ASC: 03-0347

AN ANALYSIS SYSTEM RELATING INDIVIDUAL HUMAN PERFORMANCE MEASURES TO OVERALL MISSION EFFECTIVENESS

Bryan E. Brett
Jeffrey A. Doyal
Dr. Christopher R. Hale
Science Applications International Corporation
Beavercreek, Ohio

When applied in acquisition the goal of modeling and simulation is to support technology decisions. Both constructive and virtual simulations produce great quantities of data. The information derived from these data, if obtained in a clear and timely manner, can provide acquisition professionals with insight into critical issues surrounding development and selection of new technologies. Currently, however, the derivation process is expensive, time-consuming and often unreliable. The intent of the CART data visualization effort is to improve this process by developing a suite of applications addressing critical aspects of an overall solution to the problem of translating simulation data to technology acquisition decisions. Our experience has shown that 3 levels of analysis are required to thoroughly understand the results provided by a CART simulation: specifying effects of technology alternatives on mission performance, tracing effects of human performance on mission performance and conducting detailed analysis of why failures occurred. Research addressing the first two levels is currently being conducted. Our approach in this research is embodied in the development of 6 applications: a test plan description application that serves as a simulation design database for raw data, an abstraction hierarchy application that links lower levels of human performance to high-level mission outcomes, a data repository application that allows description of data file contents in terms of elements critical to required acquisition decisions, a performance measure definition application that allows users to define performance measures at level of the abstraction hierarchy, a performance measure computation application that automates performance measure calculation, and a diagnostic hierarchy exploration application that allows analysts to trace low-level performance effects to mission outcomes in ways enabling comparisons across test conditions. As part of the discussion of our approach we will highlight methodological and conceptual considerations that have driven development of this toolset.

Introduction

The Combat Automation Requirements Testbed (CART) program (Brett, B. E., Doyal, J. A., Malek, D. A., Martin, E. A.,& Hoagland, D. G., 2000) is a simulation-based acquisition research and development project that is developing and demonstrating human performance modeling technology as a means of representing the warfighter in constructive simulations performed during weapons system acquisition. While human performance modeling is the primary focus of CART, the program also emphasizes the process by which modeling and simulation is applied to support acquisition decisions.

There are three major phases involved in the CART process. The first is mission decomposition. In this phase, the key mission(s) of interest for the system under consideration are decomposed hierarchically into increasing levels of descriptive detail. The objective of the decomposition process is to identify those operator tasks that are to be modeled along with the requirements for system and mission environment simulations that will exercise the operator models.

In the second phase, a simulation testbed is developed and applied to collect data that address crew station and other acquisition issues. CART testbeds generally consist of multiple simulation components linked via distributed simulation technology. These components represent the operator(s) of interest, the system (s) with which the operators interact, and the mission environment in which the operators and system perform. Requirements for developing the testbed are derived from the mission decomposition and a test plan that specifies concepts, technologies, and other conditions to be manipulated in testing.

The test plan also specifies performance measures that will be used to evaluate simulation outcomes and support comparisons between the different test conditions. An important component of testbed development is implementing a data collection and reduction capability that collects data from the simulation testbed, manipulates that data to calculate performance measures, and stores the results in a format that can be accessed for statistical and other analysis activities. Because of the complexity of CART testbeds, the data collection and reduction capability tends to be complex and require significant effort to develop and employ.

The third phase of the CART process involves analysis of the data collected by the testbed. This is, of course, the reason the testbed was developed in the first place. In the analysis process major outcomes such as mission performance are evaluated first to determine the overall effects of the variables manipulated in testing. Later, the data are explored in greater to detail to gain insight into the factors that contributed to the overall results. This process of data analysis and exploration is iterative. As the process unfolds and questions are raised and insights gained, new performance measures or derived data often are determined to be required. In these instances it is necessary to rework or extend the data collection and analysis capability.

As the CART program has progressed, the importance of the data collection and analysis phase has grown. The ultimate goal of CART modeling, as with all other SBA efforts, is to produce data to support decision-making. That is, the data provided by CART simulation testbeds becomes information. This information provides acquisition professionals with insight into and a greater understanding of critical issues and considerations involved in the selection and development of crew systems and other components of new equipment systems. Ideally, CART and other SBA programs need the ability to efficiently develop information from their simulation data collection systems that address specific issues of interest to different stakeholders. In practice, however, this is rarely achieved. Recent efforts on CART have turned to addressing two major challenges in efficiently and effectively developing and applying data collection and reduction capabilities. Each is described below.

