Metrics for Intelligent Autonomy

S. O'Day
Eagan, McAllister Associates
47332 Eagan, McAllister Lane
P.O. Box 986
Lexington Park, MD 20653

M. Steinberg, C. Yglesias
Office of Naval Research

K. Bonnevier, T. Moulds
NAVAIR
Patuxent River Naval Air Station, MD

R. Sincavage, M. Ricard, M. Cleary
Draper Laboratory
555 Technology Square
Cambridge, MA 02139

M. Curry, T. Sliski, J. Tierno
Alphatech, Incorporated
6 New England Executive Park
Burlington, MA 01803

J. Clark
Lockheed Martin
Fort Worth, TX

F. Snyder
Northrop Grumman Electronic Systems
Pittsburgh PA

ABSTRACT

Intelligent Autonomy (IA) is a multi-year program within the Office of Naval Research (ONR) Autonomous Operations (AO) Future Naval Capabilities (FNC) program. The primary goal of the effort is to develop and demonstrate technologies for highly automated and fully autonomous mission planning and dynamic re-tasking of multiple classes of Naval unmanned systems and minimization of human intervention in unmanned vehicle operations. This technology is being applied to both individual and teams of unmanned air, surface, ground, and undersea vehicles for a variety of mission areas including reconnaissance/search, persistent surveillance, tracking, and some limited application to strike. Autonomy technologies will be matured through a series of phased demonstrations to allow low risk transition to current and future Navy and Marine Corps systems. Demonstrations will be done using both real vehicles and simulation. Some of the major simulation demonstrations will be done within the context of a simulated warfare environment at the Naval Air Systems Command based around the Air Combat Environment Test & Evaluation Facility (ACETEF) and the Unmanned System Research and Development Lab (USRDL). The demonstrations at NAVAIR will utilize much of the architecture and many of the assets from the NCW4.0X Virtual Laboratory (V-LAB) project. Metrics for testing of IA software in this environment are currently being developed. This paper will discuss some candidate performance metrics that are currently being considered for evaluation of the Intelligent Autonomy technologies.

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE **AUG 2004**	2. REPORT TYPE	3. DATES COVERED **00-00-2004 to 00-00-2004**
4. TITLE AND SUBTITLE **Metrics for Intelligent Autonomy**		5a. CONTRACT NUMBER
		5b. GRANT NUMBER
		5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)		5d. PROJECT NUMBER
		5e. TASK NUMBER
		5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Eagan, McAllister Associates,47332 Eagan, McAllister Lane,P.O. Box 986,Lexington Park,MD,20653**		8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)		10. SPONSOR/MONITOR'S ACRONYM(S)
		11. SPONSOR/MONITOR'S REPORT NUMBER(S)
12. DISTRIBUTION/AVAILABILITY STATEMENT **Approved for public release; distribution unlimited**		
13. SUPPLEMENTARY NOTES **Proceedings of the 2004 Performance Metrics for Intelligent Systems Workshop (PerMIS -04), Gaithersburg, MD on August 24-26 2004**		
14. ABSTRACT **see report**		
15. SUBJECT TERMS		

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**5**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

KEYWORDS: *unmanned aerial vehicle, unmanned undersea vehicle, mission planning, autonomous control, JIMM*

1. INTRODUCTION

The Intelligent Autonomy Effort under the Autonomous Operations Future Naval Capability is developing a range of technologies that will be challenging to effectively evaluate using traditional metrics. The goals of the Intelligent Autonomy (IA) effort IA are to

•Provide autonomy software for highly automated and fully autonomous dynamic retasking of multiple classes of Naval unmanned systems to perform littoral reconnaissance, search, persistent surveillance, and to a limited extent tracking and strike.
•Minimize human intervention via autonomous and highly automated mission planning/replanning functions and operator aids such as alert management and plan understanding for individual vehicles and teams of vehicles.
•Enable limited automated surveillance and reconnaissance data processing for surface and shoreline object detection and classification to provide autonomous replanning based on sensed information, bandwidth reduction, and operator workload reduction.

