MEASURING FIDELITY DIFFERENTIAL IN SIMULATOR NETWORKS

Pamela Woodard
Naval Training Systems Center
Orlando, Florida

Elizabeth M. Bennett and Ron Matusof
CAE-Link Corporation
Binghamton, New York

ABSTRACT

Simulator network design has changed from proof-of-principle demonstrations to a production training medium as new contracts call for the implementation of simulator networks designed to a standard protocol. The prototype standards for Distributed Interactive Simulations (DIS) address the information content required for the interconnection of dissimilar simulations and the rules for DIS usage. These rules currently center on the control and execution of network exercises and have not yet addressed the issues of system validation and system performance measurement. This paper briefly discusses current trends in simulator interoperability. It then discusses the concept of fidelity differential and its impact on team training. It discusses various methods of measuring fidelity differential and their relative merits. Finally, it suggests a set of performance metrics that should be applied to simulation networking standards.

ABOUT THE AUTHORS

Pamela Woodard is an Electronics Engineer in the Simulation Imagery at the Naval Training Systems Center where she is a visual engineer for several simulators that are being procured, including the Special Operations Aviation Combat Mission Simulator. She has over fifteen years experience in geographic and visual simulation data bases and in software development for simulation. Mrs. Woodard holds a Bachelor of Science degree in Mathematics from George Mason University and a Masters of Science degree in Computer Engineering from the University of Central Florida. She has published several papers on the subjects of software design and interoperability.

Elizabeth M. Bennett is a Systems Engineer with CAE-Link Corporation in Binghamton, NY. She has worked the last seven years in the simulation of avionics, tactical systems, and radar systems on several programs including the E-2C, F-4, and Special Operations Aviation Combat Mission Simulator. Ms. Bennett holds a Bachelor of Science degree in Electrical Engineering from the State University of New York at Buffalo.

Ron Matusof is a Staff Engineer with CAE-Link Corporation in Binghamton, New York. He has over nine years experience in various aspects of simulation design. He has spent over four years working in the development of Link's MULTISIM simulator network, including integrating the network as part of the U.S. Army's Project Desert STAARS, and he is a regular member of the DIS Fidelity, Environmental Control, and Feedback Requirements sub-group. Mr. Matusof holds a Bachelor of Science degree in Electrical Engineering from the University of Pittsburgh. He has published several papers on the subjects of interoperability, mission rehearsal, and cue correlation.

MEASURING FIDELITY DIFFERENTIAL IN SIMULATOR NETWORKS

Pamela Woodard
Naval Training Systems Center
Orlando, Florida

Elizabeth M. Bennett and Ron Matusof
CAE-Link Corporation
Binghamton, New York

INTRODUCTION: SIMULATOR NETWORKS

The recent interest in simulator networking, spawned by the success of SIMNET, may appear to have had its genesis in the late 1980's. However, development of both experimental and production simulator networks has been on-going for over 25 years. In fact, NASA used simulator networks for training and mission rehearsal of space flights dating back to the Apollo program.[1]

In addition to SIMNET, there were other simulator networks that developed independently. Although these networks provided interoperability between a limited number of players, they did not interoperate with each other.

The majority of existing simulator networks interconnect systems of similar design. For these networks, the interconnection effort is often an exercise in defining and implementing an appropriate communication architecture between the simulators. The similarity of information in each of these networked simulators greatly reduces the quantity and magnitude of inconsistencies between each simulator's representation of the environment.

The network of the future, however, will interconnect simulators of different, or heterogeneous, design. The intent is to make these differences "seamless" so that all participants perceive and act upon the environment in a consistent manner. This network will allow any simulation to interconnect to any other simulation by simply conforming to a set of data and communication protocols.