Need to Assess Performance at Different Levels and Trace Relationships Between Levels

A variety of stakeholders will be involved in a given SBA project. The information each requires from the testbed will be different. Program managers and warfighter leadership will be concerned with mission level performance of the new system. Operations analysts will be interested in system performance in different mission segments and functions. Human factors analysts will want to understand how factors such as function allocation and crew system design and technology impact warfighter performance. Acknowledgement of the different information needs of stakeholders has led to recognition that multiple levels of analysis are required to thoroughly understand the results provided by a CART simulation testbed and to meet the information needs of different stakeholders. Specifically, three levels of analysis have been identified. These are:

1. Specify effects of system alternatives on mission performance. This level of analysis focuses on high-level performance measures that examine impacts on mission performance in terms of the alternative concepts under consideration. As used here, performance measures are aggregations of lower level event or sampled data from a simulation run or series of runs that summarize operator or system performance along critical performance dimensions. Examples of these summary performance measures are *percent targets destroyed* or *average time to identify a target*. Generally, statistical comparisons are conducted across and test conditions to determine the magnitude and reliability of effects observed.

2. Trace impacts of lower level system component performance on high level mission performance. In the CART program, the interest is in how lower level human performance affects mission outcomes. Providing these linkages has been a challenge for the human factors community. It is not enough to describe the impact of new crew systems concepts and technology on, for example, workload and situation awareness. Acquisition decision makers want to understand how human factors issues and technologies impact overall system and mission performance. Beyond human performance modelers, other users of a CART testbed might be concerned about the impact of physical system components on mission outcomes. A radar engineer, for example, will want to understand how ground mapping radar attributes such as range and resolution contribute to the success of target detection, identification, and attack processes. This intermediate level of analysis would provide such an understanding.

3. Conduct detailed analysis of performance deficiencies/failures. The two levels of assessment discussed above use summary performance measures to evaluate mission, system component, and operator performance. While summary performance measures can provide insight into particular dimensions of system and operator performance where problems occur, they do not help us understand why the problems occurred. For this, it is necessary to identify individual instances of performance failure and examine the context and factors that contributed to the failure. This approach to analysis is consistent with

ecologically oriented concepts such as Use-Centered Design (Flach, J. and Domingez, C., 1995) that stress the importance of understanding the constraints that impact human performance in a given domain. Performance failures can be viewed as instances in which a constraint or boundary was exceeded. Studying the context in which the failure occurred could provide insight into the nature of the constraint(s) that precipitated the failure. This is important in complex systems contexts where the constraints themselves can be complex and difficult to understand. The ability to conduct detailed analysis of performance deficiencies/failures requires the ability to examine scenarios individually and identify specific instances of performance failure. The context for a performance deficiency in a scenario can be recreated by generating a timeline that presents the different events and factors related to the deficiency in temporal relation to one another. Using such data, it is possible, for example, to determine that launches of a stand-off weapon consistently occur well within the weapons engagement envelope because the requirement to visually acquire and identify the target forces the pilot to use an electro-optical system with limited range. By the time the target is identified and the weapon is designated and enabled, the aircraft is well within the weapon envelope and dangerously exposed to missile threats.

Too Much Effort Required to Obtain and Analyze Simulation Data

Currently, analysis of simulation data in CART and most other simulation environments is expensive and time consuming. This is driven in part by the massive quantities of raw data generated by multi-component simulations that execute scenarios involving hundreds of entities. Even if data are collected at a relatively low rate (e.g., 1 Hz), large files can be generated in a multi-hour scenario. Often the data collected in a simulation testbed must be converted or pre-processed into a format that can be manipulated more readily by post-processing software. Also, once the raw data are converted into a format amenable for manipulation, the post-processing of this data often occurs across a series of steps that are initiated, if not performed, manually. Frequently, this post processing consists of custom developed applications and macros designed to calculate specific performance measures. Within current data collection and analysis systems considerable effort must be applied to rework post-processing and performance measure calculation software and macros when analysts decide they need new or different performance measures. Basically, the analyst must revisit many of the steps performed in the initial development of the data collection and analysis system.