The technologies demonstrated under the IA product lines will be applicable to multiple types of Naval unmanned vehicles including unmanned air, undersea, ground, and surface vehicles with a focus on air and undersea vehicles and control stations. This effort will leverage numerous DOD programs in autonomy to support specific Navy and Marine Corps unique and essential needs. Intelligent Autonomy technologies will be demonstrated through a series of phased demonstrations to allow low risk transition to current and future Navy and Marine Corps systems. The primary areas being developed and demonstrated under the IA program are:

UxV High-Level Planning/Replanning
Lead Performers: Alphatech, Draper Laboratory
Allocate mission tasks to available platform/payload types out of a team of 5-10 heterogeneous vehicles and determine an optimal sequence of mission tasks with associated time windows/constraints based on high-level tasking (platform availability, team mission tasks, priorities, and constraints).

UAV Dynamic Replanning
Lead Performer: Lockheed Martin, Ft. Worth
Produce UAV mission plan that optimizes survivability & employment of on-board capabilities while meeting an ordered set of mission objectives and constraints.

UUV On-Board Dynamic Mission Replanning
Lead Performer: Draper Laboratory
Generate minimum-cost, energy efficient, safe routes to achieve combination of mission tasks within constraints.

Alert Management & Replan Assessment
Lead Performer: Lockheed Martin, Ft. Worth
Replan Assessment to analyze mission plan changes, monitor/assess contingencies, assess contingencies, and trigger a replan or alert if necessary. Alert Management to determine the level and type of alerts received, store alerts received and forward them to be displayed, and support the operator in recovering the context of tasks interrupted.

Mixed-Initiative Interface Manager
Lead Performers: Lockheed Martin ATL, Charles River Analytics, and Aptima
Display relevant mission information, provide plan-understanding capabilities, and enable the operator task the vehicles and set the level of autonomy.

Distributed Cooperative Control
Lead Performer: Alphatech
Enable autonomous mission replanning among teams of vehicles with limited communications

Maritime Image Understanding
Lead Performer: Northrop Grumman Electronic Systems
Develop video processing technology in the river and harbor domain. Detect and classify mission relevant objects to support autonomous navigation and surveillance.

2. PROJECT DESCRIPTION AND SCHEDULE

Initially, the developers will demonstrate the functionality and capability of their products using medium to high-fidelity simulation models at their facilities. Later, software algorithms are planned to be integrated with Naval control stations or vehicles to enable testing and maturation of the software products. The algorithms will be demonstrated in both simulation and hardware/in-water demos.

NAVAIR will provide a test bed for development using the Air Combat Environment Test & Evaluation Facility (ACETEF) and the Unmanned System Research and Development Lab (USRDL). A specific Joint Integrated Mission Model ACETEF (JIMMACE) port scenario database will serve as the warfare environment for mission planning system development and the Tactical Control Station (TCS) will serve as a baseline for operator station development.

3. WARFARE ENVIRONMENT

JIMMACE may cede control of specific player tactics and system functions to external IA hardware or software assets that interface to JIMMACE shared memory. Object instances are defined in the JIMMACE scenario database (SDB). The extent of asset control is defined via statements in the control database (CDB). The types of objects (players, platforms and systems) in the environment, the tactics that these objects execute as well as the command, communications and control architecture are being defined in the JIMMACE type database.

The JIMMACE model will provide the warfare environment for integrated IA demonstrations. IA assets access and control parts of the JIMMACE warfare environment via shared memory network interfaces. DIS, HLA and UDP datagram socket connection protocols are used to transmit the data between IA assets and shared memory. The JIMMACE model creates and fills information into shared memory based on databases written in its native language (which consists of English language phrases which are combined in a straightforward syntax).

