious solutions often attempted higher fidelity and more overall simulation capability than was necessary to provide baseline training value. By contrast, game development has been consumer driven and deadline focused with less emphasis on fidelity, demonstrating good results with relatively short development cycles. Despite the game industry's success, there is still lingering doubt when it comes to connecting lower-fidelity game play to formal training objectives. The US Department of Defense (DOD) is keen to find ways to merge the approaches to provide the best training value. We feel the hybrid approach should focus on "rightsizing" the amount of simulation and fidelity for each application. Additionally we believe the future of military training must incorporate Non-Player Characters (NPCs) to derive maximum training value and satisfaction from participants. NPC development has long been viewed as "too hard" – but by applying the same "task worthy" approach, we should be able to provide the right amount of character automation for each application.

## 2. GAMES AS SIMULATIONS

### 2.1 Are Games and Simulations the Same Thing?

Key distinctions between simulations and military games can be made in terms of origin and expectations. The need for modern constructive military simulations grew from the need to replace live field exercises and tabletop war games with a solution that was both realistic and cost-effective. As these exercises moved to electronic environments, trainers envisioned multi-part systems that could not only implement a war-game, but also to allow analysis, trainee performance measurement, and After Action Review (AAR) capabilities. Validation and Verification (V&V) of these simulations was considered a vitally important and often problematic part of the system development. Initial simulations required very powerful mainframes or mini-computers. As microcomputers evolved and became ubiquitous and cheap, simulation customers saw the advantages of moving to that platform.

The growth of micro computing was also a huge economic boon to the gaming industry – everyone had a PC and many found free time to play. Games based on military or near-military themes were naturally engaging – they involved danger, weapons use, and strategy that appealed to the prime early demographic - young men. High-fidelity representation of true military tactics was not

necessary, and although some games were fairly realistic, there was never any formal V&V process. To satisfy the requirement that a game must be entertaining and immersive, much of the development effort went into creating realistic 3D worlds with animated characters and vehicles whose appearance could be tailored to game-player preferences.

In the mid 1990s, the DOD embarked on an ambitious program to create a joint simulation environment called Joint Simulation System (JSIMS) (FY01 Annual Report Joint Simulation System, 2001). JSIMS boasted a complex, distributed architecture (The JSIMS Program and Architecture, 1997) aimed at providing a single simulation solution for all the services (Tiron, 2003). Unfortunately, delays, cost overruns, internal disagreements, and system complexity spelled doom for JSIMS and the program was cancelled. An Analysis of Alternatives (AoA), conducted by the DOD detailed goals and methods for shaping the future of joint training (Gardner & Hartman, 2004). The AoA study spurred the adoption of technologies that were not traditionally considered part of the modeling and simulation community. Specifically, the AoA recommended a hybrid approach to reinvent the way training systems are developed for the DOD that included incorporating gaming solutions and innovative acquisition techniques. Already worldwide simulation customers had become increasingly interested in using game approaches. They recognized early on that the great size of the game industry, predicted to approach \$50 billion by 2011 (Szalai, 2007), was effectively driving innovation.

As the traditional simulation industry begins to adopt some game industry practices, it is important to note that, while all games may be simulations, not all simulations are games. Games may not attempt to realistically represent the emulated environment or tactics – but they could still be considered “low-fidelity simulations.” Simulations, however, are usually not developed with the specific goal of reality escape or performance reward (e.g. moving to a new level, scoring points) as games are. It is possible that reward systems might be useful for training. In coming together to produce the next generation of simulations, we must view requirements with a fresh perspective and derive a hybrid product that best fits the training needs of the DOD rather than preconceived notions of what constitutes simulation versus games. Forward thinking companies have begun to merge with other companies to satisfy the need, e.g. Kynogon and Autodesk (Autodesk Acquires Kynogon SA, 2008) and BioTech and Presagis (Engenuity Technologies Acquires BGT BioGraphic Technologies, 2005), while other companies independently develop technologies that span both industries (CityScape 1.7: Real Cities Real Fast, 2009).

## 2.2 How much simulation do you need?

Simulations, both traditional and game-based, vary greatly in the fidelity of their representation of entities and the environment in which they exist. This is usually related to the original intent of the simulation.

A simulation developed for large scale military exercises will usually support large terrain areas at a low to mid level of fidelity (10+ meter) whereas one built based on a “first person shooter” (FPS) will usually have a relatively small terrain area at very high resolution (sub 1 meter) and include building interiors.

Other areas where fidelity or implementation differs among simulations include:

- Aggregation level - are simulation objects entities or units (e.g. a soldier versus a platoon)?
- Kinematic, sensor and combat models - are they fully physics-based including flout of projectiles and weather effects on sensors or are they effects-based with a “die roll” followed by a lookup in a probability of hit/probability of kill (p-h/p-k) or detection tables?
- Level of 3D modeling and animation of entities – are individual entities observable in a 3D environment or is a symbolic representation in 2D enough?
- Number of units - are large groups large groups such as populations modeled statistically or explicitly as individual entities?
- Automation of simulation objects - are high-level orders available or do objects require significant low-level management of behaviors?

The primary challenge is selecting the right simulation components for the task at hand. This is complicated by the difficulty of managing user perceptions of their simulation needs. For example, a user that has seen a high fidelity flight simulator may be resistant to using a low to mid fidelity flight model, even though the only way to observe the modeled aircraft in the target system is as a symbol on a C2 display.

## 2.3 Interoperability

The simulation community has long understood the need to develop standards which promote interoperability between simulation components. These standards range from communications protocols such as DIS, HLA and TENA<sup>1</sup> to file and message protocols (e.g. MSDL<sup>2</sup>, C-BML<sup>3</sup>) and beyond.

While no formal standards have been developed, the game community has converged on pseudo-standards owing to the nature of the components they have needed to develop. Comparable components developed by different companies have evolved similar APIs because they perform analogous actions and they are designed to be used within the same game development frameworks.

There are two main areas of interoperability that simulations have had to deal with that many game systems have not: terrain correlation and sharing of simulation objects. It is crucial that all participants in a distributed simulation provide position data that is correlated to an agreed-upon datum. For many

<sup>1</sup> Test and Training Enabling Architecture

<sup>2</sup> Military Scenario Definition Language – SISO-STD-007-2008

<sup>3</sup> Coalition - Battle Management Language – SISO (under development)

game-based systems the notion of exporting data to another simulation (game-based or traditional) is completely foreign. Even games which are meant to be used by large numbers of players in different locations are based on a homogeneous set of servers and clients with known capabilities. Adapting a game-based simulation to accept externally controlled simulation objects can present significant challenges (Scolaro, McNamara & Little, 2008).

#### **2.4 Success of Traditional SAF Approaches to Agent Development**

Traditional SAF and Computer Generated Forces (CGF) approaches to Human Behavior Representation (HBR) have had some success in supporting the needs of DOD customers. Pew and Mavor (1998) summarized 18 months of study by the Panel on Modeling Human Behavior and Command Decision Making. The panel focused on realism in HBR based on psychological, organizational, and sociological theory and generated recommendations based on what they observed. They focused on the most ambitious, high-fidelity HBRs, ones that attempted to represent full human cognition, and were used to develop agents tailored to very specific military areas. Citing the extreme difficulty of the tasks of HBR and cognitive modeling the panel recommended the collection and dissemination of human performance data to aid in model development, creation of accreditation procedures for agents, demonstrated agent validation, and accepted analysis capabilities. The study documented the wide variety of approaches used to create CGFs and the lack of coordination and consensus in the modeling community. Also, the modeling and simulation community may have become too insular in its approach to software development. Many SAF solutions have been developed as Government Off-The-Shelf (GOTS) products or demonstrations with no real follow on market. They have been very costly to produce and have only a small customer base. Traditional HBR solutions may have also been too ambitious, trying to represent all aspects of human behavior, and cognition in particular, in all products.

Some companies have achieved wider commercial success by selling partial HBR solutions to a greater number of customers. For example pathfinders, applications that calculate a travel route through a navigation mesh, are used in many commercial games (Yap, 2002). Many companies including Presagis, Kynogon, and Xaitment now have path finding products. These path finding systems were not aimed at emulating higher level cognitive behavior; rather they were developed to solve game design and play problems by allowing artists and software developers to work together to improve the speed to market of a title by creating compelling interactive scenes for the consumer market. Game-based pathfinders had the additional burden of needing to support a wide range of game engines which, in turn, supported a production pipeline of commercial tools from a variety of manufacturers. The result was development of efficient, effective, adaptable middleware. Traditional HBR solutions had never achieved that level of modularity but

customers were beginning to demand it. This shift in thinking opened the door for other game based middleware approaches in the areas of physics, scene-graphs, artificial intelligence and interface design.

One of the earliest examples of the use of a path engine to solve military research problems was the development of the Crowd Federate at the Virginia Modeling and Simulation Center (VMASC) in 2003 (Flanagan, 2008). Since then, there have been numerous examples of the use of path engines to animate complex scenes where many NPCs would be needed to represent a population. One example is the army's use of AI-Implant to automate large crowds at the Institute for Creative Technologies (Lawlor, 2007). The use of commercial based path planning engines has mostly been limited to automation of large numbers of characters in scenes where traditional SAF-based systems would have had difficulty navigating and managing the scale of animation. However, traditional SAF approaches do have the advantage of a richer behavior representation allowing more complex interaction. It seems clear that many applications could benefit from a combination of the two approaches – or better yet a new approach borrowing the best from both.

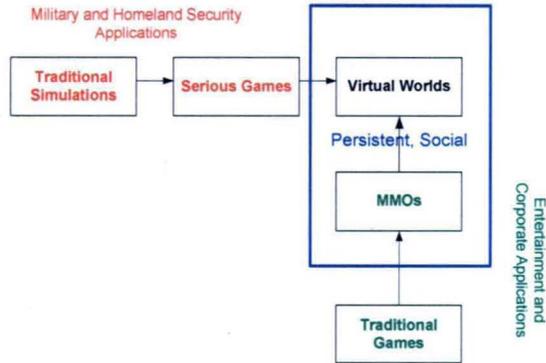
### **3. THE FUTURE: PERSISTENT, SOCIAL ENVIRONMENTS**

#### **3.1 Persistent Social Communities**

Virtual worlds are emerging as the next platform for both games and traditional simulations. The platform already boasts a market sector with revenues approaching \$2 billion a year (Economic Activity in Virtual Worlds, 2006). Beginning as extensions of Massively Multiplayer Online (MMO) games, virtual worlds have become environments that elude categorization. Both business and pleasure activities are conducted in virtual worlds, which are really seen as 3D collaborative spaces for social networking of all types. They have evolved from the traditional escapism of games to virtual extensions of the real world. Virtual Worlds Review (2006), discusses the many current applications of virtual worlds including: commercial gaming (e.g. World of Warcraft, Habbo Hotel), socializing (e.g. Second Life), education (e.g. Mokitown), political expression (AgoraXchange), and military training (Forterra Systems – On-Line Interactive Virtual Environment).

Figure 1 shows how both traditional simulations and games are converging on virtual worlds as the platform of choice. Moving to virtual worlds has many implications for the future of simulation. As the figure shows, even as virtual spaces evolve into training and analysis venues the social aspect remains. Players in virtual worlds expect social interaction no matter what the focus of the virtual space. These spaces are also persistent because they are hosted on the internet and available 24/7. The combined effects of increased social expectations and persistence must spur new technology development from both the gaming and simulation communities. For example, a classic problem with commercial virtual worlds is that, if there is not active communal

participation, use dwindles quickly. The success of these worlds relies on the active participation of human role players driving avatars to perform functions key to an objective.



**Figure 1:** Evolution from Traditional Simulation and Gaming to Virtual Worlds

### 3.2 NPCs as a Solution

The growth of virtual worlds for simulation and gaming has introduced a new requirement to provide non-player-characters (NPCs) that can move, communicate, and act in predictable and useful ways. Figure 2 shows an NPC assisting a human represented by an avatar in Second Life. The site is a Second Life location, or slurl, for the company MASAGroup (Empowering Life, 2009). The purpose of the NPCs is to show the company's products and services to visiting human-controlled avatars.



**Figure 2:** NPC Communicates with Human in Second Life

There are four key areas where NPCs could support humans in virtual worlds:

1. Presence: arrival into a virtual location can feel very similar to arriving in a real but unfamiliar location. The use of NPCs can assist in reassuring a player or visitor that they are in the right place and are supported.
2. Performance Support: NPCs also provide the equivalent of online help by having specific knowledge valuable to a player or visitor in a new location. This knowledge includes a thorough understanding of the venue and all of its features to assist in orienting the play to the space and their task.
3. Vigilance: virtual spaces focus a great deal of attention into a very narrow visual channel. NPCs can support the human by making them aware of events and ensuring that if they are not paying

attention due to fatigue or distraction that they can be brought back to task subtly.

4. Role-Playing: as serious gaming and military training use virtual spaces, on-demand NPC teammates and adversaries will grow in importance

## 4. THE IMPORTANCE OF PLAUSIBLE HUMAN-LIKE BEHAVIOR

### 4.1 HBR in the New World

While nothing matches human adaptability and flexibility, NPCs will need to supplement human role-players to enhance training in virtual worlds and beyond. For serious game applications NPCs can help to direct and focus activity to meet training objectives. They can even act as intelligent tutors providing real-time feedback and coaching.

The challenge for HBR is to find a practical middle ground between the high fidelity, high cost of many traditional HBR approaches and the relatively narrowed scope of the path finding that often constitutes game AI. This middle ground must not only provide a useful and compelling level of functionality for NPCs but also be easy to integrate within systems along the traditional simulation, game, virtual world spectrum. Additionally, HBRs should be ready to integrate with other HBRs in order to maximize the strengths of each.

### 4.2 Approaches to HBR

During the 1990s alliances between industry, government, and academia worked to create integrated cognitive architectures to be used to build CGFs in simulations such as Modular Semi-Automated Forces (ModSAF), and Joint Semi-Automated Forces (JSAF). The Agent-Based Modeling and Behavior Representation (AMBR) effort compared many of the key behavior modeling technologies (Gluck & Pew, 2005). Included in the comparison were: Elements of ACT-R, Soar, and EPIC (EASE), Distributed Cognition (DCOG), Cognition as a Network of Tasks (COGNET), and Atomic Components of Thought – Rational (ACT-R). Although the AMBR evaluation team did not specifically rank the architectures, it compared cognitive agent performance on multitasking and category learning to human performance using a common simulation test-bed. The research concluded that all of the approaches had merits and weaknesses, and that no single approach emerged as a clearly superior HBR representation, but that the state of the art was strong from a theoretical perspective.

While the architectures developed in the 1990s focus almost exclusively on cognition, the following decade saw growth in representation of behavioral factors traditionally considered to be outside cognition, such as culture and emotion. The rise of these factors acknowledges that behavior is more than a consequence of pure logic. Decision making encompasses “fuzzier” areas such as personal preference, affects, desires, and belief systems (Evertsz, Ritter, Busetta, and Bittner, 2008).

Additionally, socio-cultural factors, such as leader and follower behavior, can have a significant effect on behavior (Silverman, Bharathy, Nye, & Eidelson, 2007). As the field matures the relative influence of all factors on behavior will evolve. NPC architectures must be able to respond quickly to new advances in HBR.

AI for games has followed a somewhat different path than traditional HBR for simulations. First, the term AI has been used to represent any “intelligent” or “automatic” behavior on the part of NPCs. Game AI can be hardcoded reactions to simple stimuli, mathematical algorithms to deal with specific problems like path finding, or a more complex framework including behavior and knowledge representation (Isla & Gorniak, 2009). Game AI generally has a more pragmatic goal than traditional HBR – commercial viability. Development lifecycles must be constrained and products must be robust and reliable. Performance and attractiveness are critical as game users expect to be immersed in an entertainment experience that runs on their home computer. Fidelity of behavior representation will be sacrificed to make sure the product looks good and ships on time.

As we move toward the new world of merging traditional simulation and games we should apply lessons learned in both areas.

#### 4.3 Typical HBR Components

One thing many different HBR approaches have in common is a low-fidelity description of what AI “means.” Figure 3 depicts the generic AI representation, or “AI loop,” that shows the cycle from what is perceived in the world, to a decision engine that determines what action to take, and the resulting action performance (which in turn modifies the world and leads to new perception). The components shown are always accounted for somewhere in an HBR system whether it’s a rich cognitive architecture or a narrowly-focused movement algorithm.

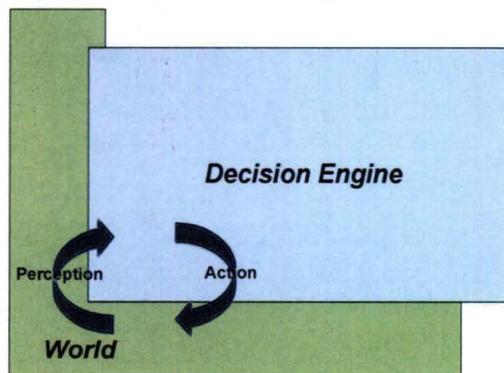


Figure 3: Generic AI Representation

In reality the contents of the “Decision Engine” box are the most interesting – and the most divergent depending on the HBR type. Figure 4 shows a decomposition of different areas that have been modeled over the past two decades. Often, an HBR will attempt to represent multiple areas, e.g. memory, cognition, and learning. Other HBRs concentrate on a single area (like path finding or action selection).

HBR types not only differ in scope and fidelity of human behavior emulation, they also differ in how they choose to represent the behavioral areas. Figure 5 shows some of the documented strategies for representation. Each representation has its advocates – from those who argue it better represents how a human brain works to those who argue it is the most efficient strategy.

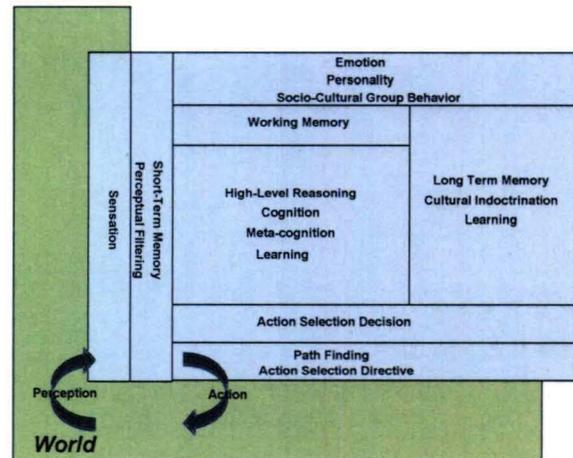


Figure 4: AI Representation Decomposition

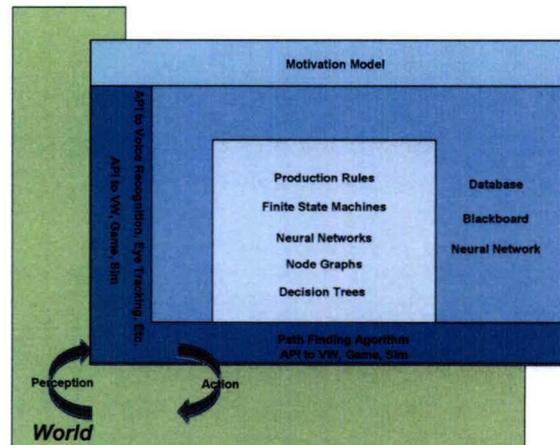


Figure 5: Heterogeneous HBR Representations

#### 4.4 Architectural Building Blocks

The state of the art in HBR could be described as fractious – with many competing strategies and definitions of the basic HBR components. Giordano (2004) describes the challenges of obtaining realistic HBR within the current state of the practice. He notes that some areas of HBR are farther along than others, including some limited conversational ability using state of the art speech recognition. Despite many advances, and many novel approaches to HBR from both traditional simulation and gaming communities, many areas are still immature, or too costly in terms of time and resources.

Rather than attempting to resolve the HBR conundrum with a one-size-fits-all solution, we should instead be looking to create an open, component-based architecture that allows different HBR players to focus on specific behavioral areas. These areas would become building blocks to a larger HBR

system. Some solutions might encompass multiple blocks. For example a neural network may encompass memory and cognition. But developers would still need to create an API to connect to the rest of the system.

Using an open architecture to deliver HBR for modern NPCs has other benefits as well. Just as developers should be providing just enough simulation to suit requirements, so should HBR providers seek to deliver just enough behavior. For some applications, or for some subset of the NPCs within an application, path-finding along with some rudimentary motivation may be enough to satisfy the goals. For example in a training application, crowds may be implemented in this way, while the key players (adversaries, allies) require a more complex brain and set of behaviors.

A valuable lesson to be learned from the game industry is that processes must be streamlined to accommodate tight development schedules. One way to accomplish this is by adopting a "just enough" approach to HBR using an open architecture. But to truly trim unnecessary cost and complexity from the process, we should also consider unifying the NPC brain authoring process. Not only would HBRs need to show they can integrate an agent into the architecture, they would also need to incorporate their authoring tools into a single development environment. Advancing a single development environment would not be a popular constraint – as it requires extra effort on the part of HBR creators. But to meet the goals set out by the JSIMS AoA and Panel on Modeling Human Behavior and Command Decision Making seemingly radical approaches may be necessary.