Beyond the manipulation of data, the analysis process can require exploration of a variety of performance measures with complex relationships between one another. At present, this exploration process is manual and involves paper reports and graphics or perhaps extensive use of spreadsheets. It can be a time consuming and cumbersome process.

The impact of the current process for manipulating and exploring simulation data is that the level of effort required to produce performance measures and data from a simulation testbed is so significant that only a subset of the potential measures and data can be created and explored. Consequently, only a fraction of the potential knowledge available from the testbed is realized.

Approach to Improving CART Data Collection and Analysis

The CART program is in the process of implementing a solution to the first two levels of analysis. The third level of analysis will be incorporated into the initial solution if funds become availabile. The approach to improving data collection and analysis consists of two major thrusts. This first involves a scheme for organizing performance measures to support comparison of mission outcomes and tracing impacts of lower levels of operator and system performance on higher-level mission performance. The second thrust focuses on implementing a set of tools that streamline and automate the process of defining and calculating performance measures and that support the visualization and exploration of results. Each is discussed below.

Organizing Performance Measures for Tracing Impacts of Operator Performance on Mission Outcomes

As described above, the mission decomposition conducted to derive CART simulation requirements employs a hierarchical decomposition methodology that proceeds from the high-level mission goal through mission functions and down to individual tasks performed by an operator. The decomposition produces a form of goal-means hierarchy that defines explicit linkages between a node at one level in the hierarchy and a node at the next higher level. Thus, the "path of influence" can be traced from a low-level operator task (a means) up to the mission goal. The goal-means hierarchy produced in the mission decomposition was chosen as an organizing structure for CART performance measures because it supports tracing relationships between different levels of performance aggregation. This approach is consistent with a concept for a diagnostic hierarchy proposed by Allender and Brett (1988). The diagnostic hierarchy consisted of different levels of linked performance measures that ranged from high-level measures of mission outcomes to low level measures of operator performance.

The process of specifying performance measures consists of defining measures for each node in the goal-means hierarchy. Multiple measures can be defined for a given node. The objective of measure definition is to assess critical dimensions of performance associated with a node. Examples of dimensions of performance are accuracy (e.g., percent hostiles identified correctly), timeliness (e.g., percent hostiles destroyed prior to releasing munitions), and completeness (e.g., percent hostiles identified). In computing these measures data are aggregated and summary performance measures are computed by test condition.

With these measures, it is possible to begin at the top of the hierarchy and trace mission performance impacts through the different nodes and levels to determine the low-level factors that produced the mission performance. Comparisons of data across different test conditions allow the analyst to see how different system concepts drive and influence different dimensions of operator performance.

An Integrated Set of Capabilities for Automated Performance Measure Computation and Visualization

Depending on the complexity associated with a goal-means hierarchy for a domain, an extensive set of performance measures can be defined. In an effort to reduce the effort traditionally associated with computing and manipulating these measures, a set of integrated database and software applications are being developed that automate the process. The applications in the set perform three basic functions. The first function of the applications is to describe key factors that that must be known to manipulate source data and calculate performance measures. These include the structure of source databases, the structure of the goal-means hierarchy, and the formulae for calculating a measure. Computation is the second function of the applications. A computation application uses information provided by descriptive applications to actually calculate a performance measure. Visualization is the third function performed by the applications. A visualization application uses information from the descriptive applications to present calculated performance measures in a manner in which effects and relationships can be observed readily. Figure 1 depicts the relationship between the applications. The applications are indicated by the ovals. Note that different applications are brought to bear in different phases of the CART process. Each application is described briefly below.

- Test Plan Description Application. Generally, a formal test plan guides application of a simulation testbed. The test plan specifies the issues to be addressed, high level performance measures, the independent variables to be manipulated, how the independent variables and their different levels are organized into test conditions, and the number of trials or data collection runs to be conducted under each test condition. Information from the test plan is important because it is used to form a structure that relates individual data collection runs to specific test conditions. Using the Test Plan Description Application, information about the test plan is entered into data tables. Later, during calculation of performance measures, this information is used to aggregate data from runs associated with a test condition or different combinations of the independent variables so measures can be calculated by condition.