The DIS intentionally ignores the actual implementation of individual simulations and instead defines the interconnection protocol (both data and communications) between networked participants. While this allows the DIS to concentrate on the mechanics of interconnecting heterogeneous simulations, it leaves unanswered the validity of such interconnections. The design of each individual simulation directly affects how a crew perceives, and then reacts to, the environment. Reactions that are not consistent with the real-world may lead to the condition commonly referred to as "negative training" since, in a team training environment, the actions of one crew directly affect the future actions of another.[2]

Interoperability refers to the ability of two or more systems to perform a set of coordinated tasks with the expressed intent of achieving a common goal. The ability of two or more systems to interoperate is a crucial part of any team exercise. Interoperability takes on added significance in light of the stated mission of the DIS. The primary mission of the DIS is to create synthetic, virtual representations of warfare environments which can be used as a substitute for some field training and testing when cost, safety, environmental, or political constraints will not permit the field training and testing required to maintain readiness.[3] As such, the DIS standard network will be used for team training, mission rehearsals, and acceptance trial of some weapon systems (through demonstration and validation of the weapon system in a virtual environment).

Joint Chiefs of Staff Publication 1 defines an acceptance trial as, "Trial carried out by nominated representatives of the eventual military users of the weapon or equipment to determine if the specified performance and characteristics have been met".[4] It is likely that when this definition was first drafted few envisioned carrying out acceptance testing within a virtual environment. A real-world acceptance trial often takes for granted certain conditions -- namely that the laws of physics apply equally to all participants of the acceptance. In the world of simulation, one must validate that the environment is, itself, a valid recreation of the real world. Therefore, we must determine the validity of the virtual environment before ensuring that a weapon meets its specified performance and characteristics. Mission rehearsals require this type of validation as well since they are, in essence, acceptance trials of specific missions.

The concept of validating a simulation environment is not new, nor has the simulation community accepted any single method of validation. The problems facing the user of a DIS network go far beyond the validation problems of the past.

FIDELITY, FIDELITY DIFFERENTIAL, AND REALISM

Often, simulators are classified by how closely they represent the real world -- an attribute many refer to as "fidelity." Finding a single, accepted definition of fidelity is like attempting to get the population of San Antonio to agree on the best film ever made. It is a futile proposition, and one that probably will produce little useful information. Rather than using a single definition of fidelity, we have chosen to use parts of several definitions.

One common definition of fidelity describes it as the degree of similarity, both physical and functional, between a training device and the actual equipment for which the training was undertaken.[5] This definition discusses fidelity solely in terms of the equipment being used and the equipment being simulated, and avoids references to trainee's perceptions and behaviors. The intent is to provide a separate method of measuring and evaluating behaviors that are not dependent on the equipment being used.

Specifying fidelity in terms of equipment performance does not guarantee that the desired interaction between crews, equipment, and environment occurs. These interactions are based upon the *accuracy* and *realism* of the simulation. Accuracy refers to the precision with which a simulation recreates the real world. Accuracy is an objective measure specified in terms of scientific principle and mathematical equations. Realism refers to how closely a simulation actually correlates to the real world in terms of both appearance and behavior.[6] Realism is a subjective measure that is expressed by a state-of-mind --- what magicians refer to as a "willing suspense of disbelief."

Crew instruction can be broken into two distinct processes: teaching of intellectual skills and training of motor skills.[7] For crew training, instruction consists of the academic skills required for successful operation and requires *accurate* information. Controlling a vehicle requires coordinated motor skills and the associated instruction requires *realistic* presentations. It is important to note that accuracy can be obtained in an unrealistic environment (such as the touch-screen representation of cockpit controls in many Computer Based Trainers), and realism can be obtained without providing accuracy (such as augmenting target visual presentations to account for limitations in a simulator's visual display). The key is that fidelity is determined by the accuracy and realism of the simulation *for a given exercise*.

Networking of simulators adds new facets to the fidelity issue. The most prominent of these is fidelity differential. Fidelity differential is a measurement of interoperability that defines the level of difference between the operations of two or more systems on a network.[8] Early proof-of-principle simulator networks could ignore the effects of fidelity differential since these networks connected simulators of essentially identical design. Fidelity differential is likely to become an increasingly important criterion to determine those tasks that should be performed on a simulator network. Some threshold fidelity differential will determine which network tasks will produce results that can be successfully transferred to the real world.