## 5. REFERENCES

- Autodesk Acquires Kynogon SA (2008). Retrieved from Efytimes website: <http://www.efytimes.com/efytimes/24943/news.htm>
- CityScape 1.7: Real Cities Real Fast (2009). Retrieved from pixelactive website: <http://pixelactive3d.com/Products/CityScape/>
- Economic Activity in Virtual Worlds (2006). Retrieved from TNL.net website: <http://www.tnl.net/blog/2006/07/31/economic-activity-in-virtual-worlds/>
- Empowering Life (2009). Retrieved from MASAGroup website: <http://slurl.com/secondlife/masa%20group/128/128/128/>
- Engenuity Technologies Acquires BGT BioGraphic Technologies (2005). Retrieved from Presagis website: [http://www.presagis.com/about\\_us/press\\_room/releases/engenuity\\_technologies\\_acquires\\_bgt\\_biographic\\_technologies/](http://www.presagis.com/about_us/press_room/releases/engenuity_technologies_acquires_bgt_biographic_technologies/)
- Evertsz, R., Ritter, F., Busetta, P., & Bittner, J. (2008). *CoJACK - Achieving Principled Behaviour Variation in a Moderated Cognitive Architecture*. In Proceedings of the 2008 Conference on Behavior Representation in Modeling and Simulation (BRIMS), April 2008, Providence, RI: SISO, Inc.
- Flanagan, S. (Ed). (2008). *From Concept to Leader: The Virginia Modeling, Analysis, and Simulation Center - a Ten Year Narrative*. Norfolk: Old Dominion University
- FY01 Annual Report Joint Simulation System (JSIMS) (2001). Retrieved July 30, 2009, from GlobalSecurity.org Web site: <http://www.globalsecurity.org/military/library/budget/fy2001/dot-e/army/01jsims.html>
- Gardner, D., & Hartman, F. (2004). *Transforming Joint Training: The Office of the Deputy Undersecretary of Defense for Readiness Discusses the Training Capabilities Analysis of Alternatives Study*. Retrieved from Training Transformation website: <http://www.t2net.org/downloads/briefs/news/MT2Issue9.6.pdf>
- Giordano, J., Reynolds, Jr., & P., Brogan, D. (2004). *Exploring the Constraints of Human Behavior Representation*. In Proceedings of the 2004 Winter Simulation Conference, December 2004, Washington, DC: IEEE
- Gluck, K., & Pew, R. (Eds). (2005). *Modeling Human Behavior and Integrated Cognitive Architectures*. Mahwah: Lawrence Erlbaum Associates.
- Isla, D., & Gorniak, P. (2009). *Beyond Behavior: An Introduction to Knowledge Representation*. Retrieved from Naimad Games website: <http://naimadgames.com/publications.html>
- Lawlor, M. (2007). *Military Humanizes Virtual Population*. Retrieved from AFCEA Signal Online website: [http://www.afcea.org/signal/articles/templates/Ssignal\\_Article\\_Template.asp?articleid=1259&zoid=203](http://www.afcea.org/signal/articles/templates/Ssignal_Article_Template.asp?articleid=1259&zoid=203)
- Scolaro, D., McNamara, J. and Little, J. (2008). *Beyond playing games: Federating game-based technology with large-scale simulation systems*. Proceedings of the Interservice/Industry Training, Systems & Education Conference (IITSEC) 2008. [CDROM]. Arlington, VA: National Training Systems Association.
- Silverman, B., Bharathy, G., Nye, B., & Eidelson, R., (2007). *Modeling Factions for 'Effects Based Operations': Part I - Leaders and Followers*. [Departmental Papers (ESE)] Retrieved from The University of Pennsylvania Library Repository website: [http://repository.upenn.edu/sw\\_gallery.html](http://repository.upenn.edu/sw_gallery.html)
- Szalai, G. (2007). *PwC: Global Gaming Market to Approach \$50 Billion by 2011*. Retrieved from GameDaily website: <http://www.gamedaily.com/articles/features/pwc-global-gaming-market-to-approach-50-billion-by-2011/70579/?biz=1>
- The JSIMS Program and Architecture (1997), Presented at the Fall 1997 Simulation Interoperability Workshop, Fall 1997, Orlando, FL: SISO
- Tiron, R., (2003). *Pentagon Cancels Program With Checkered Past*. Retrieved from National Defense Magazine website: [http://www.nationaldefensemagazine.org/ARC/HIVE/2003/APRIL/Pages/Pentagon\\_Cancels3904.aspx](http://www.nationaldefensemagazine.org/ARC/HIVE/2003/APRIL/Pages/Pentagon_Cancels3904.aspx)
- Pew, R. & Mavor, A., (1998). *Modeling Human and Organizational Behavior*. Washington: National Academy Press.
- What is a Virtual World? (2006). Retrieved from Virtual Worlds Review website: <http://www.virtualworldsreview.com/info/whatis.shtml>

# Pedagogy and Assessment in an Alternate Reality Game (ARG) Training Event

Shawn A. Weil<sup>1</sup>; Alexandra Geyer<sup>1</sup>; Paul Picciano<sup>1</sup>; Alice Leung<sup>2</sup>; William Ferguson<sup>2</sup>

<sup>1</sup>*Aptima, Inc.*  
12 Gill St, Suite 1400  
Woburn, MA 01801

<sup>2</sup>*BBN Technologies*  
10 Moulton St  
Cambridge, MA 02138

{sweil, ageyer, ppicciano} @aptima.com {aleung, wferguson} @bbn.com

**Abstract.** Alternate Reality Games (ARG) are emerging as a popular new interactive multi-person gaming genre. Distributed players look for clues and items embedded within real world artifacts (e.g., photos, advertising, web pages, and other media) in order to propel a game-narrative. This paper discusses a recent effort to transform the ARG genre from entertainment to a team training platform. Though novel for team training, ARGs can provide a highly compelling environment for social interaction, a critical component of team performance for Joint Task Force (JTF) personnel. ARGs are characterized by a highly engaged participant pool, a fuzzy demarcation between reality and fantasy, and a dynamic story that unfolds over days, weeks, or even months. Some of these qualities are ideal for training effective teamwork behaviors while others may be distracting or logistically difficult. To accommodate training requirements, the study organizers modified the ARG format. A web portal served to constrain event content and trainee interactions mitigating security concerns and permitting inclusion of explicit training objectives and a means to assess performance. In order to ensure the opportunity for learning within this disaster relief ARG-based exercise, developers crafted realistic tasks with components explicitly serving training targets. Three major learning objectives were deemed critical: Information Management, Organizational Navigation, and Interagency Interaction. Tasking orders emphasized these objectives differentially throughout the exercise to gauge trainees' knowledge and application of each. The assessment team used 5-point scales to evaluate completed tasks, the anchors reflecting the degree of fulfillment of the task requirements. Post-training questionnaires provided insight into the perceived effectiveness of this new instructional method. This approach permitted normalized comparisons across tasks, training objectives, and participants that shared few specifics. Results indicated that, with appropriate adaptations, the ARG genre may have promise for moderate intensity team training.

## 1. INTRODUCTION

Alternate Reality Games (ARG) have emerged as a new genre for entertainment. In an ARG, a small group of exercise managers create an interactive narrative that takes place over long stretches of time (e.g., weeks or months). The plot of the narrative is not overtly presented to the players but, instead, is intended to be incrementally discovered by encountering clues or experiences that are embedded in naturalistic media (e.g., websites, email, phone calls, newspaper, and graffiti) as if these incidents were occurring as part of the real world. Players find these clues and collaborate to piece together both the mystery and the solution to that mystery. The impression that these events are happening "in the wild" is a major attribute of these games, given the embedded nature of the narrative. Game organizers improvise activities, dynamically modify the scenario, and serve as confederates to add interest to the narrative. The game ends when the planned narrative has run its course and the participants have solved the mystery, uncovered the

truth, or explained the central puzzle of the game [3].

While ARGs have been predominantly used for entertainment or advertising purposes (but see [1]), they possess some characteristics which make them ideal for training certain types of environments. These characteristics which make them appealing for training include (1) the use of real-world communications media, such as e-mail and text chat, (2) the emphasis on information fusion and coordination among distributed participants, (3) the interactive experience shared by many participants while dynamically guided by a few controllers, and (4) the long event duration. In this paper, we employ the term *Helical Training* to describe an ARG-based approach to long duration, intermediate intensity Joint Task Force (JTF) staff training.

The Helical Training event sponsored by Joint Forces Command (JFCOM) and conducted in January and February of 2008 is the first of its kind in the realm of military training. Targeted towards

newly appointed members of JTFs, the Helical Training approach is intended to support interpersonal and inter-group coordination, organizational familiarity, and information management skills. Currently, individuals assigned to staffs have limited opportunity to participate in real-world exercises that would train appropriate networking skillsets. While rote, declarative training programs attempt to increase JTF knowledge, learning is often slower than desired. In contrast, a Helical exercise is intended to provide experiential training over time and leverage an immersive, action-based learning paradigm that can enhance both immediate learning and long-term retention.

While *potential* for Helical training to jumpstart JTF training had been proposed, empirical evidence is required to vet the approach pedagogically. For any training program to be deemed worthwhile, it must demonstrate improved performance among trainees in areas of focus. Careful attention must be paid towards performance achievement and, if possible, transfer of skills/knowledge from the training platform to the environment of interest. Because Helical Training is emerging from the entertainment-focused ARG genre [2], [5], there is no established methodology for performance assessment. By nature, the skills to be taught (e.g., coordination and information synthesis behaviors) do not easily lend themselves to pre/post-tests of factual information or short post-training exercises designed to demonstrate skill mastery. Instead, in the current formative evaluation of Helical Training, alternative methods for assessing performance were designed examining trends over the course of the training itself.

As part of the Winter 2008 Helical Training event dubbed *Dark Waters*, the authors were asked to create a formative assessment scheme. Using a combination of real-time observation and pre-/post-event surveys, we measured trainee performance and opinion about the experience. The *Dark Waters* scenario, event protocols, and system development were depicted in a previous report [4]. This paper describes the assessment protocols used and the major results. In addition to an evaluation of trainee performance, the Helical Training Event was an opportunity to evaluate, and thus improve, methods for evaluation. As such, our meta-ideas about future events are included in the last section.

## 2. ASSESSMENT APPROACH

Our assessment of the Helical Training Exercise will be discussed in two parts. First, observations collected during the event will be discussed. We refer to these as "In-Event Annotations." Observers

logged-in to the game environment (dubbed the *Looking Glass Interface* [LGI]) through their own accounts, and followed participant actions by reviewing posts, emails, and other actions recorded in the event. The collection of participant actions and submitted responses for a given task were then used to rate each participant according to three Major Learning Objectives described below.

A second method of assessment utilized pre- and post-event surveys. These permitted the capture of demographic information and knowledge prior to the event as well as participants' thoughts and feedback following the training. These surveys provided worthwhile insight about our participant population and contributed to lessons learned in conducting such a training exercise.

### 2.1. Participants and Observers

More than 100 individuals were registered to participate in *Dark Waters*. Many of these participants were nominally involved. Despite the large number of registrants, many did not engage sufficiently to be evaluated. At the conclusion of *Dark Waters* we found reliable interactions worthy of assessment from approximately 25% of participants.

Participant backgrounds were heterogeneous. There were two primary participant pools. The first was made up of college Reserve Officer Training Corps (ROTC) students. The second was made of employees of volunteers from the military training community.

There were two in-game observers. Both have backgrounds in experimental design, survey development, and assessment. While neither observer had first-hand experience with ARGs, both were involved in the Helical Training pilot study, one as observer, one as participant.

### 2.2. Major Learning Objectives

There were three Major Learning Objectives (MLOs) used during the Helical Training event, chosen based on an assessment of existing JTF training materials. These MLOs were (1) *Information Management*, (2) *Organization and Organizational Navigation*, and (3) *Interagency Interaction*.

#### 2.2.1. Information Management

Information Management, the first MLO, refers to the manner in which participants identified, sorted, and acted upon information presented to them during the exercise. It included the following behaviors.

- Discriminate between high and low relevance information, given a mixture of both
- Summarize and forward relevant information in a timely manner to support the Commander's Decision Cycle
- Seek or request information from appropriate sources
- Evaluate conflicting information to ascertain the reliability of a conclusion
- Document information provenance as an indicator of information integrity

### 2.2.2. Organization and Organization Navigation

The second MLO referred to the participants' ability to "work the organization." This included their ability to understand the differential roles of the various team members and to leverage those roles and skills to the greatest advantage for the team. It required them to do the following:

- Form a short-term task group and appropriately divide responsibilities
- Compensate for an unfilled position in the task group
- Hand off responsibilities to another group within your organization, at the appropriate time, providing any necessary background information

### 2.2.3. Interagency Interaction

The final MLO pertained to interagency interaction; the manner in which the participant understood the organizational components within the JTF. Their activities for this task included the following:

- Utilize liaison officers to other agencies
- Do not exceed decision making authority in communications with outside agencies
- Take into account the capabilities of other agencies during planning
- Inform relevant other agencies of near-term plans which may impact their operations

These three MLOs comprised the bulk of the evaluation and assessment.

## 2.3. In-Event Annotations

In-Game Observers judged participant performance on each of the three MLOs on a 5-point scale (Figure 1) to allow for a range of subjective assessment.

1	2	3	4	5
Below Expectations		Meets Expectations		Above Expectations

**Figure 1.** Assessment Scale Used in Helical Training Evaluation.

We selected a generic 5-point scale to rate task performance (below, meets, or above expectations) related to the Helical Training Objectives. Standards of what constituted the meeting or exceeding of each task were prescribed so ratings could be consistent across observers.

This approach affords the benefit of meaningful aggregation across not only individuals (to elucidate team performance) but also across the learning objectives themselves. Had we assigned each learning objective unique anchors, we would be prevented from drawing any conclusions from the sum of *Information Management, Organization and Organization Navigation and Interagency Interaction* because the scores would all have different meaning. Employing a rating assessment of normalized performance supports running the desired computations and descriptive statistics.

## 3. IN-EVENT ANNOTATION RESULTS

Each trainee was asked to complete 10 tasks, each requiring that they explore the "game world" and interact with their fellow trainees. Our annotation process involved collecting all pertinent information available generated by a trainee for a particular task. In addition to the response email, we also collected interaction in meetings and forums, emails to other JTF members, and other contributions to either enhance the quality of his/her own response, or the dissemination (post to forum, provide link to information, seek information beyond group) of information to assist other group members.

Based on the annotations conducted, Figure 2 shows the mean score of all participants by task. The numbers portray a modest trend of increasing mean scores which could suggest improvement as the event progressed. Each task can also be decomposed into the MLO derived constituents. Examining Figure 2, no particular learning objective consistently produced greater or lesser performance. Note that each task placed different emphasis on each learning objective. Much of the time, the scores were quite congruous with a few exceptions. Task 4 received the lowest ratings of all tasks scoring only a "1" for interagency coordination. That task specification suggested contacting groups outside the JTF. Since the task called for interaction specifically, failure to demonstrate this in the

response resulted in a low score as it would not have met expectations.

Reaching outside agencies may have been difficult for the trainees to achieve despite the necessary emails present in the organizational chart on their LGI. The typical interactions did not involve these outside organizations. Whether or not trainees found them in the chart—or felt unsure they would get a response—may explain the lack of observable inter-agency interaction. A further constraint was the evaluation team’s ability to collect *all* the interactions a particular trainee may have experienced. The sheer volume of data made it difficult to ensure all relevant interactions were marshaled for a given trainee on a task. Moreover, any actions outside the gaming environment (e.g., personal email) could not be recovered by the evaluation team.

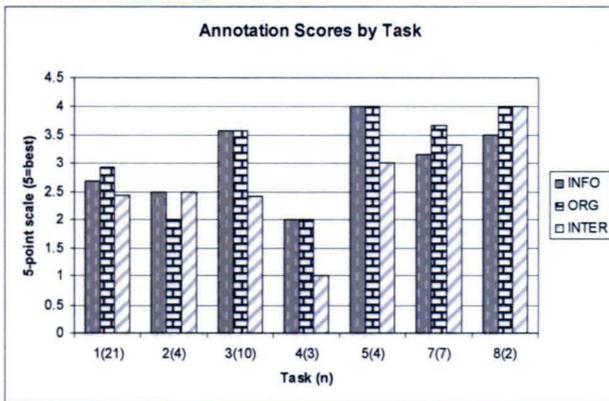


Figure 2. Annotation Scores by Task and MLO

#### 4. POST-EVENT QUESTIONNAIRES

##### 4.1. Development

In order to gather information about the trainees and their impressions of *Dark Waters*, the research team created and distributed pre- and post-event questionnaires to all trainees. The pre-event questionnaire was administered via email; the post-event survey was delivered via a secure website. Participants accessed the site through a link provided by the research team.

##### 4.2. Questionnaire Results

The mean score characterizing a trainee’s prior exposure to game-based training was low at 3.1 (SD=2.6). Based on the questionnaire responses, most trainees did not have prior experience with computer-based training, but commented they were eager to participate. Before the exercise, when asked to rate the likelihood they would value the

helical training experience, the average response was high. One representative comment states:

*I view this as an opportunity to be exposed to new information and ideas. I have been very interested in gaming as an instructional strategy and look forward to being a part of this exercise.*

The post-event questionnaire provided several interesting results. One of the items of greatest concern to observers was the participants’ perception of the LGI. We first wanted to know if the scenario and event content were realistic enough to support the exercise. Responses from more than 80% of participants indicate the scenario was believable as indicated by post-event survey Figure 3.

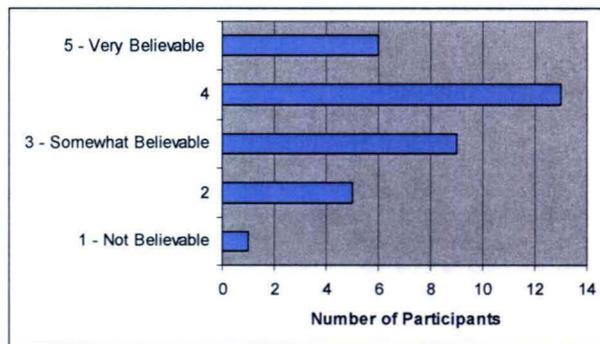


Figure 3. “How Believable or Realistic Was the Storyline?”

Some results of the questionnaire indicated the event did not provide significant training benefit in some of the prescribed learning areas. For instance, participants did not report much improvement in managing information as a result of this training exercise (Figure 4)

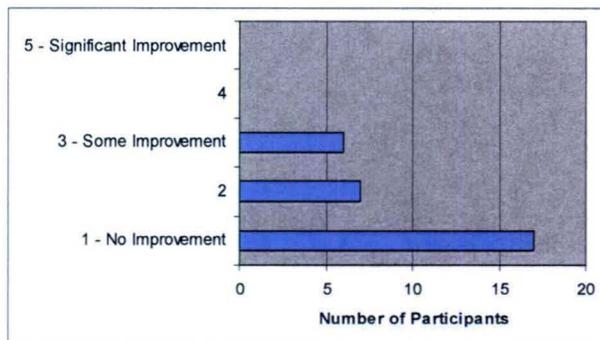
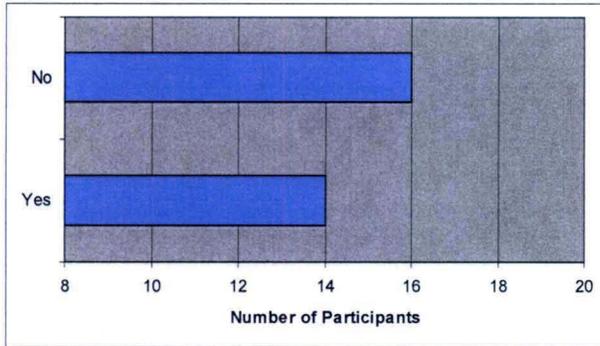


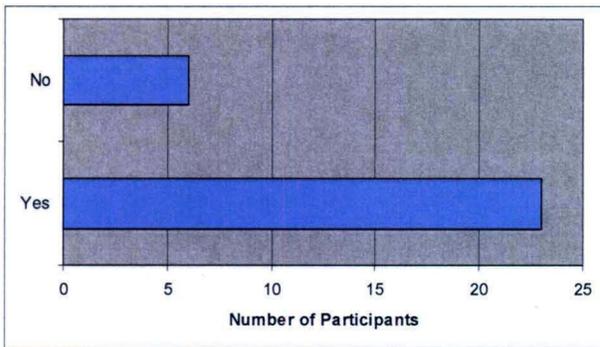
Figure 4. “To what extent did this training event improve your skills in information management?”