- Goal-Means Hierarchy Application. To date, goal-means hierarchy information has been captured informally in the CART program using spreadsheets and drawings. The Goal-Means Hierarchy Application captures a description of nodes in the hierarchy and linkages between nodes at different levels and stores the description in data tables. Later, these data are used to help guide specification of performance measures and display of simulation results.

- Testbed Data Repository Description Application. An integral component of a CART testbed is the data collection and storage subsystem. This subsystem collects data from a simulation run. Multiple simulations produce the data that are collected. Integrating these data from separate simulation data files into a common, time synchronized database that identifies data associated with a run and that can be manipulated readily is a key function of the data collection subsystem post-processing software. This integrated database will be the source data from which performance measures are calculated. In order for the automated performance measure calculation software to manipulate the source data, it must understand the structure and content of the source database. This application is used to create a meta-data description of the source data database.

- Performance Measure Definition Application. This application enables users to specify performance measures of interest. It uses the goal-means hierarchy developed earlier as the structure for specifying and organizing performance measures. The primary performance measures specified in the test plan will be defined along with measures for lower level performances. The process of defining performance measures consists of naming and describing a measure but most important, it involves specifying formulae for calculating measures. The formulae specify source data elements to be used and the methods by which the data are to be manipulated (e.g., sums, means, ratios, etc.) Later, the formulae are used by measure calculation software to point it toward the data elements to be manipulated and instruct it in how to combine the data elements.

- Performance Measure Computation Application. This application automatically conducts the calculation of performance measures defined in the Performance Measure Definition application and stores the results in a database for retrieval and display later. It uses calculation algorithms from the Performance Measure Definition Application and raw data structure information from the Testbed Data Repository Description Application to access data in the Testbed Data Repository, calculate summary data and performance measures, and stores the results in a Performance Measure Database (called Perf Meas dB in Figure 1). With this application it will be possible for analysts to quickly and easily extend and evolve their performance measure set as an analysis progresses without needing to stop and obtain the support of a programmer to update their measure calculation software.

- Diagnostic Hierarchy Exploration Application. For a given domain the goal-means hierarchy and the performance measures associated with its nodes can be very complex. Understanding where in the hierarchy the independent variables of interest had an impact and how those impacts propagate through the hierarchy can be difficult. The Diagnostic Hierarchy Exploration Application is a graphical visualization capability that allows an analyst to aggregate data according to different combinations of the independent variables and explore the means-goal hierarchy from the top down to identify where key performance impacts/differences occur. Not only will this information provide the analyst with insight into which technologies, concepts, and designs to select from among the alternatives, it will enable the analyst to explain why and how they are better. A variety of features and attributes are under consideration for the Diagnostic Hierarchy Exploration Tool. These include:

  - An interface for specifying how data are to be aggregated for calculating summary performance measures. The default aggregation will be the test conditions specified in the test plan but others will be possible. Basically, the user will be able to group the data by any subset of the independent variables used in the study.

  - A "folder tree" approach to presenting the goal-means hierarchy and associated measures. This approach uses the manner in which Microsoft Windows Explorer displays folder trees associated with files on a disk drive as the means for presenting a goal-means hierarchy. Nodes in the hierarchy that have lower levels associated with them will have a "+" symbol displayed next to them. Double-clicking one of these nodes will produce an indented listing of sub-nodes. Adjacent to the node listing will be a column that lists the performance measures specified for each node. Columns adjacent to the performance measures will present the computed values for the measures. The number of columns of values presented will depend upon the data grouping specified by the user. This display provides a basic diagnostic capability in which a user can explore the goal-means hierarchy and its associated performance measures and develop an trace how lower level performances affect higher-level outcomes.

  - The ability to add performance measures "on the fly". As an analysis progresses, an analyst might gain new insights into factors that are affecting outcomes observed in the data. These insights might lead to specification of new performance measures that more directly measure the dimensions of performance that are being affected. Users will have the ability to move fluidly from the display of performance measure data for the different nodes to screens for inserting and defining new measures. Data will be re-processed quickly and the new measure will appear in the hierarchy.