4. OPERATOR ENVIRONMENT

The IA program is focused on operator functions related to mission management only. It is assumed that either the vehicles are highly autonomous or that there are additional operators or operator functions concerned with vehicle management issues, such as shipboard recovery and air traffic management issues for UAV's. The system operators will interface with the IA hardware and software assets through the TCS interface and the new IA operator interface modules. Serving as a network interface between the operator interface (TCS and the new IA OI's) and the warfare environment will be the Unmanned Simulation System (USS) stimulator. The USS will provide a way for the evaluators to insert real-time scenario modifications such as vehicle or sensor cautions, warning, and emergencies. This will be used as part of the evaluation metrics in testing the interaction between the IA technologies and the operator(s) in relation to SA and workload.

The baseline operator environment will be a 1-3 person team that may consist a mission commander, vehicle operator, and/or a sensor operator. The number of operators depends on the type and number of unmanned vehicles and the complexity of the mission tasks. A secondary baseline derived from the JUCAS concepts will be a five-person team that will dynamically split vehicle control aspects with sensor control and C4I aspects.

The USS will interface with JIMMACE via a HLA interface. The USS is a CORBA based architecture and supports custom interfaces via UDP and TCP.

5. CANDIDATE METRICS

This section will describe some of the major candidate metrics that are being considered for use on the IA effort. There are a variety of useful measurements that can characterize the engineering quality of unmanned vehicle simulation, intelligent autonomy, and operator control station software. These can be roughly categorized as:

- response of system components to a range of initial input parameters
- human factors of operator mission management and situational awareness
- response of system components to changes in the simulated warfare environment during execution

Metrics relating the unmanned vehicle simulation system performance to mission goals include:

- optimization of the number and mixture of unmanned vehicles to maximize the number of successful missions
- impact of mission re-planning time on mission success
- optimum distance between assets and targets for maximal mission success
- impact of reactive/creative maneuvers on mission success
- loss of assets in mission completion/objectives completed
- operator/vehicle ratio

The response of the system to different initial conditions can be measured in terms of time and impact on mission success. Top-level parameters, which can be varied, include:

- geographic size of gaming area and placement of assets
- weather, terrain and other environmental factors
- number of missions
- number of each type of unmanned vehicle
- mission re-planning cycle time
- extent of UV intelligent tactics
- types of missions within a particular scenario

Each proposed system has a number of engineering metrics which also relate to mission flexibility and success.

5.1 *Mission Software Component Metrics*

The IA project accepts certain scenario dependent mission coverage metrics to evaluate the IA mission-planning components. The time it takes to turn around a mission plan or re-plan is an obvious metric. This can be dependent on the number of constraints, number of total missions underway and on a variety of warfare environment parameters. The goal is then to maximize the number of simultaneous missions that can be executed within the context of the following:

- Number of simultaneous mission tasks that the system can handle
- Number of mission task types that the system can handle
- Planning exception rate (dropped tasks over total tasks)
- Fraction of mission constraints not met (if feasible)

In the area of dynamic performance metrics, it is hoped that the time elapsed between task appearance and completion is minimized within the context of the following scoring ratios $S(t)$ where t is the time from the beginning of the scenario:

- Optimization of mission-specific cost-measures
- Discounted optimization of mission-specific cost-measures to emphasize timeliness

Stability/sensitivity metrics are used to avoid frequent changes in plans that may have a detrimental effect on mission performance and operator situation awareness. One such measure is that of "thrashing" in tasking where one takes the time elapsed between execution of one task and the last change of the preceding task. This can also be examined as a function of communication bandwidths, error rates, and scenario variation.