The event did seem to make an impact with one of the intended major learning objectives. Almost half of the participants reported learning something about the JTF Organization as a result of the Helical Training Exercise (Figure 5).



**Figure 5.** “Did you learn anything about the JTF Organization that you did not know before?”

Despite some of the scores concerning the self-reported learning improvements, the Helical Training Event was viewed positively by most of the trainees. Overall, participants largely agreed such a training event could be a valuable addition to military training protocols (Figure 6).



**Figure 6.** “Do you feel that this type of training tool is valuable for the military in general?”

## 5. DISCUSSION AND LESSONS LEARNED

The Helical Training approach used in *Dark Waters* is an innovative solution to a problem faced in many corners of the military: how to train officers to work effectively in an operations center or JTF environment. In the past, pre-deployment training for staff officers was comprised mainly of text laden PowerPoint briefs and minimally-attended seminars. It was hoped that the multi-week, low intensity nature of Helical Training would provide incentive (via an enjoyable training experience) as well as a more naturalistic environment for trainees to learn less tangible—but critical—aspects of JTF operations such as information integration and network-building.

As a training genre, Helical Training is in its infancy. As such, there are no established protocols for assessment. This report is an attempt to capture the initial assessment scheme and top-level results

from the *Dark Waters* event. These protocols will be refined in subsequent Helical Training events allowing for more rigorous measurement and assessment of both trainee achievement and observer activities.

The Helical Training Event accomplished its mission as a proof of concept. This first attempt at instantiating the scenario, executing tasks, and collecting assessment material demonstrated the feasibility and effectiveness of ARG-based training. The participant activity log indicates that the LGI was used to find facts, share information, make inquiries and clarifications across the game space, and execute quality responses to task orders. In four of eight tasks, the mean observer rating exceeded the midpoint (3), indicating many participants exceeded the stated task requirements. Instead of simply meeting the minimum criteria, they demonstrated the interest, motivation, and commitment to *Dark Waters*.

In addition to the success in engaging participants, the event also demonstrated that observers could create, collect, and analyze performance data of interest in a helical training environment. This was a nontrivial matter, and though we have learned a great deal to improve the process moving forward, the capability for this first event was impressive. The ability to collect and align data with the learning objectives resulted from implementing a flexible plan that allowed the prudent adaptations as dictated by the unfolding event. In future training exercises, we will augment our ability to imbed the performance measures in the scenario from the beginning providing focused performance measurement opportunities, leveraging the knowledge gained through this event.

*Dark Waters* was a technical and logistical success. Individuals were able to log in, the scenario narrative was able to move forward, and the LGI functioned well [4]. From a pedagogical perspective, it is our opinion that *Dark Waters* was a qualified success. Subjectively, participants exhibited a slight positive trend over the course of task assignments, as measured by the in-game observers. However, the responses by the participants in the post-event questionnaires indicated that learning was minimal. These modest training results should not be overstated for a number of programmatic reasons. For example, the participants used in this event were significantly different from the target trainee population. Some had significant foreknowledge of JTF operations or previous experience in coordination and information management—they began the exercise at ceiling performance and thus did not show improvement. Second, the level of

participation required for realistic coordination to occur effectively was not achieved. A critical mass of people is needed for organizational dynamics to emerge; it is expected that an operational Helical Training event would have this critical mass.

There are two major areas of lessons learned to be reported, with the explicit hope that training impact can be increased in future helical events. The first relates to the MLOs, scenario design, and task assignments. The second relates to the assessment protocols and system interaction.

### 5.1. Participant Performance

Based on the results of the in-game evaluation, the pre- and post-event questionnaires, and the author's subjective assessment of performance, several key points have emerged with regards to participant interaction in a helical training event.

- Participants were willing and excited about involvement prior to the event and maintained respectful levels of participation and effort throughout the duration.
- Embedding performance measures in the scenario tasks in the development stage to assess learning objectives will augment evaluation capabilities.
- Learning objectives should be clearly defined and characterized early in the process. It will also help trainees to understand the objectives prior to the event.
- Task completion and participation may benefit from more realistic task allocation. Rather than a homogeneous allocation—where each person gets the same task—different tasks should be assigned to different participants.
- There are opportunities for both individual and group task assignments. Each should be considered carefully and leveraged to fulfill evaluation/training/practice objectives.

### 5.2. Evaluating the evaluation

In conducting the evaluation, significant knowledge was gained concerning things that worked well and areas to improve the evaluation process in future exercises. Below is a list of some of the important lessons learned from the evaluators' perspective. Following the list are notes collected during the exercise pertaining to specific functions.

- The learning objectives proved sufficiently orthogonal supporting the annotation process.

- Pre- and post-event questionnaires provided good information about the participants and the processes.
- A more formalized method of linking participant actions to the task would be beneficial in collecting relevant material.
- Observers will find it difficult to conduct evaluation in real-time because of the volume of data.
- The LGI as a central repository and system boundary for the exercise is critical.

In all, the *Dark Waters* event proved to be an exciting first-of-its-kind event, and while the development of an effective assessment protocol had challenges, we believe that the methods used could be effectively modified for future use.

## 6. ACKNOWLEDGEMENTS

DARPA and JFCOM funded this work. It was managed by the Office of Naval Research and the Joint Advanced Distributed Learning Co-Lab. This work was performed under contracts N61339-06-C-0084 and N00014-07-0339. The opinions expressed here do not necessarily reflect the official position of the sponsors or the Department of Defense.

## REFERENCES

1. Enspire Learning, (2007). This is Not a Game: Using Alternate Reality Games in Corporate Training. Retrieved June 19, 2008 from [http://enspire.com/about/white\\_paper\\_using\\_alternative\\_reality\\_games\\_to\\_teach\\_data\\_security.html](http://enspire.com/about/white_paper_using_alternative_reality_games_to_teach_data_security.html).
2. Hon, A. (2005). Through the Rabbithole: ARG Lecture. Retrieved November 23, 2006 from <http://mssv.net/2005/11/21/through-the-rabbithole-arg-lecture/>
3. Kim, Jeffrey Y., Allen, Jonathan P., & Lee, Elan. (2008). Alternate Reality Gaming. Communications of the ACM, vol. 51, p. 36-42.
4. Leung, A., Ferguson, W., Roberts, B., Pantaja, E., Weil, S., Geyer, A., & Picciano, P. (2008). Alternate Reality Game (ARG)-Inspired Training for Staff-Level Skills. Proceedings of the 2008 Interservice/Industry Training, Simulation, and Education Conference (IITSEC), Orlando, FL.
5. Szulborski, Dave. (2005). This Is Not A Game: A Guide to Alternate Reality Gaming, New Fiction Publishing.

# Improving Customer Waiting Time at a DMV Center Using Discrete-Event Simulation

Georges M. Arnaut (Ph.D. Student)\*; Shannon Bowling (Ph.D.)  
Engineering Management and Systems Engineering - Old Dominion University (Norfolk, VA)  
[garna001@odu.edu](mailto:garna001@odu.edu)

**Abstract.** Virginia's Department of Motor Vehicles (DMV) serves a customer base of approximately 5.6 million licensed drivers and ID card holders and 7 million registered vehicle owners. DMV has more daily face-to-face contact with Virginia's citizens than any other state agency [1]. The DMV faces a major difficulty in keeping up with the excessively large customers' arrival rate. The consequences are queues building up, stretching out to the entrance doors (and sometimes even outside) and customers complaining. While the DMV state employees are trying to serve at their fastest pace, the remarkably large queues indicate that there is a serious problem that the DMV faces in its services, which must be dealt with rapidly. Simulation is considered as one of the best tools for evaluating and improving complex systems. In this paper, we use it to model one of the DMV centers located in Norfolk, VA. The simulation model is modeled in Arena 10.0 from Rockwell systems. The data used is collected from experts of the DMV Virginia headquarter located in Richmond. The model created was verified and validated. The intent of this study is to identify key problems causing the delays at the DMV centers and suggest possible solutions to minimize the customers' waiting time. In addition, two tentative hypotheses aiming to improve the model's design are tested and validated.

## 1. INTRODUCTION

The usage of simulation has increased noticeably in the recent years due to the advancement of computer technology. The act of simulating behaviors and situations has been adopted in multiple areas like military, social behavior, flight simulators, robotics, etc. In this paper, we use Discrete Event Simulation (DES) since our aim is modeling the DMV system as it progresses over discrete times in a non-continuous fashion. In this model, we attempt to mimic the behavior of the real DMV center system by building our model from variables that are generated from data that is collected from experts of the DMV headquarter in Richmond, VA. The model is examined thoroughly and conclusions and solutions are produced from this study. Additionally, the study identifies two possible scenarios to enhance the system, and determines if they present a statistical significance to the model.

### 1.1 OBJECTIVES OF THE STUDY

The study is conducted in order to (1) minimize the customer waiting time at the DMV – 850 Widgeon

Road, Norfolk, (2) give insights towards minimizing the customer waiting time at all the DMV centers, statewide, or around the country, (3) attempt to improve the existing model (i.e. *Should we add another check-in window?*), and (4) give suggestions aiming for optimizing the system. The focus of the study will be on reducing the following three delays: 1-The ticketing waiting time: time needed to obtain a service ticket. 2-The service window waiting time: time needed to reach the service window and be serviced. 3-The transaction time: time needed to be serviced.

## 2.0 THE MODEL

The model built using Arena (version 10) from Rockwell Systems, is a miniature of the existing DMV center located at 850 Widgeon Rd, Norfolk, VA.

### 2.1 Model Details

Customers arrive in a stochastic way according to an inter-arrival rate produced by exhaustive observations conducted at the Widgeon center. The model has a main queue called the "check-in"

queue. This is the main queue where all customers have to pass through in order to get their tickets and proceed to the nearby seating area, waiting for their ticket number to be announced. When a customer's ticket number is announced, the customer proceeds to one of the service windows in order to be served in a FIFO manner. As there are 14 service windows, according to our extensive observations, only 10 windows are being used at the same time. Thus, for the sake of the study, we consider 10 windows with 10 servers serving on them. There is a separate M/M/1 queue for each of the windows (i.e. 10 separate queues) that we will emphasize graphically in our model. In the real system, we cannot observe these queues separately as the customers are seated all together. There is a range of 10 to 14 servers serving at these windows according to a 'real' weekly schedule obtained from the Widgeon DMV center. The waiting times, transaction times, and other delays are generated from the collected data (please refer to the Data Collection section). The model runs for 8 hours on weekdays (serving from 9am until 5pm) and 4 hours on Saturday (8am until 12pm). Each DMV agent is allowed to take one '30 minutes break' during the day on weekdays, and no breaks on Saturdays. The breaks are divided between three groups. First group of employees can take the break between 11am and 2pm. The second group can take the break between 1pm and 2pm. Finally, the third group of employees can take the break between 2pm and 3pm. As a summary, the model works like the following: Customers arrive to the system, wait at the check in (ticketing) queue in order to obtain a ticket. Then, customers go wait in line in order to obtain a ticket (ticketing wait time) before moving to one of the 10 available service windows. The customers wait for another delay until they are serviced by a DMV agent, which is the service wait time. Then, the customers are serviced with a service delay called Transaction Time. Finally, the customers leave the system.

## 2.2 Model Constraints

Our study has several limitations due to the time factor and the nature of the study:

- The types of services that customers request are ignored. Due to the limitation of the data collected, the types of services are overlooked and all the service delays are recorded as one service type delay. For example, the time that a customer spends for obtaining new car license plates is combined with the time of another

customer trying to obtain a driving license (which is remarkably longer).

- Customers that leave the DMV center for any reasons (e.g. missing documents) are still counted in the time statistics but not modeled in our system.
- The customer's inter-arrival rate and the waiting time at the main ticketing queue are obtained by interviewing experts as well as extensive observation at the DMV Widgeon center [3].
- Holidays as well as the busiest days of the month (i.e. first day and last day of the month) are not counted in our model. However, Fridays and especially Saturdays are considered busier than the other days.
- The DMV employees' weekly schedule is considered static although it changes weekly.
- We assume that customers arrive one at a time to the DMV center.
- All the units used in this study are in *minutes*.

## 2.3 Model Design

The model built using Arena 10.0 is represented in figure1. The key model variables are:

- 10 service windows
- 10 Service Queues - (Waiting Time)
- Customers arriving in a stochastic way to the center
- Service Time (or Trans Time)
- Main queue – customer check in (time needed to acquire a ticket)
- Number of DMV agents defined by a schedule ranging between 10-14 during the day (including breaks)

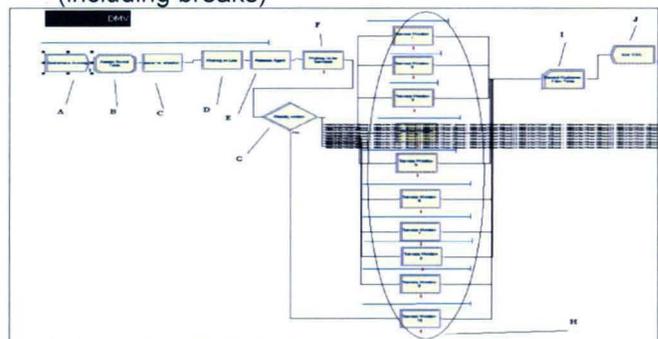


Figure 1: DMV Model Design in Arena 10.0

- A:** Entity creating the users arriving to the DMV in a stochastic way.
- B:** Recorder that catches the time of arrival of the customers for further calculation
- C:** Module that seizes the agent on the main check in queue in order to serve the customer

**D:** Module that delays the service time of the agent according to a distribution explained in the Data Collection section

**E:** Module that releases the agent after finishing from servicing the customer at the check in queue

**F:** Module that delays the customer service time before reaching the service window according to the next available window queue (refer to the Data Collection section).

**Table 1: DMV Weekly Schedule**

		26-Jan	27-Jan	28-Jan	29-Jan	30-Jan	31-Jan
		MON	TUES	WED	THURS	FRI	SAT
o	Rose	8:00-5:30	8:00-5:30	sdo	8:00-5:30	8:00-5:30	8:00-1:00
c	Brenda	8:00-6:00	9:15-6:00	9:15-6:00	9:15-6:00	9:15-6:00	sdo
o	Debra	8:00-5:30	8:00-5:30	8:00-5:30	sdo	8:00-5:30	7:30-12:30
c	Lillie	8:00-6:00	8:30-6:00	8:30-6:00	8:30-6:00	sdo	8:00-1:00
	Andrea	8:00-5:45	8:30-5:45	8:30-5:45	sdo	8:30-5:45	7:45-12:45
	Aqwanda	8:00-5:45	sdo	8:30-5:45	8:30-5:45	8:30-5:45	7:45-12:45
	Brian	8:00-5:45	8:30-5:45	sdo	8:30-5:45	8:30-5:45	7:45-12:45
	Carolyn	8:00-5:45	sdo	8:30-5:45	8:30-5:45	8:30-5:45	7:45-12:45
	Gia	sdo	8:30-5:45	8:30-5:45	8:30-5:45	8:30-5:45	7:45-12:45
	James	sdo	8:30-5:45	8:30-5:45	8:30-5:45	8:30-5:45	7:45-12:45
	Marvita	8:00-5:45	8:30-5:45	8:30-5:45	sdo	8:30-5:45	7:45-12:45
	Melvina	8:00-5:45	8:30-5:45	8:30-5:45	8:30-5:45	sdo	7:45-12:45
	Quanet	AT	AT	HPTON	8:45-5:30	8:45-5:30	sdo
	Stephen	8:00-5:45	8:45-5:30	8:45-5:30	8:45-5:30	8:45-5:30	sdo
	Theresa	8:00-5:45	8:30-5:45	8:30-5:45	8:30-5:45	8:30-5:45	7:45-12:45
	Shaney	8:00-5:45	TRNG	TRNG	TRNG	sdo	7:45-12:45
	P-14's						
	Tarameka	11:00-4:00	8:00-4:00	8:00-4:00	11:00-4:00	sdo	7:45-12:45

**G:** Module that decides which is the next available service window according to a small program to calculate the service window containing the least number of customers waiting in its queue.

**H:** Ten service windows that serve the customers.

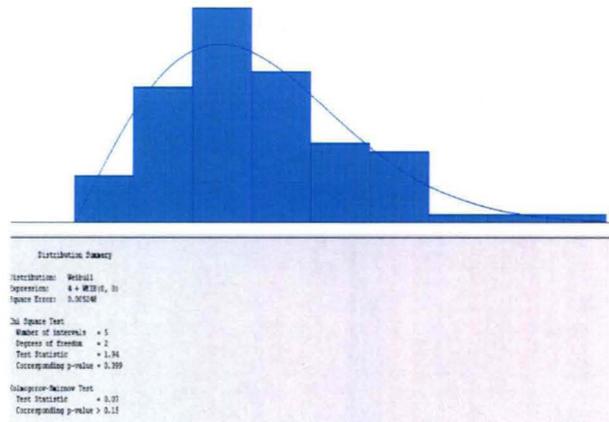
**I:** Recorder that calculates the flow time of the customers in the system using the previous recorder (B).

**J:** Module that allows the customers to exit and leave the system.

### 3. DATA COLLECTION

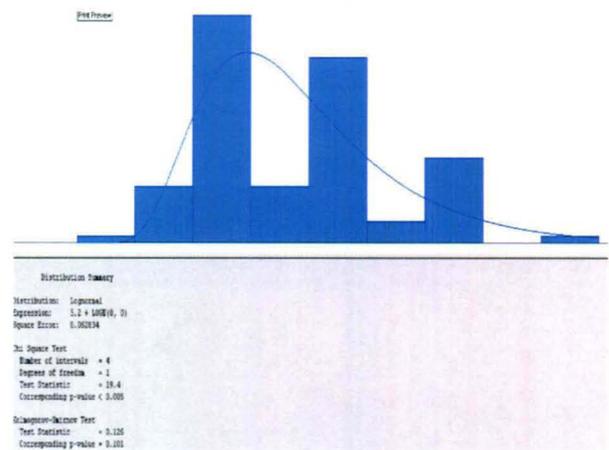
The data is mainly collected from extensive observation of the center, interviewing experts from the headquarters, and from the DMV weekly data sheets provided by the DMV experts [2]. The weekly data collected from the DMV experts provided with around 100 data points that were used to generate the distributions of the related delays. The customers *inter-arrival rate* is generated from a schedule that resulted from extensive observation and interviewing the system's experts. The arrival schedule is implemented from Monday through Saturday. On weekdays, the customers arrive between 9am and 5pm. On Saturdays, the customers arrive between 8am and 12:00pm. The *ticketing waiting time* is generated from a Triangular distribution that resulted from extensive observation of the center

and the behavior of customers. The expression used is *EXPO (0.62)*. The *service waiting time* delay is generated from a Weibull distribution that resulted from plotting the data points on a histogram (refer to Figure 2), provided by the DMV weekly data. The expression produced is:  $4 + WEIB(17, 2)$ . We considered this distribution a good fit for our collected data because it has a very low Square Error (which is 0.005248) and the p-value is remarkably larger than 0.05.



**Figure 2: Histogram of Service Waiting Time Delay**

The *transaction time (service time)* is generated from a log normal distribution that resulted from plotting the data points provided by the DMV weekly data in the histogram shown in Figure 3. The expression produced is:  $5.2 + LOGN(1.81, 0.803)$ . Figure 3 plots the histogram of the data collected and includes the distribution summary. We consider this distribution a good fit for our collected data because it has a very low Square Error (which is 0.0628) and the p-value: 0.101 is remarkably larger than 0.05.



**Figure 3: Histogram of the Transaction Time Delay**

#### 4. OUTPUT ANALYSIS

After running the model for 10 replications where each replication represents a week composed of 6 business days, we came up with the following results:

- The weekly average of customers coming to the DMV center of Widgeon Road is 2054 customers.
- The total average waiting time for each customer (or customer flow time) is 41.65 minutes.

The queuing delays are represented in details in Table 2 below:

**Table 2: DMV Model Statistical Results**

Waiting Time	Average	Half Width	Minimum Average	Maximum Average
seize for checkin Queue	3.7772	1.63	1.5112	7.9814
Service Window 1 Queue	2.3072	0.72	1.0726	4.3667
Service Window 10 Queue	29.0041	7.14	11.3725	49.2350
Service Window 2 Queue	15.7113	4.16	7.0563	27.7253
Service Window 3 Queue	18.4753	5.05	8.0317	34.3370
Service Window 4 Queue	20.9472	5.42	9.1003	36.8350
Service Window 5 Queue	23.0120	6.09	9.1411	40.8050
Service Window 6 Queue	25.0501	6.60	9.4656	44.2974
Service Window 7 Queue	26.3464	6.90	9.7793	46.2830
Service Window 8 Queue	27.2823	6.87	10.3778	46.2237
Service Window 9 Queue	28.4466	6.99	10.9392	47.6204

We observe that there is an average of 3.77 minutes wait time at the check-in queue (ticketing) which we categorized as fair. In addition, looking at the queuing occurring at the service windows, the average transaction time was around 21.65 minutes. Here, it is necessary to mention that the service window delays are not the delays only related to the service time at the window, but also includes the waiting time before the customers get served by an agent. The average service time is around 8 min which in our opinion does not need further improvement.