There has been much discussion in the training community regarding the requirement to produce a *fair fight or level playing field* for network exercises. The idea is to develop a simulation network where crew and equipment performance can be measured without the fidelity differential affecting the outcome. A study using a network of dissimilar, high fidelity simulators found that the fidelity differential gave an unfair and unrealistic advantage to the simulator of lower fidelity.[9] In this case, the lower fidelity player's weapon scoring algorithms allowed the crew to score kills for weapon impacts well outside of the real-world circular error probability (CEP) while the higher fidelity player was constrained to the accuracies of the real world. Additionally, this study found that fidelity differentials between the simulations and the real world also accounted for decreased realism in certain air-to-air engagements.

The designers of multi-fidelity networks must carefully implement their system such that an un-

Fidelity must be measured with respect to a specific cue or environmental quality. Confidence that the networked simulations will be an effective training tool can only be obtained after training tasks have been defined, and fidelity for simulators in the relevant areas has been assessed. Even then, absolute time delays and other factors can affect the presentation of the correct environmental cues at the correct time. The importance of cues may be summarized as follows: "cues occurring in the present are perceived and assessed for their meaning and impact at both the current time and for an anticipated time, even though that time may only be an instant away." [6] Since the goal of networked simulations is to increase the realism and complexity of the training environment (in part so that higher cognitive and decision skills may be improved) the perception of the situation at any time (based on cues) is very important.

The BDS-D program views this problem in terms of an objectives versus cost trade-off. The basis of this trade-off is referred to as fidelity anchoring (originally defined by the Institute for Defense Analysis). Fidelity anchoring is an approach to deciding how a simulation should be designed to meet its objectives with minimum cost.[16] Under this scheme, fidelity is defined in terms of effectiveness, user acceptance, and affordability, all of which are subjective measurements. The desired outcome of fidelity anchoring is that every component of the simulation should have the exact degree of fidelity required by its intended application. Unfortunately, fidelity anchoring cannot be fully assessed until after a system has been produced, and therefore is difficult to use in a procurement specification.

To produce a procurement specification for network interoperation, fidelity differential must be measured relative to known specific qualities of the individual simulations comprising the network, and of the network interconnection itself. Depending on the task being trained, measurement of fidelity with respect to any one of the following environmental qualities may be necessary:

Simulation Characteristics such as update rate, dynamic response, system latency, etc.

Network Characteristics such as simulator to network latency, network node-to-node latency, network delay dispersion (difference in network delay across time), etc.

Simulation Accuracy which includes the physical replication of the crew station, handling qualities of the vehicle dynamics, correlation of the simulated terrain with the real world, scoring accuracies for simulated weapons, etc.).

Simulation Realism including the subjective measurements of training effectiveness, user acceptance, mission performance, and objective measurements of correlation with other player's environments.

Fidelity measurement may be achieved by self-reporting systems such as the "Data Logger" used to help assess the AIRNET device. AIRNET is the rotary wing version of SIMNET, and is intended to be a low cost, low fidelity interactive networking simulation.[17] Recorded data includes player vehicle position, weapons data, and kill ratios. A set of performance metrics for this network is also presented, based on "operational relevance, shooting, moving and communicating, and sensitivity to mission effectiveness, who won the war?" These metrics include threat interactions, tactical piloting and navigation, communications, and battle resource management.

CONCLUSIONS

A network of dissimilar simulations has enormous potential for both training and operational evaluation. The network will inherently have limitations for which the user must account and compensate. The concept of providing a benchmark for users to determine interoperability is necessary. The benchmark by itself, however, provides little useful information. A benchmark is supposed to be a standard measure of performance that enables two simulations to be compared with each other. For any application it is performance on that application that counts, and benchmarks are relevant only so much as they resemble it.[18] Therefore, care must be taken when applying these benchmarks or the user has a significant risk of running a simulation without validation or correct response.

Four factors affect how well simulations will interoperate in a networked environment and how level the playing field is. They are:

1. The fidelity of individual simulations.
2. The differences in simulation fidelities.
3. The accuracy of the simulations.
4. The realism of the simulations.

The allowable fidelity of the individual simulations, the fidelity differentials, and the accuracy and realism of the network environment will all vary dependent on the training objectives desired. It is important that we define a method by which the interoperability of the network can be measured, and how these measurements can be applied to a particular network exercise.