5.2 *Operator Station Metrics*

Since the IA program is focused on operator functions related to mission management only, there are some differences from the types of metrics traditionally used for operators directly controlling the vehicle. Of particular concern is the neglect-tolerance of the system. This concerns both how well the autonomous system behaves when there is limited human intervention and how well the human operator is able to maintain situation awareness when not constantly in the loop or when managing multiple vehicles and mission tasks. The impact of human performance on the overall system can appear at several levels:

- system interoperability level (with external assets)
- software system level (e.g. efficiency and accuracy of UV mission prosecution by this system)
- operator station component level

System level measures can be used to identify the decision-making roles in which the human is most influential and effective relative to the capabilities of the automation. Measures should be made under varying levels and types of human intervention for factors such as

- Speed and accuracy for decisions and actions
- Time to respond to critical events
- Duration of mission activities
- Ratio for completion of "Mission-Critical Objectives" vs. "Secondary Objectives"

Task loading metrics are critical for estimating the required number of operators required for a mission. This can be drawn out from the speed and accuracy of task completion for different levels of task demands associated with the mission (e.g., the number and rate of required tasks for successful mission completion; complexity of the mission, etc.). Objective measures can also be used to identify the points at which the operator begins either shedding tasks or failing to achieve accurate task completion.

Another important type of metric is subjective workload measures (i.e. NASA-TLX). These enable operators to rate their experience of mission difficultly/cognitive demands for both the overall workload of the mission and the workload associated with select critical incidents and mission phases. These measures are helpful for identifying the appropriate distribution of task load and organizational structure for a team of operators and the areas where additional automation may be desired.

Examples of relevant metrics are:

- the quality and extent of operator station training that is needed for operators to be effective in using the system
- speed of task completion vs. mission completion requirement speed
- accuracy of task completion
- identification of points at which critical tasks are dropped
- mission workload (overall and for critical tasks)
- reduction of required operators without impact to mission effectiveness

For the operator to achieve effective mission management of the system, it is important to maintain situation awareness to the progress of the mission. There are a number of subjective measures that can be used where operators rate their understanding of the situation. There are also objective measures such as "blanking the screen" and asking the operator to answer questions about key features of the situation & make predictions about expected mission progress. A final area is SA for critical incidents. These measures can be used to identify the effectiveness of user interface displays in allowing the operator to monitor key events relevant to mission tasks.

Some subjective human performance measures include:

- operator understanding of the mission complexity
- SA during the mission
- operator correctly using automated capabilities
- operator trust of automated capabilities
- efficiency and accuracy of decision-making
- operator effectiveness in mission prosecution

These measures can also be used to judge the effectiveness of the system concept of employment.

Factors, which will affect the outcome of such measurements, are:

- the background of the operator
- the number of operators
- the quality and extent of operator training
- quality of the operator interface software
- the commonality to other operator interfaces of relevant systems
- complexity and tempo of the mission
- the level of autonomy of the vehicle

• the complexity of the scenario

5.3 *Maritime Detection and Classification Metrics*

Images collected and processed by unmanned vehicles are useful to mission planning if they provide detailed topological and/or object location and identification information.

These capabilities may be described by the following measures:

- accurate spatial digitization of objects and environment by the image processing algorithms
- robustness of image processing algorithms in varied environments
- ability of algorithms working with COTS products to do real-time image processing and object recognition
- ability to derive understanding/real-time map of harbor and river environments from vehicle ISR image processing
- efficiency with which real-time image processing results can feed mission re-planning

5.4 *JIMMACE System Modeling*

The JIMMACE model in conjunction with a shared memory interface can be used to model an idealized mission planner and fully automated mission control operator station. JIMMACE tactics can be employed to simulate a systems with varying degrees of autonomy. JIMMACE simulated players can assume the role, mission and function of operators and mission planners. Metrics can be extracted from the model output for comparison with the mission planning components and manned operator stations to identify inefficiencies.

6. SUMMARY

There is still a great deal of uncertainty about how best to use metrics to evaluate future autonomous system. This paper discussed a range of approaches to metrics that are currently being examined for use in planned demonstrations with a wide variety of autonomy components. Experimentation with different metrics over the course of these demonstrations will be help better define under what circumstances a particular metric is appropriate and useful.

7. REFERENCES

[1] J. McMaster, S. O'Day, J. Sarratt, "Millennium Challenge 2002 Fleet Battle Experiment Juliet Analytical Excursions with JIMMACE", 9th ITEA Modeling and Simulation Workshop, Las Cruces, NM, December 8-11, 2003.