That leaves the deficiency of the system to only one variable which is the excessive arrival rate of the customers, which in turn, affects all the other delays causing the excessive queues.

#### 5. MODEL VERIFICATION AND VALIDATION

Our V&V was conducted in parallel with the system's experts at the DMV headquarters. By comparing our model's results with their weekly data, we've found that our generated distributions, our model (with its variables), and our results were valid. For the inter-arrival rate of the customers, according to the DMV weekly sheets (of the real system), an average of 2105 customers visited the DMV at Widgeon weekly. According to our model, the average was 2054 customers which is

considered very close, and therefore valid. As for the other delays (e.g. Transaction delay), the distributions were verified via emails with a senior analyst at the DMV headquarters in Richmond [reference needed]. The final results of the study and the possible solutions were submitted to the headquarters upon their request and are being studied by their analysts.

#### 6. ALTERNATIVE SCENARIOS

Observing the customers' waiting time at the ticketing window in our model's animation, in Table 2, and in the real system, inspired us to come up with two different alternatives different from having one main ticketing queue.

##### 6.1 Alternative 1

We considered *having an additional ticketing window resulting in two parallel check-in queues that are served by two agents*. After implementing this addition to our model and running it for 10 replications (just like the original model), we realized that the customer waiting time was slightly reduced from 41.65 minutes to 40.28 minutes. In order to find out if this alternative was worth implementing, we conducted a Paired-t test on the two approaches (this one and the original model). We concluded that the change was not statistically significant (the two means overlapped at 0.05 level).

##### 6.2 Alternative 2

The second alternative was *to increase the number of agents by having 14 agents working at all times (including breaks)*. This alternative has two sub-alternatives. One, by increasing the number of agents to the original model without adding another ticketing queue (i.e. having two ticketing windows), and the other one by adding another ticketing window. After implementing the changes (in both sub-alternatives) and running the model for 10 replications, we concluded that this scenario is also not statistically significant.

#### 7. CONCLUSION

At the beginning of the study, we were considering the proposal of a *Self-Check in kiosk* to speed up the ticketing phase, as a parallel approach to the main check in service window. Our reason for proposing such an approach was our belief (prior to running the model) that having two separate (but parallel) check-in queues will speed things up and minimize the customers' waiting time. After

implementing and running the model, and after experimenting with the two alternatives proposed (refer to section 6), we concluded that having a self-check in kiosk will not have a significant positive difference on the existing model, and therefore decided to drop that suggestion. Thus in our opinion, this limits the delay to two main system design gaps. Either the service time (transaction time) is relatively high, or the arrival rate is just too excessive. For the first gap, the service time can be reduced by increasing the number of agents but also, increasing the number of service windows proportionally. This would reduce the service time remarkably and affect the overall waiting time of the visiting customers. Here, it must be noted that the pace of the service is relatively fine. The serving pace does not need to be enhanced since according to our records, the average transaction time for each customer is 21.65 minutes (including the time waiting to be serviced), which is relatively fair. Therefore, the queuing is not occurring from the transaction time (i.e. service time). As for the second gap, the

arrival rate can only be reduced by offering more online services (but also keeping the option of physically visiting the DMV center for these services). This will reduce the arrival rate of customers since, with the digital age and the ease of access to go online, customers would most probably prefer conducting the transactions online (e.g. from their work office) rather than spending time to visiting the DMV center. Several DMV centers started giving appointments to their customers in order to balance and control their inter-arrival rate. This approach is also feasible and could be implemented at the Widgeon center.

#### REFERENCES

1. DMV. (2009). Retrieved April 25, 2009, from [www.dmv.state.va.us/webdoc/about/index.asp](http://www.dmv.state.va.us/webdoc/about/index.asp)
2. Powell, Minni, personal communication, June 2009
3. White, Robert, personal communication, July 15, 2009



# Implementing And Simulating Dynamic Traffic Assignment With Intelligent Transportation Systems In Cube Avenue

**Peter Foytik, Mike Robinson**  
VMASC Old Dominion University  
*pfoytik@odu.edu, rmrobinson@odu.edu*

**Abstract.** As urban populations and traffic congestion levels increase, effective use of information and communication tools and intelligent transportation systems as becoming increasingly important in order to maximize the efficiency of transportation networks. The appropriate placement and employment of these tools within a network is critical to their effectiveness. This presentation proposes and demonstrates the use of a commercial transportation simulation tool to simulate dynamic traffic assignment and rerouting to model route modifications as a result of traffic information.

## 1. INTRODUCTION

Modeling and Simulation (M&S) of transportation is critical to developing and assessing proposed ideas and technologies. Simulations of past transportation events allow planners to better understand what really happened. By simulating future changes, decision makers can greatly improve the roadways of tomorrow.

Alternatives are frequently proposed in different locations of cities for the future development of city and federal roadways. Proper testing of proposed plans must be done to assure best solution. One area of transportation system improvements that has largely not benefited from M&S testing is the installation or improvement of Intelligent Transportation Systems (ITS).

IEEE's Intelligent Transportation Systems Society defines ITS as systems that utilize synergistic technologies and systems engineering concepts to develop and improve transportation systems of all kinds. ITS refers to efforts to add information and communications technology to transport infrastructure. It strives to apply advanced technology to resolve the problems of surface transportation by improving efficiency, safety, and mobility. Other objectives include reducing energy, economic costs, and damage to the environment [2]. To better improve the planning of a large area such as the region of southeastern Virginia, ITS should be tested over the entire network to assess the improvements in traffic flows and congestion levels. This document will describe efforts and research to implement ITS and vehicle driver effects from ITS in a mesoscopic model using Avenue from the Cube family of transportation software.

## 2. TECHNICAL BREAKDOWN OF CUBE

### 2.1 Cube A Transportation Tool

Cube family of transportation tools developed by Citilabs is chosen for this projects as the tool of choice because it is already selected as the planning standard by the Virginia Department of Transportation (VDOT). Cube provides a macroscopic transportation modeling tool, a microscopic modeling tool and a mesoscopic modeling tool, each of these tools can integrate together by sharing loaded networks. It also allows the modeler additional control with its scripting language allowing the ability to program in vehicle reactions that the software tool was not developed or intended to do through the default user interface. The scripting language is proprietary, and offers flexibility to make changes to road networks and Origin Destination matrices (OD).

#### 2.1.1 Microscopic Dynasim

The Cube microscopic tool Dynasim is like other transportation micro simulation in that the user can simulate individual vehicle behavior creating a very accurate simulation. The problem with a micro simulation is the amount of time required to develop and run a scenario [4]. This problem usually requires the simulated area to be reduced to a more manageable size so that the simulation can run in a reasonable amount of time. Therefore, if the interest of the study is to see the effects of ITS on an intersection, a microsimulation would work quite well. This type of simulation will show you the local effects in a very small area very well, but what if the planners need to see effects in a larger scale in multiple locations at the same time? A microscopic simulation could accomplish this but require much more time to set up and to run.

### 2.1.2 Macroscopic Voyager

The macroscopic tool within Cube, Voyager, is probably the most used and well known tool within the suite. Voyager can calculate volumes of traffic over large networks. It offers a number of modules that users can use to simulate transportation demand macroscopically. Each module requires its own input files, and using either a script file generated by the module to perform a default task, or a script created by the user to do a unique task can produce many outputs. Once the script runs the task the module can generate a number of outputs files of various formats that can be used as inputs to other modules or as strict outputs that visualize data.

Voyager runs the highway module which produces the calculated values on each segment of the network for a period of time chosen by the user using a gravity model [1]. The highway module takes as in input a daily demand matrix, then uses the command pathload to run the volume over the network and using a gravity model to find an equilibrium over the network. The pathload command within highway takes a few inputs, one being the path variable. The path variable is used to set impedence over the roadways that are being simulated with pathload. The model developer can select different roadways to run with different pathload statements allowing multiple impedances over the entire road network. A typical example would be to see the average congestion for the Hampton Roads area for an entire day. In this case, the user could color code the road segments to display a range of colors representing the value of congestion. This type of output is useful for showing daily traffic and is able to highlight the roadways that are being overused. Using Voyager to model ITS is very possible and can show the change in volume on roads due to ITS in a static sense. However if the planner wants to simulate over time how vehicles are changing direction and routes, then the macroscopic model will not completely accomplish that.

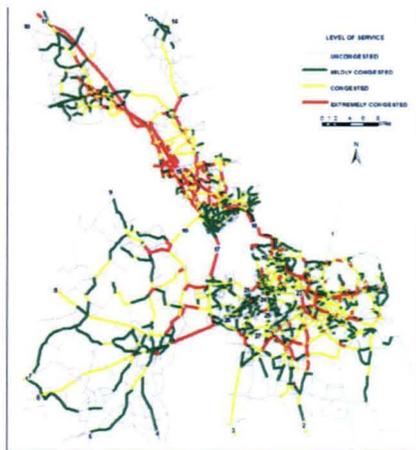


Figure 1: Shows an image of a Voyager output network with links color by level of service.

### 2.1.3 Mesoscopic Avenue

Mesoscopic models are in between a micro and a macroscopic model, allowing traffic volume to change over time through a large scale system [4]. While some mesoscopic simulation tools have more micro than macro features, Cube has more macro than micro features in that it simulates volume over the road network. Visualization of the output animation appears as a microscopic simulation where but it visualizes packets of vehicles based off of the volume calculations instead of individual vehicles. The mesoscopic tool in Cube is very closely related to the macroscopic tool, so close that it is actually just another module added to Voyager.

Avenue is capable of reading in a list of OD matrices, one for each time segment and a network file. Time segments are defined time steps that the simulation advances and also are the defined moments when new volume can be added to the system by a new OD matrix. At each time segment the simulation will run the volume over the network as a discrete event simulation finding equilibrium then doing the same thing for the next time segment. All of this is done through Avenue's Dynamicload statement. Much like the highway module's pathload, Dynamicload uses a gravity model to calculate equilibrium, but instead of for one OD Matrix for one time period Dynamicload will calculate equilibrium for each time segment using the calculated equilibrium from the previous time segment. The output files that Avenue produces are matrix files, network files, data/text files, path files, packet log files and a few other types of outputs. The most important output file is the network file which contains values on all of the road segments from the last simulation run. Most of the values that are outputted on the road segments have a value for each time segment that the simulation ran. For example, volume, queue length, speed, and time are default outputs that Avenue provides, with each segment representing each variable as *variable\_t* where *t* equals the time segment it represents and variable representing the name of the variable. This allows the user to color code the road segments over time using the Bandwidth chart display. A bandwidth chart is a display that gives the user controls to advance time and to see how that particular value changes.

SPEEDS_2	60	60
SPEEDS7_2	60	60
VS1_2	474	43
VS2_2	529	66
VS3_2	36	26
VS4_2	56	14
VS5_2	39	19
VS6_2	52	17
VS7_2	46	12
QUEUEVS1_2	0	0
QUEUEVS2_2	0	0

Figure 2: Shows an output from one of the road segments in a simulation run providing values of Volume at each time segment.

Packet log output is a text file containing a record of locations where the packets have traveled for each time segment. The user can load this file over the network and view animations of the transportation that was simulated. The animation is a view of packets represented as rectangles over the loaded network. Users can control time to advance at different speeds and as time progresses the animation displays packets traveling the routes they were simulated to take towards their destinations. Since the packet log file is a text file, script files can be written with a matrix module to parse the log file and determine data that can be presented in a user created output file. An example of a parsing task would be to locate the amount of time segments it took each packet to arrive at its destination, and then average that value to obtain the average travel time for the simulation.

Avenue's dynamic assignment and flexibility along with its informative outputs and helpful visualizations make it a great choice for modeling vehicle behavior from ITS. Large areas can be simulated in a reasonable amount of time and in a time stepped simulation. Mesoscopic simulation benefits the non technical planners who need to understand how the simulated system is affected by driver behaviors.

### **3. IMPLEMENTING A DAILY MESOSCOPIC MODEL**

#### **3.1 A Hampton Roads Mesoscopic Model**

Implementing a daily mesoscopic model for a major metropolitan and especially in the Hampton Roads can be challenging. Demand must be generated for all origins and destinations at each time segment. This is typically done by taking percentages of the daily demand over each time segment. The problem with this technique in a Hampton Roads model is that the traffic patterns here resemble two peak load curve for most routes. One peak represents vehicles going one direction in the morning, and another representing the same traffic returning in the evening. The best solution will have demand values for the morning, lunch and evening traveling in the appropriate direction. Because the tests being done now are prototype tests, smaller percentages of the daily demand will work.

#### **3.2 Mesoscopic Model For Testing**

Specific tests require manually injecting traffic in a test area and applying congestion to one of the roadways in the way of this injected demand. This process is much like the process of doing a microscopic test in that only a small test area is being worked on. The demand is also set up much like the microscopic simulation where an origin destination matrix needs to be defined for each time segment. Once the test area is performing as expected the same type of driver intelligence can be applied in multiple areas for

the entire Hampton Roads network, or in one spot with demand over the entire network to cause different reactions to the test area.

## **4. CAUSING CONGESTION**

### **4.1 Implementing Incidents**

In order to realistically model driver behaviors and the influence of ITS, congestion must also be simulated. There are two ways that congestion can be accurately portrayed in Avenue. The easiest method is to overload the system with large amounts of vehicle traffic volume, a more realistic method involves injecting a simulated incident by reducing road capacity. The most precise way to create congestion where needed is to create an incident. Overloading the system with traffic volume is effective but can be unpredictable as traffic could overly congest areas that are not of interest.

In Cube there is not a default function to apply incidents, so to implement incidents the modeler has to be able to use the Cube scripting language. The incidents modeled in Avenue require that the incident last as long as the time segment. A modeler cannot request Avenue to reduce the capacity of a roadway for one half of the time segment because all dynamic changes and calculations are done by time segment. Therefore to model fifteen minute incidents Avenue would require fifteen minute time segments or a different capacity reduction would need to be calculated. This new capacity would equal the capacity effects of a fifteen minute incident but at a one hour value [5]. The locations and severity of the incidents can be selected from historic data of road segments in Hampton Roads that are more likely to have an incident [5], or can be assigned to specific areas that a planner would like to study.

The command Avenue uses to reduce capacity on road segments is Dynamic. The Dynamic command only works with the variable C. In Avenue to change the capacity of the road segment the C variable has to be changed which represents the capacity on the road segment over the entire simulated time. When the Dynamic command is used with C, a value of C is calculated for each time segment. The modeler can then alter the C value for any road segment at any defined time by using Dynamic and C. The script line needs to contain the A node (which refers to the starting point of a segment) and B node (which represents the end point of a segment) of the link segment, this is to assure that the incident occurs in the right direction and on the right road segment. The capacity is usually reduced by multiplying the current C variable by a reducing factor. When the time segment of the incident is over, the C variable is calculated by its normal calculation, capacity multiplied by lanes multiplied by simulation length.

$IF(li.A = 66537 \text{ and } li.B = 66791)$

$DYNAMIC C[28] = li.C \times 0.5$

The equation above shows a conditional statement that locates the link with an A node equal to 66537 and a B node equal to 66791 then sets its C variable to half of its normal value at time segment 28.

## 5. SIMULATING DYNAMIC BEHAVIOR

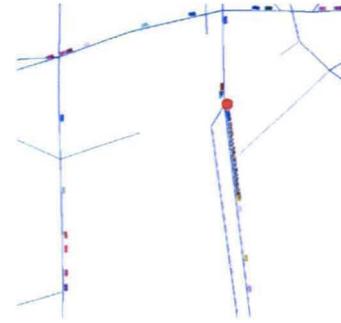
### 5.1 Road Impedance To Control Behavior

Currently in Avenue dynamic behavior is already being simulated. The idea of this study is to better control that behavior and accurately simulate what is really happening. As stated in the documentation of Avenue, impedance for all road segments can be defined by the user. This impedance can be altered based on the road cost value, time to traverse link values, and user defined values. Smarter traffic can also be simulated by running multiple iterations of Avenue allowing the gravity model better equalize the network reducing congestion by having vehicles choose different routes based on the knowledge of the previous run simulation. This is accurate to a point, if an incident is being simulated then a multi iteration run is not going to be realistic. Therefore Avenue needs to be manipulated to allow the impedance of some specific roads to change at different times.

When the evacuation model of the Hampton Roads was developed a compliance variable was implemented to control the behavior of vehicles sticking to the evacuation routes [5]. This same principal can be applied to this study. Early implementations involved using compliance variables over the network for each time segment. These compliance variables could then route vehicles around incident areas as if there were information alerting vehicles of these areas.

The problem is that vehicles need to approach the accident segments as if knowledge of the accident doesn't exist. Then once the accident knowledge can be distributed the vehicles need to make an attempt to divert. This clearly shows that the impedance needs to change dynamically. The only way to change the impedance is to change the path variable within Dynamicload.

Dynamicload is Cube's dynamic analog of the static PATHLOAD which is the heart of the macroscopic simulation, and takes as an input a list of volumes for each time segment as well as a path impedance (1). The path variable can be set to COST, TIME, or a list of working link variables (LW). LW variables can contain values of link impedance or impedance equations and can be set, then altered after each time segment in the ADJUST phase of Avenue. This provides a nice dynamic adjustment to the link impedance providing a more controlled environment to produce dynamic behavior in a simulated ITS event.



**Figure 3:** Shows an Avenue output network with an incident occurring on a road segment indicated by the large red dot.

#### 5.1.1 Impedance set to Cost and Time

By using a mixture of Cost and Time a simple and deterministic ITS system can be simulated. By using two Dynamicload statements ITS behavior can be achieved by using one statement for the time segments where ITS is being simulated and the other statement for normal times when ITS is not active or needed. This is accomplished by using Cost as the impedance for the Dynamicload simulating the ITS time segments and using Time as the impedance for the other Dynamicload statement to simulate impedance without ITS. This method also requires that your demand cooperates with the Dynamicload statements. For example during normal road impedance the regular demand matrices will be applied to the Dynamicload that has its path equal to Time, and during those time segments the other Dynamicload's demand matrices should equal zero. When ITS effects need to be simulated then the Dynamicload with path equal to Cost takes the regular demand matrices and the Dynamicload with path equal to Time takes the zero demand matrices. This method keeps the regular volume flowing onto the network at all times and seamlessly simulates deterministic behavior due to ITS.

#### 5.1.2 Impedance set to LW variables

Deterministic behavior really isn't enough to simulate the true behavior of ITS effects so instead of using just Cost and Time as impedance, LW variables are used. LW variables for each time segment give the ability to simulate different behavior to the entire system. To simulate certain dynamic behaviors caused by ITS, specification of road segments to road impedance needs to take place. New variables can be assigned to the road network to give weighted values that can specify which roadways will be effected and which roadways will stay the same. By setting all normal roadways to a weight of 1 and the effected roadways to a value greater than one, a multiplicative operation to the impedance equation will result in a greater value for the effected road segments and a normal value for the non effected segment.

## 6. FUTURE WORK AND CONCLUSION

Development and testing of these scenarios have a ways to go, but current tests show control over the traffic in a way that can be made more realistic to mimic real driver behaviors. Using driver survey's to obtain data that can produce frequencies of when drivers decide to abandon a normal route because of information or congestion and to reroute either to a known or an unknown route. These frequencies can then be applied to the LW variable equations to create realistic simulations. Then using the data from the surveys the model can be validated to the number of vehicles that potentially would reroute. More tests of manipulating the LW values to be altered by time values per segment as well as congestion values are being done. Currently Avenue allows time to dynamically alter impedance variables but the results are inconsistent and need to be verified.

## REFERENCES

1. *Cube Voyager Reference Guide*, Version 5.0, Citilabs inc, 2007-2008.
2. W. Barfield and T. Dingus, "Human Factors in Systems", Lawrence Erlbaum Associates. August 1997
3. A. De Palma and F. Marchal, "Real Cases Applications of the Fully Dynamic Metropolis Tool-Box: An Advocacy for Large-Scale Mesoscopic Transportation Systems", Kluwer Academic Publishers, 2002.
4. M. Ben-Akiva, M. Bierlaire, H.N. Koutsopoulos, and R. Mishalani, "DynaMIT: a simulation-based system for traffic prediction and guidance generation", Presented at Tristan III, June 1998.
5. R.M. Robinson, "Hampton Roads Hurricane Evacuation Transportation Study," Final Report, VMASC, Suffolk VA, 2008.