Conclusions

The CART data visualization toolset is aimed at streamlining and automating the processing and display of complex operator-centric simulation testbed data and results. This capability should increase the amount of knowledge and information that can be gleaned from a simulation testbed. Equally important, the toolset uses a goal-means hierarchy as the organizing structure for defining performance measures and presenting results. The goal-means hierarchy provides linkages between low-level operator tasks and high-level system goals. These linkages can be used to trace the impact of operator performance on mission outcomes. This organization of performance measures can help the human factors practitioner explain why human factors-related technologies and issues are important to overall system performance. Finally, this scheme for structuring performance measurement around a hierarchical system decomposition should integrate readily with analysis methods currently applied by cognitive systems engineers (Rasmussen, J.A., Pejtersen, M., & Goodstein, L.P., 1994*)* and add a new capability to their toolkit.

Acknowledgments

The Air Force Research Laboratory Human Effectiveness Directorate has supported the work reported in this paper under the Combat Automation Requirements Testbed program (contract number: F41624-98-C-6012).

References

Allender, L. E. and Brett, B .E. (1988). Toward a Diagnostic Operator Performance Assessment Scheme. *Proceedings of the 30th Annual Military Testing Association Conference*.

Brett, B. E., Doyal, J. A., Malek, D. A., Martin, E. A., Hoagland, D. G. (2000) *The Combat Automation Requirements Testbed (CART) Task 1 Final Report: Implementation Concepts and Example* (Air Force Research Laboratory Human Effectiveness Directorate Technical Report Number AFRL-HE-WP-TR-2000-0009). Wright Patterson Air Force Base, OH 45433-7022.

Flach, J. M. and Dominguez, C.O. (1995). Use-Centered Design: Integrating the User, Instrument, and Goal. *Ergonomics in Design, July 1995*, 19-24.

Rasmussen, J.A., Pejtersen, M., & Goodstein, L.P. (1994*)*. *Cognitive Systems Engineering*, New York, New York: John Wiley & Sons.

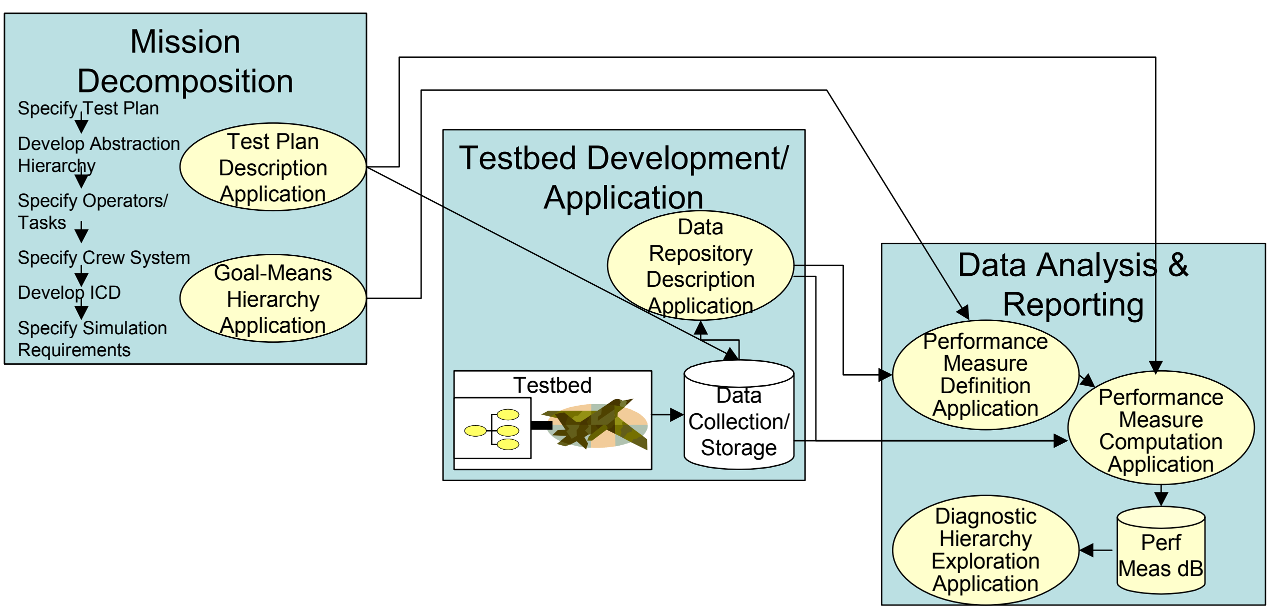

Figure 1. The relationship between applications in the CART data visualization capability.