# An Optimization Model for Scheduling Problems with Two-Dimensional Spatial Resource Constraints

Christopher Garcia, Ghaith Rabadi

*Engineering Management and Systems Engineering  
Old Dominion University*

[cgarc001@odu.edu](mailto:cgarc001@odu.edu), [grabadi@odu.edu](mailto:grabadi@odu.edu)

**Abstract.** Traditional scheduling problems involve determining temporal assignments for a set of jobs in order to optimize some objective. Some scheduling problems also require the use of limited resources, which adds another dimension of complexity. In this paper we introduce a spatial resource-constrained scheduling problem that can arise in assembly, warehousing, cross-docking, inventory management, and other areas of logistics and supply chain management. This scheduling problem involves a two-dimensional rectangular area as a limited resource. Each job, in addition to having temporal requirements, has a width and a height and utilizes a certain amount of space inside the area. We propose an optimization model for scheduling the jobs while respecting all temporal and spatial constraints.

## INTRODUCTION

Scheduling problems arise in many areas of business and industry. Common to all types of scheduling problems is the need to assign a set of resources or jobs to a set of time slots. Beyond this, different types of scheduling problems present their own unique sets of objectives and constraints and require individualized formulation and solution methods. In this paper, we examine a problem that involves scheduling jobs that have two spatial dimensions, width ( $x$ ) and height ( $y$ ), in addition to having required processing times, deadlines, and earliest start times. These jobs must be processed inside a two-dimensional processing area which has its own width and height. Consequently, solving this problem involves simultaneously determining both the time each job should be processed as well as the spatial location and layout of each job within the processing area.

To demonstrate an instance of this problem, consider the following jobs listed below in Table 1. Assume that we are given a processing area having a width of 10 and a height of 8 in which these jobs must be processed, and that our

objective is to minimize the total tardy time. In order to solve this problem we must determine both a start time and a coordinate for each job. To complicate matters further, we also assume that the *layout* of any job can be changed, which is accomplished by rotating the job by 90 degrees. This results in a swapping of the job's width and height. For instance, Job 1 has a width of 4 and a height of 5. If its layout is changed, Job 1 will have a width of 5 and a height of 4. An optimal solution to this problem is given in Table 2. The lower-left corner of the processing area can be understood as the (0, 0) coordinate, enabling the solution to be visualized as shown in Figure 1.

Although much literature exists for both box-packing problems and scheduling problems, there is relatively little literature that directly addresses spatial scheduling. Literature addressing this topic directly addresses a more specialized problem encountered in shipbuilding [1], [2]. To our knowledge the problem proposed in this paper has not been discussed in previous literature, and we are aware of no previous problem instances.

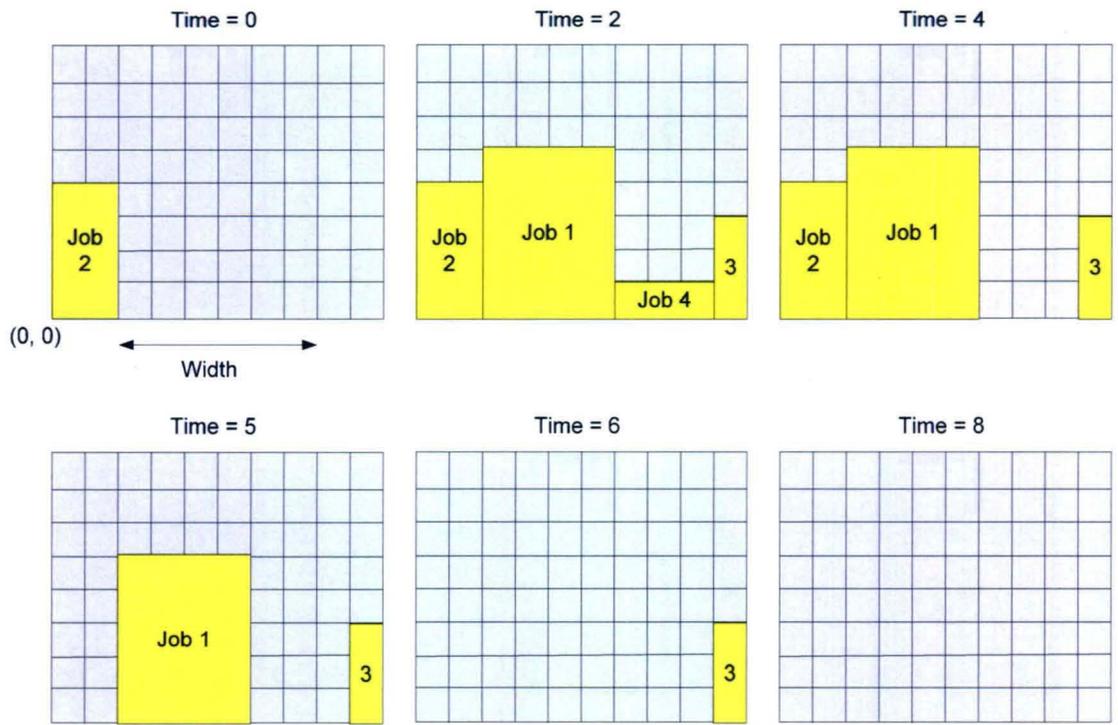
**Table 1:** A small problem instance

Job	Width	Height	Earliest Start	Processing Time	Deadline
1	4	5	2	4	8
2	2	4	0	5	6
3	1	3	2	6	10
4	3	1	0	2	9

**Table 2:** An optimal solution to the small problem instance, in ascending start-time order

Job	Start	End	Due	Tardy	X	Y	Width	Height
2	0	5	6	0	0	0	2	4
1	2	6	8	0	2	0	4	5
4	2	4	9	0	6	0	3	1
3	2	8	10	0	9	0	1	3

**Figure 1:** A visualized solution to the small problem instance



## A GREEDY APPROXIMATION HEURISTIC FOR OPTIMIZATION

The incorporation of spatial resources into scheduling problems introduces a considerable amount of complexity [6]. Additionally, because we were not aware of any prior work on this problem our aim was to develop a computationally fast approximation heuristic to provide baseline solutions for evaluating the performance of future methods. The algorithm we developed combines a two-dimensional box-packing algorithm with an earliest-deadline-first scheduling algorithm. Each job is understood to be a structure having an (X,Y) coordinate, width, height, processing time (duration), earliest allowable start time, deadline, start time, and end time.

### Spatial Operations

The spatial operations are used to determine the spatial location and layout for each job within the processing area. We assume we are given a width and height of the processing area as well as constants  $m$  and  $n$ . The area is divided into  $(m \times n)$  discrete units –  $m$  units in the width dimension and  $n$  in the height dimension. Thus, each unit's width is  $(width / m)$  and its height is  $(height / n)$ . An  $(m \times n)$  matrix of binary values called *Area* is used to keep track of which units are covered by a job. Here, 1 denotes that the corresponding unit is occupied and 0 denotes that it is not. The packing procedure returns either a coordinate for the job (in the case of success) or  $\emptyset$  if it is unable to fit the job into the area.

Three specific procedures are used: PACK, GREEDY\_PACK, and REMOVE. PACK is the top-level packing procedure and attempts to find a feasible coordinate and layout for a given job. GREEDY\_PACK attempts to fit a job into the available space without changing the layout. REMOVE frees the space occupied by a job. These procedures are described in pseudocode in Table 3. Comments follow the # symbol.

**Table 3:** The PACK, GREEDY\_PACK, and REMOVE procedures

<pre> Procedure PACK (Job) DO:   Coordinate := GREEDY_PACK(Job)   IF Coordinate = <math>\emptyset</math> DO:     # Change layout – rotate job 90 degrees     SWAP(Job.width, Job.height)     RETURN GREEDY_PACK(Job)   ELSE     RETURN Coordinate   END END </pre>
<pre> Procedure GREEDY_PACK (Job) DO:   FOR <math>1 \leq i \leq m - \lfloor \frac{(m)(Job.width)}{width} \rfloor</math> DO:     FOR <math>1 \leq j \leq n - \lfloor \frac{(n)(Job.height)}{height} \rfloor</math> DO:       <math>P := \{ p \mid i \leq p \leq i + \lfloor \frac{(m)(Job.width)}{width} \rfloor \}</math>       <math>Q := \{ q \mid j \leq q \leq j + \lfloor \frac{(n)(Job.height)}{height} \rfloor \}</math>       IF Area[p][q] = 0 for all <math>p \in P</math> and <math>q \in Q</math>       DO:         Area[p][q] := 1 for all <math>p \in P</math> and <math>q \in Q</math>         Job.X := (width)(i) / m         Job.Y := (height)(j) / n         RETURN (Job.X, Job.Y)       END     END   END   RETURN <math>\emptyset</math> END </pre>
<pre> Procedure REMOVE (Job) DO:   <math>i := Job.X / m</math>   <math>j := Job.Y / n</math>   <math>P := \{ p \mid i \leq p \leq i + \lfloor \frac{(m)(Job.width)}{width} \rfloor \}</math>   <math>Q := \{ q \mid j \leq q \leq j + \lfloor \frac{(n)(Job.height)}{height} \rfloor \}</math>   FOR ALL <math>p \in P</math> and <math>q \in Q</math> DO:     Area[p][q] := 0   END END </pre>

## The Scheduling Algorithm

The scheduling algorithm is concerned with two types of events: 1) the next time one or more jobs are eligible to be processed/added to the area, and 2) the next time one or more jobs are complete and can be removed from the processing area. Two job lists are utilized to keep track of these concerns. *Open* is a list of all unscheduled jobs, sorted in ascending order by

earliest start time. *In\_Processing* is a list of all jobs currently inside the processing area and is sorted in ascending order by end time. It should be noted that ADD and DELETE operations on these lists preserve their order. The algorithm assigns job times as it moves jobs from *Open* into *In\_Processing*. The pseudocode for this procedure is found in Table 4 below, and comments follow the # symbol.

**Table 4:** The scheduling algorithm

```
Procedure SCHEDULE (Jobs). DO:
  Open := SORT Jobs BY earliest start time
  In_Processing := ∅
  Time := 0

  WHILE Open IS NOT EMPTY DO:

    Next_Finished_Job := NEXT_ELEMENT(In_Processing)

    # Remove jobs that are finished
    WHILE In_Processing IS NOT EMPTY AND Next_Finished_Job.end_time ≤ Time DO:
      REMOVE(Next_Finished_Job) # Procedure defined in Table 3 above
      DELETE Next_Finished_Job FROM In_Processing
      Next_Finished_Job := NEXT_ELEMENT(In_Processing)
    END

    # Add jobs until we run out of space or all open jobs have been added
    Next_Job := NEXT_ELEMENT(Open)
    Time := Next_Job.earliest_allowable_start_time
    Coordinate := PACK(Next_Job) # Procedure defined in Table 3 above

    WHILE Open IS NOT EMPTY DO AND Coordinate ≠ ∅ DO:
      Next_Job.start_time := Time
      Next_Job.end_time := Time + Next_Job.duration
      DELETE Next_Job FROM Open
      ADD Next_Job TO In_Processing
      Next_Job := NEXT_ELEMENT(Open)
      Time := Next_Job.earliest_allowable_start_time
      Coordinate := PACK(Next_Job)
    END
  END
END
```

## RESULTS

We developed a Java implementation of this greedy spatial scheduling algorithm. We also developed an algorithm to generate random problem instances. For each job, the width and

height were generated using a uniform distribution over [1, width or height of area]. Durations were generated using a uniform distribution over [5, 25]. Earliest allowable start times and deadlines were generated using an incremented *current time* and a *tightness* factor

ranging from 1 to 10, with 10 generating the most tightly-packed problems. Earliest allowable start times were generated by

$$E = \text{current time} - r$$

where  $r$  is a random number uniformly distributed over  $[0, (\text{current time}) / \text{tightness}]$ . Deadlines were generated by

$$D = \text{current time} + \text{current job duration} + (r * \text{tightness})$$

where  $r$  is a random number uniformly distributed over  $[0, \text{current job duration}]$ . Finally,  $\text{current time}$  is initialized to 0 prior to the generation of any job and subsequently incremented after each job generation by

$$\text{Increment} = 10 * r / \text{tightness}$$

where  $r$  is a random number uniformly distributed over  $[0, \text{current job duration}]$ .

Ten generated problem instances of varying sizes were selected to be solved by the spatial scheduling algorithm, and the results are reported in Table 4 below. In each problem instance a width of 10 and a height of 7 were specified for the processing area. As can be seen, a higher tightness parameter results in a lower maximal deadline for a given number of jobs. Thus, a higher tightness parameter results in a more tightly-packed problem instance.

**Table 4:** The results for several generated problem instances

Problem Instance	Number of Jobs	Tightness	Maximal Deadline	Total Tardiness (Objective Function)	Computational Time (milliseconds)
E-100	100	3.3	2608	0	16
H-100	100	9.9	877	497	16
E-500	500	3.3	11,231	0	94
H-500	500	9.9	3640	51219	94
E-1000	1000	3.3	21,971	0	234
M-1000	1000	8.5	8393	132	312
H-1000	1000	9.9	7239	180,473	328
E-10000	10,000	3.3	222,491	0	10,172
M-10000	10,000	8.9	80,259	22,903	12,812
H-10000	10,000	9.9	73,381	3,426,391	13,062

## FUTURE WORK

Much literature exists for many types of scheduling and box-packing problems. There is relatively little literature, however, that directly addresses the topic of spatial scheduling. The work in this paper is quite preliminary in nature, and there are many aspects of this scheduling sub-discipline to be explored. Future theoretical work includes the development of mathematical models for different types of spatial scheduling problems as well as an analysis of the complexity of these problems. Future applied work includes the development of new

algorithms and heuristics that can provide reliably near-optimal solutions within a reasonable amount of time.

## REFERENCES

1. *Modeling and Solving the Spatial Block Scheduling Problem in a Shipbuilding Company*. Kyungchil Park, Kyungsik Lee, Sungsoo Park, Sunghwan Kim. 1996, Computers & Industrial Engineering 30, pp. 357-364.

2. *A Spatial Scheduling System and its Application to Shipbuilding: DAS-Curve*. **Kyoung Jun Lee, Jae Kyu Lee**. 1996, *Expert Systems with Applications* 10, pp. 311-324.

3. *Spatial Block Arrangement in Shipbuilding Industry Using Genetic Algorithm for Obtaining Solution for Anticipated Bottleneck*. **Ranjan Varghese, Duck Young Yoon**. Seoul, Korea : The International Society of Offshore and Polar Engineers, 2005. Proceedings of the Fifteenth International Offshore and Polar engineering Conference.

4. *A new exact method for the two-dimensional bin-packing problem*. **Francois Clautiaux**,

**Jacques Carlier, Aziz Moukrim**. 2007, *European Journal of Operations Research* 183, pp. 1196-1211.

5. *Two- and three-dimensional parametric packing*. **F.K. Miyazawa, Y. Yakabayashi**. 2005, *Computers and Operations Research* 34, pp. 2949-2603.

6. *Adjacent Resource Scheduling: Why Spatial Resources are so Hard to Incorporate*. **Jacob Jan Paulus, Johann Hurink**. 2006, *Electronic Notes on Discrete Mathematics* 25, pp. 113-116.

# Fundamental Travel Demand Model Example

Joel Hanssen

Old Dominion University

[joel.hanssen@yahoo.com](mailto:joel.hanssen@yahoo.com)

**Abstract.** Instances of transportation models are abundant and detailed “how to” instruction is available in the form of transportation software help documentation. The purpose of this paper is to look at the fundamental inputs required to build a transportation model by developing an example passenger travel demand model. The example model reduces the scale to a manageable size for the purpose of illustrating the data collection and analysis required before the first step of the model begins. This aspect of the model development would not reasonably be discussed in software help documentation (it is assumed the model developer comes prepared). Recommendations are derived from the example passenger travel demand model to suggest future work regarding the data collection and analysis required for a freight travel demand model.

## INTRODUCTION

In simple terms of supply and demand, the transportation network is the supply and the need to travel is the demand. The complexity of the transportation system is due to the fact that the demand has a multipurpose and the supply is inherently multimodal. The model of such a system attempts to forecast the travel demand of people by automobile (single occupancy or high occupancy) and by public transit (bus, subway, or intercity rail) when they choose not to travel by walking, biking, or airplane [1]. A second category of travel demand is freight. The supply is multimodal (ship, rail, and truck) and the demand has a multipurpose (each commodity is a purpose). The objective of this paper is to look at the inputs required and illustrate the basic process for modeling the first category of travel demand. Models of large, complex geographical regions already exist. These examples are useful for illustrating simulation techniques. The example presented in this paper is smaller in scale and allows a more manageable study of the supporting data and the development of the model itself using Cube software by Citilabs.

## GENERATION

Trip generation is a data collection and analysis process that is used to estimate the number of trips each region will produce and attract. In order to accomplish this, the region is divided into zones and zonal data is collected. Next, a trip survey assesses the travel habits of the population. Finally, regression analysis provides coefficients for trip purpose equations which are used to calculate the number of trips produced and attracted by each zone.

## 1 - Zones and Zonal Data

In this example, the study area is divided into three regions: city, suburb, and rural. These regions are subdivided into smaller traffic analysis zones (TAZs). The sizes of TAZs are typically driven by population, hence smaller and more numerous TAZs in the city region. Each TAZ has a homogeneous land use and can be described by only a few variables. In this example, these variables are the number of households (HH), the total population, the area in square miles (SQ\_MI), the mean household size (MEANHHSIZE), a factor which represents the density of the population (DENTHOU), the number of people that work in the area regardless of whether they live there (TOTAL\_EMP), and the mean number of vehicles per household (VEH\_HH). Typically, a metropolitan planning organization (MPO) collects zonal data for a study area. The study area in this example is heuristically developed using the 2007 Data Book compiled by the Hampton Roads Planning District Commission. Variables of interest in the Data Book include population, area in square miles, mean household size, total employment, and total number of vehicles registered. Additional variables are required to translate the characteristics of real world zones to the zones in the example: density (Equation 1.1), an employment factor expressing the number of people working in terms of area (Equation 1.2), the number of households (Equation 1.3), and the mean number of vehicles per household (Equation 1.4).

$$\text{DENTHOU} = \frac{\text{POPULATION}}{1000 \times \text{SQ\_MI}} \quad (1.1)$$

$$\text{EMP\_SQMI} = \frac{\text{EMP}}{\text{SQ\_MI}} \quad (1.2)$$

$$\text{HH} = \frac{\text{POPULATION}}{\text{MEANHHSIZE}} \quad (1.3)$$

$$\text{VEH\_HH} = \frac{\text{NUM\_VEH}}{\text{HH}} \quad (1.4)$$

The map is drawn to scale using ArcMap and the area of each zone is determined. Based on the author's familiarity with the Hampton Roads region, seven real world areas are selected to represent the characteristics of various area types. Based on each example zone's area type (i.e. rural, suburb, or city), the following real world variables are assigned: DENTHOU, MEANHHSIZE, EMP/SQ\_MI, and VEH\_HH. The following additional variables were calculated: population (Equation 2.1), the number of households (Equation 1.3), and total employment (Equation 2.2).

$$\text{POPULATION} = (\text{DENTHOU} * 1000) * \text{SQ\_MI} \quad (2.1)$$

$$\text{TOTAL\_EMP} = \text{EMP\_SQMI} * \text{SQ\_MI} \quad (2.2)$$

## 2 - Trip Survey

A relationship must be established between these variables and the number of trips. The number of trips is determined by circulating a trip survey to a representative sample of the population. A typical survey collects information about each household (number of people, number of vehicles, etc) and then asks them to keep a trip diary. Simple linear regression analysis determines the relationship between the dependent variable (number of trips) and the independent socioeconomic variables. The example model surveyed 3% of the population. The number of surveys collected from each zone is proportional to the fraction of the total population living in that zone in the model. As additional constraints: the average number of vehicles per household in the survey closely matched the VEH\_HH for the corresponding zone in the model, while the average number of people per household in the survey closely matched the MEANHHSIZE size for corresponding zone in the model. The dependent variable (number of trips) is manipulated to create a significant relationship with VEH\_HH and MEANHHSIZE (Figure 1).

## 3 - Trip Purpose Equations

In this example, there is only one trip purpose. The trips are home-based-work (HBW) and are produced according to the availability of a car and the number of people in a household factored by socioeconomic data that indicates 60% of the population is part of the work force; represented by Equation (3). The attraction for these HBW trips is a factor of the employment in a particular zone; represented by Equation (4).

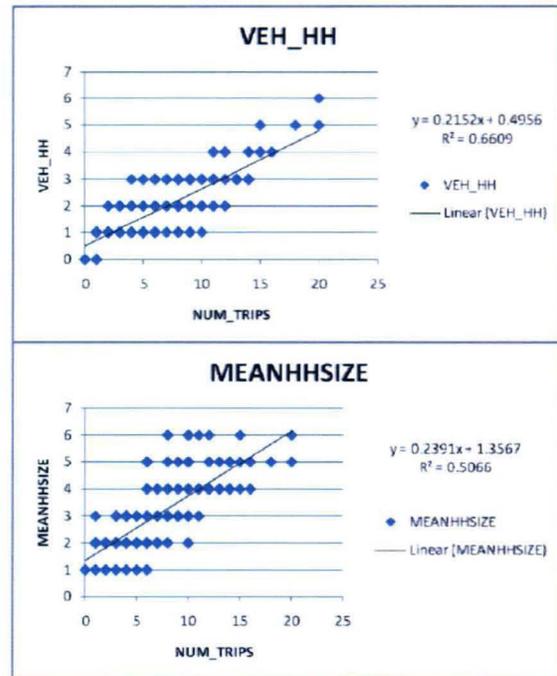


Figure 1. Regression Analysis

$$((0.4956 + 0.2152 * \text{VEH\_HH}) + ((1.3567 + 0.2391 * \text{MEANHHSIZE}) * 0.6) * \text{HH}) \quad (3)$$

$$(\text{TOTAL\_EMP} / 360342) \quad (4)$$

## 4 - Generation Table

The result of the Generation program in Cube is a simple text file. The first column contains the zone number, each subsequent pair of columns contains the values for the number of trips produced and attracted for each trip purpose. In this example there is only one trip purpose (Table 1).

Table 1: Generation

1	8901	5707
2	38423	70290
3	41518	37644
4	90411	90534
5	60464	44518
6	8104	4449
7	36533	31213

## DISTRIBUTION

Trip distribution is the process used to estimate how the trips that each zone will produce and attract are distributed across the other zones. The required inputs are the trip generation table, a level of service (LOS) matrix, and a friction factor table.

## 1 - Level Of Service

This matrix is obtained through a process called skimming. During skimming, the Cube Highway program extracts the costs (i.e. time and distance) of journeys between zones. In order to do this, the highway program requires a highway network file. A network is built from a zonal map (polygon shape file) developed for trip generation and a highway map (line shape file). For this example, a highway map was developed with several speed classifications represented by two digit codes. The first digit represents facility type (i.e. freeway, major arterial, or minor arterial). The second digit represents the area type (i.e. rural, suburban, or city). Cube combines the two maps and calculates the centroid for each zone and builds centroid connectors to the highway. This "new" map is the highway network.

## 2 - Friction Factors

Travel function values are contained in a lookup file. The lookup file is formatted as a series of curves; one for each trip purpose. In this example the friction values increase with time until a trip of greater time than 60 minutes will not be distributed.

## 3 - Origin/Destination Matrix

Utilizing these three inputs, the Cube Distribution program produces the trip Origin/Destination matrix.

## ASSIGNMENT

Trip assignment is the process used to estimate the routes that will be used to travel from origin to destination. The inputs required are the Origin /Destination matrix file and the Highway network file. Also a cost function should be specified. In this example the cost function gives a lesser weight to DISTANCE (see Equation 5). The result is illustrated by the map in figure 2. Note that the high congestion is in the city and low congestion in the rural area. Because this example does not include external trips; the links leading out of the region are unused. This may also explain the low congestion on the freeway.

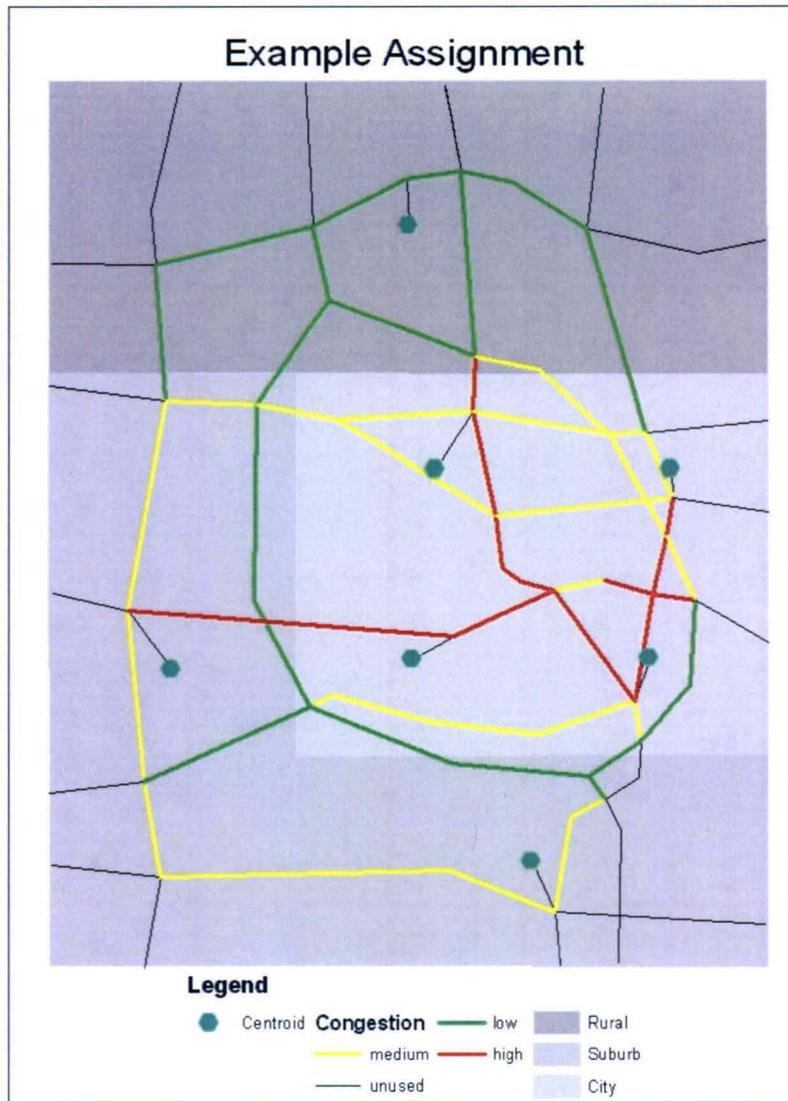
$$\text{TIME}^*1 + \text{LI.DISTANCE}^*0.3 \quad (5)$$

## CONCLUSION

It should be apparent from this example that the greatest preparation for the model precedes the first step: trip generation. Zones must be defined, zonal data collected, trip counts determined, and finally, a relationship must be established between some socioeconomic variable or land use variable and the number of trips. Each trip purpose will have a unique relationship. Although the example in this paper does not include a public transit network, it is desirable to model the public transportation system with a fourth step and simulate the choice of people to utilize that alternative. Together, these four steps (Trip Generation, Trip Distribution, Mode Choice, and Traffic Assignment) are referred to as the "four-step model" [2]. It is also desirable to model the freight transportation system and forecast the travel demand of freight. The choice of freight to compete with people on the highway network is of particular interest.

## FUTURE WORK

The premise for freight-travel-demand is fundamentally different from people-travel-demand: the purpose of the trip is to move a commodity from origin to destination. The concept of a trip survey would be very different and may not even be applicable. The origins and destinations of commodities encompass a region of far greater scale. It is noteworthy that TAZs are not valid for freight trip generation. Freight Analysis Zones (FAZs) must be developed and the corresponding zonal data must be determined [3]. The trip generation of FAZs would be followed by trip distribution. Mode split would simulate the choice of freight to compete with people on the highway network. Because commodity trips may utilize the highway network for a sub-link of the trip, a model of the freight system would need to include logistic nodes to simulate mode splits. Finally, trips that are coarsely distributed across FAZs must be finely distributed across TAZs in the region of interest, followed by simple all-or-nothing assignment. Again, the portion of trips utilizing the highway network would be of particular interest. Future work needs to be done to find variables that have significant relationships to commodities (land use, socioeconomic, or some hybrid). It is important to know what data is needed and to collect the right data, before attempting to model freight travel demand.



**Figure 2.** Assignment

#### REFERENCES

1. Fontaine, M. (2009) "Transportation Modeling and Simulation: Real World Examples", *Principles of Modeling and Simulation*, A John Wiley & Sons, New Jersey, pp. 181-198.
3. Harris, G. A. et al (2009) "Developing Freight Analysis Zones at State Level: Cluster Analysis Approach", *Transportation Research Board*, Washington, D.C.
2. Martin, W. A.; McGuckin, N. A. (1998) "NCHRP Report 365: Travel Estimation Techniques for Urban Planning", *National Cooperative Highway Research Program*, Transportation Research Board, Washington, D.C.
4. (2008) "Cube Help", Cube Reference Guide, Citilabs America, San Francisco.
5. (2008) "Hampton Roads Data Book", Hampton Roads Planning District Commission, Virginia.

# Visualization of Traffic Accidents

Jie Wang; Yuzhong Shen; Asad Khattak

Virginia Modeling, Analysis, and Simulation Center, Old Dominion University  
{jwang016, yshen, akhattak}@odu.edu

**Abstract:** Traffic accidents have tremendous impact on society. Annually approximately 6.4 million vehicle accidents are reported by police in the US and nearly half of them result in catastrophic injuries. Visualizations of traffic accidents using geographic information systems (GIS) greatly facilitate handling and analysis of traffic accidents in many aspects. Environmental Systems Research Institute (ESRI), Inc. is the world leader in GIS research and development. ArcGIS, a software package developed by ESRI, has the capabilities to display events associated with a road network, such as accident locations, and pavement quality. But when event locations related to a road network are processed, the existing algorithm used by ArcGIS does not utilize all the information related to the routes of the road network and produces erroneous visualization results of event locations. This software bug causes serious problems for applications in which accurate location information is critical for emergency responses, such as traffic accidents. This paper aims to address this problem and proposes an improved method that utilizes all relevant information of traffic accidents, namely, route number, direction, and mile post, and extracts correct event locations for accurate traffic accident visualization and analysis. The proposed method generates a new shape file for traffic accidents and displays them on top of the existing road network in ArcGIS. Visualization of traffic accidents along Hampton Roads Bridge Tunnel is included to demonstrate the effectiveness of the proposed method.

## 1. Introduction

Driving is an essential part of day-to-day life for most people. However, with so many people on the road with varied physical and psychological conditions, driving an automobile is usually a risky and dangerous activity. According to the report released by police, annually there are approximately 6.4 million vehicle accidents in the US and nearly half of them result in catastrophic injuries [1]. Besides various precautions against traffic accidents, prompt and efficient traffic accident handling is very important for saving lives and reducing the adverse impact caused by traffic accidents.

Geographic information systems (GIS) are computer-based systems used to store, manipulate, display, and analyze geographic information. Various GIS software tools have been widely used in transportation research and managements for many years due to their powerful geo-visualization capability, analytical capability and database management features [2]. Visualization of traffic information greatly improved the performance and efficiency of many traffic related applications such as traffic supervision, traffic scheduling, and so on. Among these applications, transportation safety can significantly benefit from the use of GIS [3]. Visualization of traffic accidents using GIS can greatly facilitate handling and analysis of traffic accidents in many

aspects. The locations of traffic accidents are especially critical for rapid response of traffic accidents. Visualizations of traffic accidents over a large area can provide useful statistics and insight into the condition and state of the transportation network.

Environmental Systems Research Institute (ESRI), Inc. is the world leader in GIS software [4]. ArcGIS, a software package developed by ESRI, has the capabilities to display events associated with road networks. But during processing of the event locations related to the road network, the existing algorithm used in ArcGIS ignores the route direction information of the location and results in erroneous visualizations of the event locations. For example, the algorithm used in ArcGIS considers the route I-64 E and route I-64 W as the same route although in fact they are two separate routes. This software bug causes serious problems for applications in which accurate location information is critical for emergency responses, such as traffic accidents. In this paper, we describe a method that can extract correct event locations and create accurate visualizations of traffic accidents. The proposed method generates a new shape file for traffic accidents and displays them on top of the existing road network in ArcGIS.

The remainder of this paper is organized as follows. Section 2 introduces the method used by

ArcGIS that produces erroneous results. Section 3 discusses the proposed method and its results and compares them with those produced by ArcGIS. Finally conclusions are drawn in Section 4.

## 2. ArcGIS Algorithm

Route events can be added to ArcGIS in two ways: 1) using the *Make Route Event Layer* geo-processing tool; or 2) accessing the *Add Route Events* wizard from the ArcMap tool menu [3]. In ArcGIS, information about route events is stored in an associated table. Route is a link feature class with an identifier field for every record, which can be used as a reference in the event table. The route information is stored in a shape file that has associated attributes. To display route events on a road network, parameters of the relationship between the table storing events and the shape file that stores the routes should be defined first. The event table should also contain a field that indicates the measure or the precise location of the event. Figure 1 shows the graphical user interface in ArcGIS that is used to select the route file, route identifier, event table and measure field. In this paper traffic accident data provided by the Virginia Department of Transportation is utilized to illustrate and compare the results produced by ArcGIS and the proposed method.

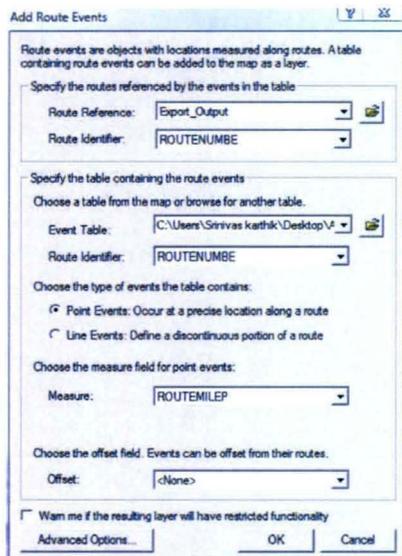


Figure 1: Interface for adding route events in ArcGIS

In Figure 1, the drop-down box *Route Reference* is used to select the shape file that contains the road network, including all routes. The drop-down box *Route Identifier* is used to specify the attribute of the road network shape file that identifies the route, which happens to be *ROUTENUMBER* in the example shown in Figure 1. The file containing the route events is then selected via the drop-down box *Event Table*. The same file should also contain a field for identifying the route. As shown in Fig. 1, route events can be either point type or line type. For traffic accidents, the event type is point type representing the traffic accident locations. A *Measure* is needed to specify the locations of point events and in this case *RouteMilePost* is used to specify the traffic accident locations, which is the distance from the starting point of the route. ArcGIS only makes use of the attributes specified in Table 1 and Table 2 for visualization of route accidents. It can be seen that the directions of the routes are ignored by ArcGIS, e.g., routes I-64 East and I-64 West are treated as the same route. Thus accidents are placed randomly along either direction of a route depending on the record location in the binary shape file. In the example shown in Figure 3, all traffic accidents are erroneously placed on I-64 East.

Table 1: Event table utilized by ArcGIS

Object ID	Route Number	--	Route Mile Post
0	0064	--	264
1	00264	--	249
2	00664	--	259
3	0064	--	294
4	00265	--	14
5	0095	--	4
6	0064	--	264

Table 2: Route Table Utilized By ArcGIS

FID	Shape	Object ID	--	Route Number
0	Polyline M	0	--	0064
1	Polyline M	1	--	00264
2	Polyline M	2	--	00664
3	Polyline M	3	--	0064
4	Polyline M	4	--	00265
5	Polyline M	5	--	0095
6	Polyline M	6	--	0064

### 3. Proposed Method

To address the problems in ArcGIS for traffic accident visualization, we developed a program that utilizes all the relevant information in the route table and event table for identifying accident locations. Specifically, the attribute *Route Suffix* in both tables are utilized, which indicates the directions of the routes. Some attributes of the event table and route table are shown in Tables 3 and 4, respectively. The block diagram of the proposed method is shown in Figure 2.

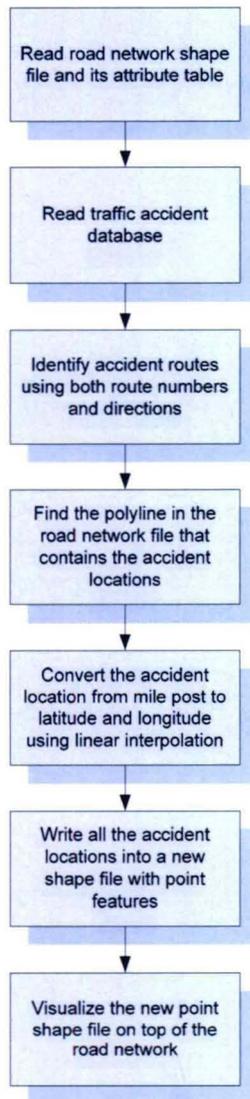


Figure 2: Block diagram of the proposed method

The proposed method first read the input files, including the shape file containing the road network, its associated attribute table, and the traffic accident database. Then it identifies traffic accident routes using both route numbers and directions, i.e., the attribute *Route Suffix*. In a shape file that contains a road network, each polyline is a part of roadway that consists of a series of vertices. The minimum and maximum mile posts of a polyline are stored as attributes of that polyline. Thus, based on the attribute *RouteMilePost* of the accident location, the polyline in the road network file that contains the accident location can be found by comparing *RouteMilePost* and the minimum and maximum mile posts of the polyline. After the polyline containing the accident location is identified, we need to further to find the two consecutive vertices of that polyline between which the accident is located. Each vertex of the polyline in a road network has three attributes: latitude, longitude, and mile post. The latitude and longitude of the accident location is computed as linear interpolation of two vertices based on the mile posts. Finally, the accident locations are written into a new shape file with point features and it is imported into ArcGIS and visualized on top of the original road network. Figure 4 shows the results generated by the proposed method. It can be seen that traffic accidents are correctly displayed along both directions of Interstate I-64. Figures 5 and 6 show enlarged views of Figure 3 and 4. It can be clearly seen that the proposed method corrected the error in ArcGIS.

The implementation of the proposed method utilized the Shapefile C Library [5], which provides the ability to read, write, and update ESRI shape files (.shp) and associated attribute files (.dbf).

Table 3: Event table utilized in the proposed approach

Object ID	Route Number	--	Route Mile Post	Route Suffix
0	0064	--	264	W
1	00264	--	249	E
2	00664	--	259	W
3	0064	--	294	E
4	00265	--	14	W
5	0095	--	4	N
6	0064	--	264	W

**Table 4:** Route table utilized in the proposed approach

FID	Shape	Object ID	--	Route Number	Route Suffix
0	Polyline M	0	--	0064	W
1	Polyline M	1	--	00264	E
2	Polyline M	2	--	00664	W
3	Polyline M	3	--	0064	E
4	Polyline M	4	--	00265	W
5	Polyline M	5	--	0095	N

#### 4. Conclusions

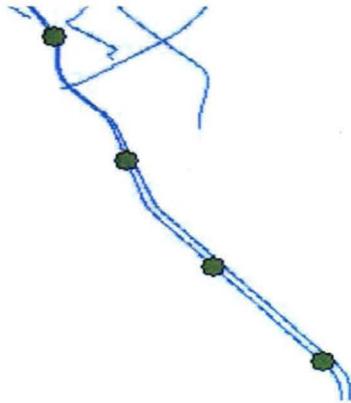
Visualization of traffic accidents can greatly facilitate handling and analysis of traffic accidents. ArcGIS is the leading GIS solution provider and its software ArcGIS supports visualization of route events, such as traffic accidents. However, the default algorithm of ArcGIS does not consider route directions when displaying route events, and thus producing erroneous visualizations. This paper proposed a method that utilizes all relevant information of traffic accidents, namely, route number, direction, and mile post, and extracts correct event locations for accurate traffic accident visualization and analysis. The proposed method generates a new shape file for traffic accidents and displays them on top of the existing road network in ArcGIS. Included visualization of traffic accidents along Hampton Roads Bridge Tunnel demonstrated the effectiveness of the proposed method.



**Figure 3:** Traffic accidents displayed along HR Bridge Tunnel (I-64) generated by ArcGIS. All the accidents are erroneously placed along I-64 East because ArcGIS discards the route directions.



**Figure 4:** Visualization of traffic accidents by the proposed method. Traffic accidents are placed correctly along both directions of I-64.



**Figure 5:** Enlarged view of traffic accidents displayed along HR Bridge Tunnel (I-64 East) generated by ArcGIS. All accidents are placed along I-64 East.

**REFERENCES**

1. National Highway Traffic Safety Administration. Traffic Safety Facts, 2000: Overview. DOT HS 809 329. Washington, DC: Department of Transportation, National Highway Traffic Safety Administration.
2. Thill, J.C., (2000) Geographic information systems for transportation in perspective. *Transportation Research Part C*, 8(1-6): p. 3-12.



**Figure 6:** Enlarged view of traffic accidents produced by the proposed method. Accidents are placed correctly along both I-64 East and I-64 West.

3. Kim, K. & N. Levine, (1996) Using GIS to improve highway safety. *Computers, Environment and Urban Systems*. 20(4-5): p. 289-302.
4. ESRI, (2009) <http://www.esri.com/>.
5. Shapefile C Library V1.2, (2009) <http://shapelib.maptools.org/>.



# High-Fidelity Roadway Modeling and Simulation

Jie Wang<sup>1</sup>; Yiannis Papelis<sup>1</sup>; Yuzhong Shen<sup>1</sup>; Ozhan Unal<sup>2</sup>; Mecit Cetin<sup>2</sup>

<sup>1</sup>Virginia Modeling, Analysis, and Simulation Center, Old Dominion University

<sup>2</sup>Dept. of Civil and Environmental Engineering, Old Dominion University

{jwang016, ypapelis, yshen, onun001, mcetin}@odu.edu

**Abstract:** Roads are an essential feature in our daily lives. With the advances in computing technologies, 2D and 3D road models are employed in many applications, such as computer games and virtual environments. Traditional road models were generated by professional artists manually using modeling software tools such as Maya and 3ds Max. This approach requires both highly specialized and sophisticated skills and massive manual labor. Automatic road generation based on procedural modeling can create road models using specially designed computer algorithms or procedures, reducing the tedious manual editing needed for road modeling dramatically. But most existing procedural modeling methods for road generation put emphasis on the visual effects of the generated roads, not the geometrical and architectural fidelity. This limitation seriously restricts the applicability of the generated road models. To address this problem, this paper proposes a high-fidelity roadway generation method that takes into account road design principles practiced by civil engineering professionals, and as a result, the generated roads can support not only general applications such as games and simulations in which roads are used as 3D assets, but also demanding civil engineering applications, which requires accurate geometrical models of roads. The inputs to the proposed method include road specifications, civil engineering road design rules, terrain information, and surrounding environment. Then the proposed method generates in real time 3D roads that have both high visual and geometrical fidelities. This paper discusses in details the procedures that convert 2D roads specified in shape files into 3D roads and civil engineering road design principles. The proposed method can be used in many applications that have stringent requirements on high precision 3D models, such as driving simulations and road design prototyping. Preliminary results demonstrate the effectiveness of the proposed method.

## 1. Introduction

Roads are an essential component of civilization and many different types of roads exist, e.g., highways, freeways, expressways, arterial streets, and rural roads. With the advances in computing technologies, road models are widely used in many applications such as games and virtual environments. In early computer applications involving road models, almost all 2D and 3D road models were generated by professional artists manually using modeling software tools such as Creator, 3ds Max, and Maya. Good results can be achieved through manual road modeling at the expense of extensive labor and time.

Various methods have been proposed to automate the road modeling process. Procedural modeling methods allow automatic generation of objects according to specifically designed procedures and they have been utilized for automatic road generation. Parish and Muller introduced the L-system for natural phenomena modeling in their procedural modeling software CityEngine for road generation, building construction, and building

face creation [1]. Sun et al. extracted four kinds of common road patterns (Population-Based, Raster, Radial, and Mixed) from existing road networks and generated road networks for virtual city modeling based on these pattern templates [2, 3]. Watson et al. used an agent-based technique to generate road network as a part of virtual cities generation [4, 5]. Glass et al. attempted to combine various existing procedural techniques including Voronoi diagrams, subdivision, and L-system to replicate the identified features of road patterns in south African informal settlements [6]. Chen et al. put forward a tensor field based procedural method to model the street networks of large urban area interactively [7]. However, all these methods focused on the creation of artificial road networks for virtual environments to achieve visual satisfaction; the detailed road geometric design and structure based on civil engineering principles were not addressed in these methods.

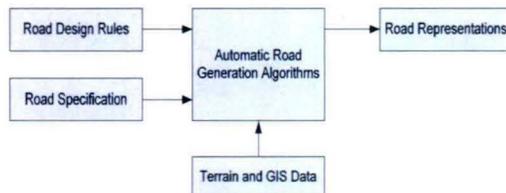
Automatically generating high-fidelity roadways from real terrain and GIS data with special focus on combining procedural modeling and civil engineering principles is a very challenging problem but with many applications, such as,

such as driving simulation and road design prototyping. To this end, this paper proposes a method that takes into account road design principles practiced by civil engineering professionals, and as a result, the generated roads can support not only general applications such as games and simulations in which roads are used as 3D assets, but also demanding civil engineering applications, which require accurate geometrical models of roads. The inputs to the proposed method include road specifications, civil engineering road design rules, terrain information, and surrounding environment. Then the proposed method generates in real time 3D roads that have both high visual and geometrical fidelities. This paper discusses in details the procedures that convert 2D roads specified in shape files into 3D roads as well as civil engineering road design principles.

The remainder of the paper is organized as follows. Section 2 describes the system design of the proposed road generation method. Section 3 discusses the implementation of the system design in detail using the generation of a parametric road curve as an example. Section 4 shows some experimental results. Lastly, Section 5 concludes this paper and discusses future research directions.

## 2. System Design

The major goal of this work is to automatically generate high-fidelity roadways from real geographical information, satisfying civil engineering requirements. The overall structure of the proposed method is illustrated in Figure 1. Inputs to the proposed method include road design rules, road specification, and terrain and GIS data and the outputs are road representations produced by the proposed method.

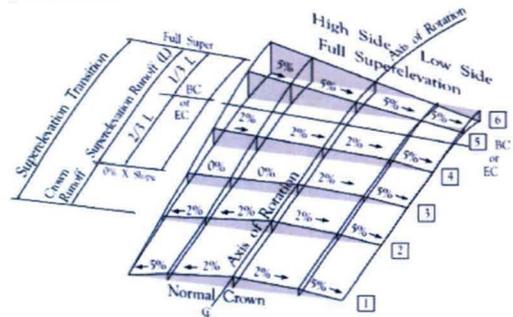


**Figure 1:** Structure of the proposed automatic road generation system

Road design rules vary with countries and states. Thus, based on the target area for which the roads are to be generated, appropriate design rules should apply. For

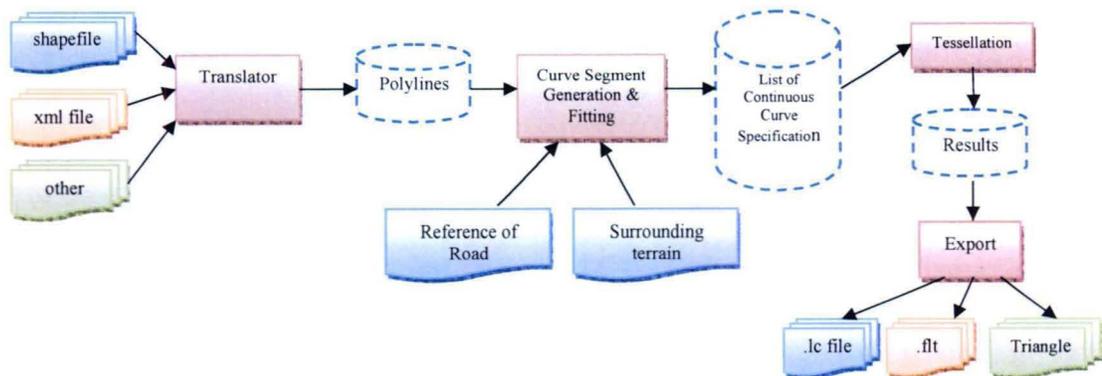
roads in the US, the US Department of Transportation (DOT) Federal Highway Administration developed Federal Lands Highway Project Development and Design Manual [8]. In addition, state DOT's also provide highway design manuals and guidelines. These road design manuals describe every aspect of road design, such as road cross sections, interchanges, maximum and minimum grades, overbridges, underbridges, abutment, merges and diverges, roundabouts, and turn radius. Clearly it is not necessary and infeasible to use all design rules in our automatic road generation system for modeling and simulation applications. As a result, it is necessary to determine a minimum set of design rules for our automatic road generation system.

The road specification describes the requirements and features of the road to be generated, including items such as road functional class (local, connector, arterial, freeways), number of lanes, speed limit, lane width, road signs, road environment (urban or rural), surface type, interchanges types, curvature, visibility, and superelevation. These parameters of road specification define the requirements for the items included in the minimum set of design rules. For example, Figure 2 illustrates basic elements for describing a superelevation transition [9]. Terrain and GIS data can complement the parameters in the road specification. For example, the GIS data may provide the shape file for road centerlines as well as the number of lanes. Other GIS data such as land use and satellite images can also be useful for road generation.



**Figure 2:** Elements of a superelevation transition [9]

At the core of the proposed method is the automatic road generation component, which consists of two major parts: data structures and algorithms.



**Figure 3:** Functional blocks in the proposed method

Data structure will be developed to represent the road design rules, road specifications, terrain, GIS information, as well as the roads to be generated. In addition, different levels of representations can be utilized. For example, two different representations of the generated roads can be utilized: high level representation and low level representation. High level representation stores the road topology (links and nodes), attributes of each link (number of lanes, speed limit, length, etc.), and low level representation can be a triangular mesh of a short road segment for high-fidelity driving simulator applications. The automatic road generation algorithms produce the roads based on the road design rules and road specifications. Different algorithms will be developed to generate different road features, such as road surfaces, interchanges, merges and diverges, and traffic signs. A hierarchy of algorithms will be developed. For example, the top level (class) of the hierarchy will generate the composition of the entire road (route), while the middle and low classes will generate the details of each component. Other classes will produce the final outputs of the generated roads. Procedural methods such as the L-system will be investigated and used if necessary. The proposed algorithms also need to adjust the terrain so that the generated roads fit their surrounding environments well.

### 3. Implementation

The whole road generation process is shown in Figure 3. First road specifications, in the form of various file formats, e.g., shape file and XML file, are transformed by a translator into polylines, which represent the coarse position of the road centerline and will be used as road prototype. These polylines are then refined and

optimized into road segments in Segment Generation & Fitting based on reference of road, namely, road design rules and additional road specifications, and surrounding terrain. All of these road segments are represented by a set of parameters and then tessellated into different levels of discrete road representations and finally can be exported into various file formats. In the following, generation of a road curve will be used as a typical example to illustrate the implementation and some preliminary results will be shown in section 4.

#### 3.1 Segment Generation & Fitting

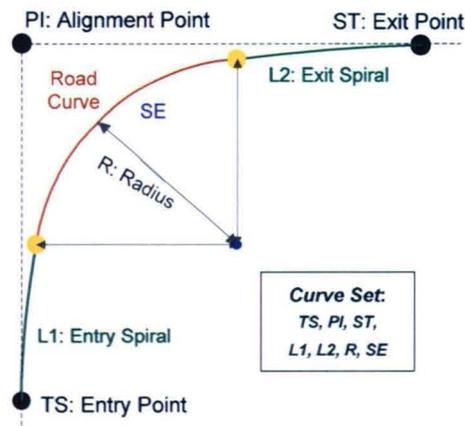
Among all the functional blocks in the proposed method, the segment generation is one of the focal points. The input, output, and main functions of this functional block are described as follows:

**Input:** The input of this process is a polyline representation obtained from a translator which reads and translates various file formats, such as shape files and XML files.

**Output:** The output of this process is parametric representations of road segments. An example of the parametric specification includes entry point(TS), alignment point(PI), exit point(ST), radius(R), length of entry spiral (L1), length of exit spiral(L2), and superelevation (SE), as illustrated in Figure 4.

To design a road, civil engineering professionals make use of a set of parameters to specify the road geometry, such as superelevation, spiral length, and radius. To generate high-fidelity roads that comply with civil engineering principles, such parameters are generated by the proposed method. The polyline road representations in the shape files contain a series of points or vertices for each

road segment, which is defined by connecting the consecutive points or vertices. Thus polyline road representations are linear approximations of ideal road curves. To extract the road curve parameters from the polyline representations, the proposed method first divides the polylines into different segments through a segmentation process and then finds parametric representations of each segment using the least square method.



**Figure 4:** An example of the parametric representation of a curved road segment. A road curve contains three point positions, which are the entry point (TS), alignment point (PI) and exit point (ST), two length values, which are entry spiral (L1) and exit spiral (L2), a radius (R), and the superelevation (SE).

### 3.1.1 Road Polyline Segmentation

In order to divide the road polylines into different segments, we identify three types of critical points which are used to segment the road polylines. The critical points are defined based on their geometrical features as follows.

- Sharp turn: The angle between the adjacent line sections of this point exceeds a predefined threshold, resulting in a sharp turn when driving or passing through this point.
- S-turn: Curvature can be calculated for each point of the polyline and the curvature of each point has a radius. A tangent circle of each point is a circle with the same radius as that of the curvature. The center of the tangent circle can be on either left or right side of the polyline. S-turn is such a point at which the center of its tangent circle changes from one side of the polyline to the other side.

- Turn start/end: The polyline representations of roads tend to have dense points for road segments with large curvature and sparse points for relatively straight road segments. Thus the lengths of two line segments at a point can be used to indicate a transition from a straight line to a curve, or vice versa. The proposed method calculates the ratio of the lengths of the two adjacent line segments at a point and compares it with a predefined threshold. If the ratio is greater than the threshold, that point is the start or end of curved road segments.

### 3.1.2 Parametric Representations Using Least Square Fitting

In this section, we present the results for fitting a set of points  $(x_i, y_i)$ ,  $i = 1, \dots, N$  into a line or circle using the least square method. The derivations are omitted due to space limit.

- Line Fitting

A line can be represented by the equation  $y = ax + b$ . The parameters  $a$  and  $b$  can be calculated using the least square method as follows.

$$a = \frac{\sum x_i y_i \sum x_i - \sum y_i \sum x_i^2}{(\sum x_i)^2 - N \sum x_i^2},$$

$$b = \frac{\sum x_i \sum y_i - N \sum x_i y_i}{(\sum x_i)^2 - N \sum x_i^2}$$

- Circle Fitting

A circle can be determined by its center  $(A, B)$  and radius  $R$ . The parameters  $A$ ,  $B$ , and  $R$  can be calculated using the least square method as follows.

$$A = \frac{a}{-2}, B = \frac{b}{-2} \text{ and } R = \frac{1}{2} \sqrt{a^2 + b^2 - 4c}$$

where

$$a = \frac{HD - EG}{CG - D^2}, b = \frac{HC - ED}{D^2 - GC},$$

$$c = -\frac{\sum(x_i^2 + y_i^2) + a \sum x_i + b \sum y_i}{N},$$

and

$$C = (N \sum x_i^2 - \sum x_i \sum y_i),$$

$$D = (N \sum x_i y_i - \sum x_i \sum y_i),$$

$$E = (N \sum x_i^3 + N \sum x_i y_i^2 - \sum (x_i^2 + y_i^2) \sum x_i),$$

$$G = (N \sum y_i^2 - \sum y_i \sum y_i),$$

$$H = N \sum x_i^2 y_i + N \sum y_i^3 - \sum (x_i^2 + y_i^2) \sum y_i.$$

### 3.2 Tessellation

This section mainly discusses parametric representations of curved road segments. A

spiral curve is a curve of constantly changing radius. Spiral curves are generally used to provide a gradual transition in curvature from a straight section of road to a curved section. Figure 5 shows the components of a spiral curve. Spiral curves are necessary on high-speed roads from the standpoint of comfortable operation and gradually bringing about the full superelevation of the curves.

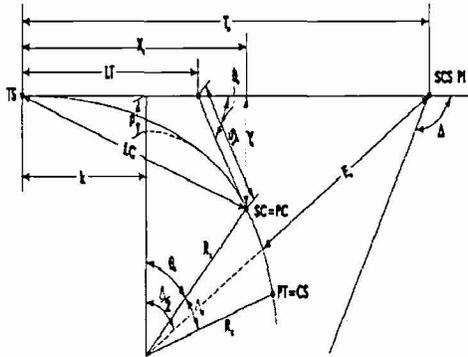


Figure 5: Components of a spiral curve

#### Definitions

- SCS PI = Point of intersection of main tangents.
- TS = Point of change from tangent to spiral curve; SC = Point of change from spiral curve to circular curve; CS = Point of change from circular curve to spiral curve; ST = Point of change from spiral curve to tangent.
- LC = Long chord; LT = Long tangent; ST = Short tangent.
- PC = Point of curvature for the adjoining circular curve; PT = Point of tangency for the adjoining circular curve.
- $T_s$  = Tangent distance from TS to SCS PI or ST to SCS PI.
- $E_s$  = External distance from the SCS PI to the center of the circular curve.
- $R_c$  = Radius of the adjoining circular curve.
- $\theta_s$  = Central (or spiral) angle of arc  $l_s$ .
- $\Delta$  = Total central angle of the circular curve from TS to ST.
- $\Delta_c$  = Central angle of circular curve of length  $L$  extending from SC to CS.
- $k$  = Abscissa of the distance between the shifted PC and TS.
- $Y_c$  = Tangent offset at the SC;  $X_c$  = Tangent distance at the SC;  $x$  and  $y$  = coordinates of any point on the spiral from the TS.

#### 4. Experimental Results

#### 4.1 Segmentation and Fitting Results

Figure 6 shows an example of road segmentation and fitting using part of Interstate I-64 as an example. Figure 6(a) displays the original polylines imported from a shape file. Figure 6(b) shows the results after segmentation. Polyline are divided into several segments separated by critical points: red points for curve segment and black points for straight line segments. Figure 6(c) shows the results line and circle fitting using the least square method. Two circles with radii of 1433.1m and 686.7m are extracted from points on the curved parts and three straight lines are extracted from points on the line segments. It can be seen the proposed segmentation and fitting methods are very effective, producing parametric representations of the road segments that fit very well the original polyline representations.

#### 4.2 Tessellation Results

This part uses horizontal alignment as an example. In order to create the alignment of the road surface, one needs to know the coordinates of centerline at any arbitrary intervals. Usually, these intervals are 25 ft or 50 ft. Design of a spiral curve begins by specifying the following key parameters.

- 1)  $R$  or  $D$  (degree of curve) is given by design considerations (limited by design speed).
- 2)  $L_c$  is chosen with respect to design speed and the number of traffic lanes.

In addition to these parameters,  $\Delta$ , Chainage of PI, and TS, are determined in the field. All other spiral parameters can be determined by computation and/or by use of spiral tables.

**Example:** A simple curve is to fit a road curve with a radius of 1000 ft for a D.S. of 50 mph. The intersection angle  $\Delta = 30^\circ$ . Determine the spiral offset ( $y$ ) from TS to SC in  $\frac{1}{2}$  station intervals (50 ft) if station of TS = 150+00.

#### Length of the spiral:

$$l_c = \frac{1.58V^3}{R}, X_c \cong l_c, y_c = \frac{l_c^2}{6R}, D = \frac{5730}{R}, \theta_s = \frac{l_c \times D}{200}$$

$$p = y_s - R_c(1 - \cos\theta_s), E_s = (R_c + p) \times \sec\left(\frac{\Delta}{2}\right) - R_c,$$

$$X_c = \sqrt[3]{y_c \times 6 \times R_c \times l_c}, k = X_c - R_c \sin\theta_s,$$

$$T_s = (R_c + p) \tan\left(\frac{\Delta}{2}\right) - k, y = \frac{x^3}{6R_c l_c}$$

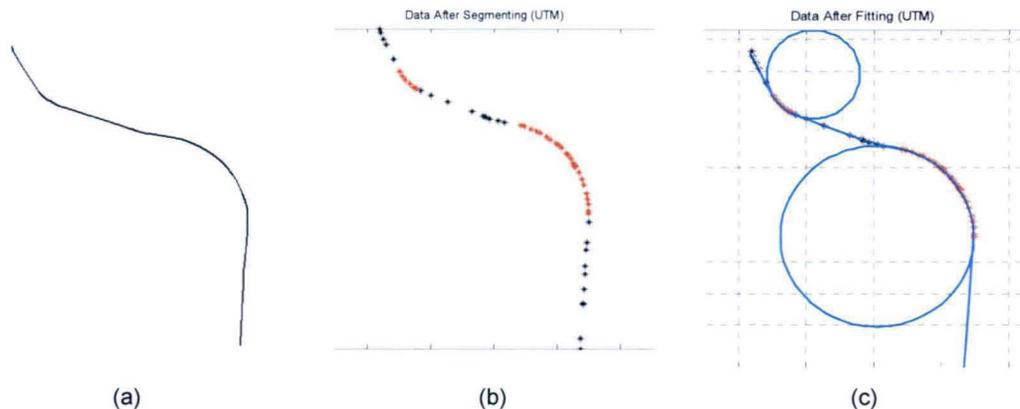
After finding all these values, the coordinates of TS can be computed. Then, the distances found above can be used for calculating the

coordinates of the stations. When reaching at point SC, the circular arc begins. Based on the radius and angle of the circular curve, all the elements of the circular arc can be computed.

## 5. Conclusion

This paper presented a method to automatically generate high-fidelity roadways based on shape files, civil engineering rules, terrain information and surrounding environment. Using curve generation as an example, detailed procedures

for polyline segmentation and parametric representations using the least square method was presented. Road curve computations based on civil engineering principles were also included. Preliminary results demonstrated the effectiveness of the proposed method. This work is an ongoing project and we will continue to improve our algorithms and representations, model more road features and apply our work to other applications as well.



**Figure 6:** Results from Polyline segmentation and fitting. (a) shows the polylines for a part of I-64 HOV lane based on the input shape file. (b) Segmentation results. Polylines are divided into several segments: red points for curve segment and black points for straight line segment. (c) Data after curve and line fitting. Radiuses for curve parts are extracted from discrete point data.

## REFERENCES

1. Parish, Y. I. H. & Müller, P. (2001) "Procedural modeling of cities," in *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, ACM.
2. Sun, J., Yu, X., Baci, G. & Green, M. (2002) "Template-based generation of road networks for virtual city modeling," in *Proceedings of the ACM symposium on VRST*. ACM: Hong Kong, China.
3. Sun, J., Baci, G., Yu, X. & Green, M. (2004) "Image-based Template Generation of Road Networks for Virtual Maps," *International Journal of Image and Graphics*, 4(4): p. 701-720.
4. Lechner, T., Watson, B., Wilensky, U. & Felsen, M. (2003) "Procedural City Modeling," in *1st Midwestern Graphics Conference*.
5. Lechner, T., Watson, B., Ren, P., Wilensky, U., Tissue, S. & Felsen, M. (2006) "Procedural Modeling of Urban Land Use," in *ACM SIGGRAPH 2006 Research posters*. ACM: Boston, Massachusetts.
6. Glass, K. R., Morkel, C. & Bangay, S. D. (2006) "Duplicating Road Patterns in South African Informal Settlements Using Procedural Techniques," in *Proceedings of the 4th international conference on Computer graphics, virtual reality, visualization and interaction in Africa*. ACM: Cape Town, South Africa.
7. Chen, G., Esch, G., Wonka, P., Müller, P. & Zhang, E. (2007) "Interactive Procedural Street Modeling," *International Conference on Computer Graphics and Interactive Techniques*, ACM Press New York, NY, USA.
8. Federal Highway Administration, (2008) *Federal Lands Highway: Project Development and Design Manual*, U.S. Department of Transportation. Report Number: FHWA/DF-88/003
9. California, Highway Design Manual, (2008) <http://www.dot.ca.gov/hq/oppd/hdm/hdmtoc.htm>

## **Appendix A – MODSIM World 2009 Conference Organizing Committee**

### **CP3 Members**

Chairman, Josh Jackson, SAIC

Vice Chairman, William Younger, MYMIC

President, Thomas Reese, VMASC

Secretary/Treasurer, E. Dana Dickens, III, HRP

Director, Dan Abahazy, JFCOM – Tom Frost, Proxy

Director, Jeff Bergenthal, Lockheed Martin

Director, Wayne Buck, NATO

Director, Don Combs, EVMS – Robert Alpino, Proxy

Director, Bob Harper, Northrop Grumman

NASA Liaison, Sharon Welch

### **Conference Committee**

MODSIM World 2009 Program Chair, Matt Spruill, SAIC

MODSIM World 2009 Program Deputy Chair, Technical Program, Dr. Thom Pinelli, NASA

MODSIM World 2009 Program Deputy Chair, Arrangements, Andrew Sinclair, HRP

Pre-Conference Workshops, Pete Schrider

K-20 Student Outreach, Jody A. Strausser, TCC

### **Logistics**

Administrative Services, Sheila Flanagan, VMASC

Communications, D'An Knowles, VMASC

Marketing Services, Tony Pino, Wisdom Harvest



## Appendix B – MODSIM World 2009 Track Chairs and Track Organizers

### DEFENSE & HOMELAND SECURITY TRACK

**LTC Keith M. Perkins** – Chair  
Chief, Innovation Section, Solutions Division  
Joint Warfighting Center, Joint Forces Command  
112 Lake View Parkway  
Suffolk, VA 23435  
NIPR: [keith.perkins@jfc.com](mailto:keith.perkins@jfc.com)  
SIPR: [keith.perkins@hq.jfc.com.smil.mil](mailto:keith.perkins@hq.jfc.com.smil.mil)  
Tele: 757.203.5916 DSN 668-5916  
FAX: 757.203.5851

**Maj. Ionel Vlasie** – Deputy Chair for Defense  
HQ SACT/NATO  
7857 Blandy Road  
Norfolk, VA, 23551  
Email: [ionel.vlasie@act.nato.int](mailto:ionel.vlasie@act.nato.int)  
Tele: 757.747.4323 (Office)  
Fax: 757.747.3271

**Mr. Bruce C. Milligan** – Deputy Chair for Homeland Security  
Federation of American Scientists  
1725 DeSales Street, NW, 6th Floor  
Washington, DC 20036  
Email: [bmilligan@fas.org](mailto:bmilligan@fas.org)  
Tele: 202.546.3300  
Fax: 202.675.1010

### 2009 Defense & Homeland Security Track Description

This track focuses on the the exchange of ideas and information, fostering research and collaboration, and on maximizing the application and contributions of modeling and simulation -- broadly defined to include modeling, simulation, visualization, game-based learning, and systems thinking/systems dynamics -- to defense and homeland security. For 2009, the track is structured as two separate but interconnected and closely related components. Roughly half of the sessions are devoted to each, and at least one session will be devoted to issues pertinent to defense and homeland security (at the federal, state, and local levels). This approach will provide attendees greater flexibility in terms of deciding which sessions and tracks to attend. Modeling and simulation have a long history of association and application to defense and, more recently, to homeland security. The 2009 Defense and Homeland Security Track will focus on the following five key aspects of modeling and simulation: (1) currently available products and training tools; (2) new and emerging technologies and related tools; (3) best practices and case studies of successful applications of modeling and

simulation; (4) the use of modeling and simulation for decision-making, support, and risk assessment; (5) and discussions and predictions about the future of modeling and simulation for use in defense and homeland security. Others aspects of interest include the development and advancement of modeling and simulation for (1) defense- and homeland security- related training, assessment, and evaluation; (2) technology integration; (3) protection of critical infrastructure and infrastructure resiliency; (4) disaster (contingency) planning, implementation, response, and recovery; (5) interoperability; and (6) the development of policy and strategies for managing global, national, regional, and local "incidents and events".

## EDUCATION & TRAINING TRACK

**Dr. Rasha Morsi** – Chair

Associate Professor, Electronics Engineering  
Director, Creative Gaming and Simulation (CGS) Lab  
Norfolk State University  
700, Park Avenue  
Norfolk, VA 23504  
Email: [rmorsi@nsu.edu](mailto:rmorsi@nsu.edu)  
Tele: 757-823-9514 (Office)  
Tele: 757-823-0047 (at the MCAR)  
Fax: 757-823-2698

**Mr. Phil Jones** – Deputy Chair

Manager, Training Solutions  
MYMIC, LLC  
200 High Street, Suite 308  
Portsmouth, VA 23704  
Email: [Phillip.jones@mymic.net](mailto:Phillip.jones@mymic.net)  
Tele: 757-391-9200 x104 (Office)  
Fax: 757.391.9098

**Mr. Mark Clemente** – Deputy Chair

Educator-in-Residence  
National Institute of Aerospace/Virginia Beach City Public Schools  
100 Exploration Way  
Hampton, VA 23666  
Email: [Mark.Clemente@nianet.org](mailto:Mark.Clemente@nianet.org)  
Tele: 757.325.6801

**Ms. Bonnie Matthews** – Deputy Chair Representing VSTE

1622 Fallen Timber Trail  
Powhatan, VA 23139  
Email: [bgmatthews@verizon.net](mailto:bgmatthews@verizon.net)  
Tele: 804.794.9546

**Mr. Jesse W. White** – Deputy Chair Representing VTEA  
Career and Technical Education  
Hampton City Schools  
School Administration Center  
One Franklin Street  
Hampton, VA 23669-3570  
Email: [jeswhite@sbo.hampton.k12.va.us](mailto:jeswhite@sbo.hampton.k12.va.us)  
Tele: 757.727.2466 (o)  
Fax: 757.727.2468

### 2009 Education & Training Track Description

This track focuses on presenting current developments and advancements in the use of modeling and simulation (M&S) in education and training, and the preparation and certification of M&S professionals and specialists. This includes pedagogies for modeling and simulation as content and methodology for K-12, community college, undergraduate, and graduate level education; professional development; and continuing education. Education and training are broadly defined to include instruction, instructional-related issues and techniques, curricula, standards, accreditation, and resources. The education and training track serves as a forum for the (1) exchange of knowledge; (2) interchange among individuals and parties; and (3) presentation, discussion, and resolution of issues pertaining to new developments, pedagogy, and educational modeling and simulation.

### CROSS MARKETING PARTNER FOR EDUCATION

Virginia Society for Technology in Education (VSTE)  
Virginia Technology Education Association (VTEA)

### ENGINEERING & SCIENCE TRACK

**Mr. Mark Lucas** – Chair  
Technical Director  
Naval Surface Warfare Center – Dam Neck  
1922 Regulus Ave  
Virginia Beach, VA 23461-2097  
Email: [mark.lucas@navy.mil](mailto:mark.lucas@navy.mil)  
Tele: 75.492.7563 (Office)  
Fax: 757.492.7275

**Mr. Bill Pritchett – Deputy Chair**  
Product Development Manager  
Simulation and Training Group  
Raytheon Command and Simulation Systems  
5510 Cherokee Ave, Suite 350  
Alexandria, VA 22312  
Email: [wpritchett@raytheonvtc.com](mailto:wpritchett@raytheonvtc.com)  
Tele: 703.333.6229 (o)  
Tele: 703.946.1403 (c)

**Ms. Tammy R. Van Dame**  
Director, Corporate Communication  
Naval Surface Warfare Center, Dam Neck  
1922 Regulus Ave  
Email: [tammy.vandame@navy.mil](mailto:tammy.vandame@navy.mil)  
Virginia Beach, VA 23461-2097  
Tele: 757.492.6155 (office)  
Tele: 757.675.6394 (cell)

### 2009 Engineering & Science Track Description

This track offers a forum for information exchange among individuals and groups concerned modeling and simulation (MODSIM) theory, tools and techniques, and its application and practice in and across all engineering disciplines. Topics of interest include, but are not limited to, MODSIM in life cycle engineering, in the application of MODSIM for the RDT&E of defense and commercial products and services, and environmental protection and alternative energy development. This track also includes the use and application of MODSIM in the natural and physical sciences and in basic and applied research. Areas of interest include, but are not limited to, theoretical and experimental developments in MODSIM, verification and validation, integration and interoperability, open architecture, information assurance, current and future application of MODSIM in new and emerging scientific disciplines.

### **HEALTH & MEDICINE TRACK**

**Ms. Reba Moyer Childress – Chair**  
Director, Clinical Simulation Learning Center -  
<http://www.nursing.virginia.edu/students/cslc.aspx>  
Founding Board Member, International Nursing Association for Clinical Simulation and Learning - <http://inacsl.org/>  
Founder & Director, Virginia State Simulation Alliance (VASSA)  
University of Virginia School of Nursing  
PO Box 800782  
Charlottesville, VA 22908-0782  
Email: [rmc6h@virginia.edu](mailto:rmc6h@virginia.edu)  
Tele: 434.924.2748; 434.924.0086

**Mr. Jeff McCrindle** – Deputy Chair  
Vice President Sales and Marketing  
Education Management Solutions, Inc.  
The Oaklands Corporate Center  
440 Creamery Way, Suite 100  
Exton, PA 19341  
Email: [Jeffrey.McCrindle@ems-works.com](mailto:Jeffrey.McCrindle@ems-works.com)  
Tele: 610.701.7002 x260

**Ms. Beth H. Pettitt** – Deputy Chair  
Chief, Solider Simulation Environments  
U.S. Army RDECOM  
12423 Research Parkway  
Orlando, FL 32826  
Email: [Beth.H.Pettitt@us.army.mil](mailto:Beth.H.Pettitt@us.army.mil)  
Tele: 407.384.3934 (Office)  
Fax: 407.384.5440

**Dr. Michelle Roper** – Deputy Chair  
Sentech, Inc, Department of Energy  
Office of Energy Efficiency and Renewable Energy - Room 5E-042, Mail Stop EE-1  
1000 Independence Avenue, SW  
Washington, DC 20585  
Email: [Michelle.Fox@ee.doe.gov](mailto:Michelle.Fox@ee.doe.gov)  
Tele: 202. 287.5517

**Dr. Mary Ann Notarianni** – Deputy Chair  
Associate Professor  
School of Nursing  
Old Dominion University  
Health Sciences Building 2130  
Norfolk, VA 23529  
Email: [mnotaria@odu.edu](mailto:mnotaria@odu.edu)  
Tele: 757.683.5093

### 2009 Health & Medicine Track Description

Natural and man-made biological threats, public health emergencies, and uncontrollable costs threaten to overwhelm today's educational and healthcare delivery systems, and society. In response to these challenges, subject matter experts, administrators, simulation developers, educators and practitioners collaborate to develop cost-effective serious games and simulations that allow for the artful use of modeling and simulation, and better analysis of complex systems for improved incident action plans, decision making, and training. The Health & Medicine Track provides a

forum for experts and stakeholders from a variety of disciplines to share advances in simulation and modeling to identify and solve problems, and foster future research opportunities that expand the role of modeling and simulation to dramatically improve patient outcomes and overall disaster response.

### THE HUMAN DIMENSION (Modeling Human Behaviors and Interactions)

**Dr. Kara A. Latorella** - Chair  
Crew Systems Branch, M/S 152  
NASA Langley Research Center  
Hampton, VA 23681-0001  
Email: [k.a.latorella@nasa.gov](mailto:k.a.latorella@nasa.gov)  
Tele: 757.864.2030  
Fax: 757.864.7793

**Mr. Phillip N. Jones** – Deputy Chair  
Director, Analysis, Training & Assessment  
MYMIC LLC  
200 High Street, Suite 308  
Portsmouth, Virginia 23704  
Email: [Phillip.Jones@mymic.net](mailto:Phillip.Jones@mymic.net)  
Tele: 757.391.9200 x104

#### 2009 The Human Dimension (Modeling Human Behaviors and Interactions) Track Description

Understanding the human dimension cuts a broad swath; from individual performance to thought to the behaviors of humans in complex human/machine systems, cultures, and personal interactions. This conference track seeks papers/presentations that model/simulate individual human behavior (information seeking and monitoring, decision making, and controlling) and the influence of affect on behavior, as well as the interaction of individuals; especially as participants in complex human/machine systems and human social networks and organizations. Papers/presentations should address the utility of models and simulations in terms of their application to solving design and analysis problems such as: characterizing individual differences; predicting human performance & errors; designing and evaluating human/system integration and human/automation interaction; team construction & performance; and predicting communication dynamics, or other emergent social behaviors. Equally important are methodological papers/ presentations that provide innovative approaches to requirement definition, validation and verification, as well as data collection and analysis.

## SERIOUS GAMES

**Ms. Jennifer McNamara** - Chair  
Director of Strategic Partnerships  
BreakAway  
10150 York Road, Suite 250  
Hunt Valley, MD 21030  
Email: [jmcnamara@breakawayltd.com](mailto:jmcnamara@breakawayltd.com)  
Tele: 410.316.9693 (Direct)  
Tele: 610.420.8812 (Cell)  
Fax: 410.683.1706

**Mr. Brent Smith** - Deputy Chair  
VP/Chief Technology Officer  
Engineering & Computer Simulations, Inc.  
3501 Quadrangle Blvd., Suite 260  
Orlando, FL 32817  
Email: [brents@ecSORL.com](mailto:brents@ecSORL.com)  
Tele: 407.823.9991 x 305 (Office)  
Tele: 407.497.7340 (Cell)

**Mr. Joseph M. Saur** - Deputy Chair  
Sr. Research Scientist, GTRI  
Joint Systems Integration Command  
116 B Lake View Parkway  
Suffolk, VA 23435  
Email: [joseph.saur@jsic.jfcom.mil](mailto:joseph.saur@jsic.jfcom.mil)  
Tele: 757.203.4016 (Office)  
Fax: 757.

### 2009 Serious Games Track Description

Even before computers, people performed complex modeling and simulate activities – directing state-of-the-art technologies – to explore and understand complex phenomena. Over the years, computer hardware and software developments have made increasingly advanced computation covering a broader range of issues possible. Recently, the application of game concepts, design and technologies has brought new perspectives to modeling and simulation. This track focuses on the area of “serious games” – the discipline that leverages game concepts, design and technology for modeling and simulation, training, advertisement, education and social change. Presentations in this track will focus on exploring all dimensions of serious games including those relating to the broader MODSIM World 2009 tracks: Defense & Homeland Security, Education & Training, Engineering & Science, Health & Medicine, Transportation & Logistics, and the Human Dimension. By exploring the tools, technologies, design and implementations of serious games this track intends to challenge traditional perspectives to induce new ideas for how these capabilities can be applied to modern day challenges.

## TRANSPORTATION & LOGISTICS TRACK

**Dr. Cenk Tunasar** - Chair  
Senior Associate  
Booz Allen Hamilton  
8283 Greensboro Drive  
McLean, VA 22102  
Email: [tunasar\\_cenk@bah.com](mailto:tunasar_cenk@bah.com)  
Tele: 703.377.7961 (Office)  
Tele: 202.352.1030 (Cell)

**Ms. Selena D. Griffett** – Deputy Chair for Transportation  
Senior Civil Engineer-VDOT Projects Manager -- Engineering Services  
22 Lincoln Street  
Hampton, VA 23669  
Email: [sgriffett@hampton.gov](mailto:sgriffett@hampton.gov)  
Tele: 757.728.2038

**Dr. Bradley D. Silver** – Deputy Chair for Logistics  
Program Manager, Mathematical Modeling Group  
Logistics Management Institute  
2000 Corporate Ridge  
McLean, VA 22102  
Email: [BSILVER@lmi.org](mailto:BSILVER@lmi.org)  
Tele: 1.800.213.4817 x-117866

### 2009 Transportation & Logistics Track Description

Today's business environment is increasingly dependent upon complex transportation networks and extensive supply chains. Positioning and distribution of raw materials and finished goods represent a significant investment. By extension, they also represent a significant opportunity for improved competitiveness through the application of modeling and simulation techniques. In the public sector, management of our nation's transportation infrastructure, such as our highways, seaports, airports, and air traffic control network, affects all our lives on a daily basis. The logistics of providing public services, ranging from emergency response and relief to routine distribution of aid, is under intense scrutiny. Modeling and simulation tool offer managers in the public sector advanced capabilities to help deliver vital services more safely, effectively, and efficiently. The Transportation and Logistics Track will explore complex network simulations, advanced decision support systems, visualization technologies, and other such modeling and simulation tools and techniques and their specific applications to the transportation and logistics challenges in business and the public sector. This track's scope includes analysis efforts such as traffic congestion on roads, at seaports, and across the aviation system; supply chain and inventory optimization; security and cost of multimodal transportation operations; fleet maintenance concepts; and incident response and evacuation



REPORT DOCUMENTATION PAGE			Form Approved OMB No. 0704-0188		
<p>The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.</p>					
1. REPORT DATE (DD-MM-YYYY) 01-03-2010		2. REPORT TYPE Conference Publication		3. DATES COVERED (From - To)	
4. TITLE AND SUBTITLE Selected Papers Presented at MODSIM World 2009 Conference and Expo			5a. CONTRACT NUMBER		
			5b. GRANT NUMBER		
			5c. PROGRAM ELEMENT NUMBER		
6. AUTHOR(S) Pinelli, Thomas E. (Editor)			5d. PROJECT NUMBER		
			5e. TASK NUMBER		
			5f. WORK UNIT NUMBER 736466.01.04.07.02		
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) NASA Langley Research Center Hampton, VA 23681-2199			8. PERFORMING ORGANIZATION REPORT NUMBER  L-19843		
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) National Aeronautics and Space Administration Washington, DC 20546-0001			10. SPONSOR/MONITOR'S ACRONYM(S)  NASA		
			11. SPONSOR/MONITOR'S REPORT NUMBER(S) NASA/CP-2010-216205		
12. DISTRIBUTION/AVAILABILITY STATEMENT Unclassified - Unlimited Subject Category 66 Availability: NASA CASI (443) 757-5802					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT Selected papers from MODSIM World 2009 Conference and Expo are contained in this NASA Conference Publication (CP). MODSIM World 2009 was held in Virginia Beach, Virginia, October 14-16, 2009, at the Virginia Beach Convention Center. The theme of the 2009 conference and expo was "21st Century Decision-Making: The Art of Modeling & Simulation." The 79 submitted papers were peer-reviewed and 64 were accepted for presentation at MODSIM World 2009. As a condition of acceptance, the first author was responsible for securing/obtaining all permissions associated with the general release and public availability of the paper. Further, the first authors also had to grant NASA the right to include their papers in the NASA CP. There are 53 papers in this NASA CP.					
15. SUBJECT TERMS Modeling and simulation; Agent-based modeling; Game theory; Discrete event simulation; Cognitive modeling; Visualization					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			STI Help Desk (email: help@sti.nasa.gov)
U	U	U	UU	350	19b. TELEPHONE NUMBER (Include area code) (443) 757-5802