All missions are not the same, and therefore the required fidelity of interactions in a network between the simulated environment, other crews, and the own vehicle varies between exercises. Determining the allowable tolerance for fidelity differential of a particular network involves considering the training objectives for the simulation, the technical impact of providing varying degrees of fidelity and correlation, and the associated costs. Meeting these challenging and often conflicting considerations will require analysis of the trade–offs.

The FAA utilizes a Simulator Evaluation Specialist to determine compliance with the Federal Aviation Regulations (FAR) Part 121 requirements established for airplane simulators. The Simulator Evaluation specialist is an FAA technical specialist trained to evaluate simulators and to provide expertise on matters concerning airplane simulation.[19] The FAA evaluation uses both objective and subjective testing methods to determine if a simulator is capable of meeting a particular set of tasks. Recognizing that all simulators are not intended to perform the same function, the FAA has defined several levels of simulation that have different criteria for evaluation.

Networked simulations can be evaluated in a similar manner. The individual network entities and the network itself can be evaluated by qualified subject matter experts to determine its ability to perform a specific set of tasks. If the network is found to adequately perform these tasks, it would be rated as qualified for that set of tasks. A network, under this scheme, could have multiple ratings, if it were intended to perform multiple tasks. Note that the network, in this case, refers to both the interconnection of simulators as well as the simulations themselves. Addition or deletion of any single simulation would require the network to be requalified.

We recommend that a mechanism and an associated governing organization be established, perhaps under the auspices of the DIS, DMSO, IEEE to classify devices relative to fidelity, fidelity differential, accuracy, and realism, and to provide benchmarks to determine which devices can faithfully operate together for a given task. The governing organization must be composed of qualified, subject matter experts, who can evaluate the network from both an objective and subjective standpoint. As a starting point, we recommend modifying the fidelity anchoring scheme currently being used for the BDS–D to account for the simulation characteristics, network characteristics, accuracy, and realism of the networked exercise.

The DIS standard represents an important step for the simulation community. It provides a viable means of using simulation resources, be they for operational testing or training, in an increasingly realistic manner. But the use of the DIS standard will only be valuable if the network interconnections can be validated. The measurement of fidelity differential is an important part of validating network interactions. Without an understanding of fidelity differential and its impact upon team exercises, the DIS will amount to little more than a proof–of–principle.

REFERENCES

[1]McCafferty, R., Merrit, N., and Lovitch, A., "Dress Rehearsal for a Lunar Voyage", TNB General Precision Aerospace, Volume 8, 4th Quarter, 1965, Little Falls, NY, December, 1965.

[2]Grodsky, M.A., Moore, H.G., and Flaherty, T.M., "Crew Reliability During Simulated Space Flight", AIAA/AFLC/ASD Support for Manned Flight conference, Dayton, OH, April, 1965.

[3]"Distributed Interactive Simulation Operational Concept [Draft]", University of Central Florida Institute for Simulation and Training, February, 1992.

[4]Department of Defense Dictionary of Military and Associated Terms, Joint Chiefs of Staff Publication 1, Government Printing Office, Washington, DC, June, 1987.

[5]Hays, R.T., "*Simulator Fidelity: A Concept Paper*", Army Technical Institute Technical Report 490, Alexandria, VA, November, 1980.

[6]Miller, D.M., "Foundations for Tactical Training: A Challenge to Industry", AIAA Flight Simulation Technologies Conference, Boston, MA, August, 1989.

[7]Barnes, A.G., "The Compromise Between Accuracy and Realism in Flight Simulation", AIAA Flight Simulation Technologies Conference, New Orleans, LA, 1991.

[8]Knight, S.K., "Issues Affecting the Networking of Existing and Multifidelity Simulations", Second Workshop on Standards for Interoperability of Defense Simulations, Kissimmee, FL, January, 1990.

[9]Dees, J.W., and Cornett, T.R., "Simulator Networking in Helicopter Air-to-Air Combat Training", AIAA Flight Simulation Technologies Conference, Boston, MA, September, 1989.

[10]Sawler, R.J., and Matusof, R., "Issues Concerning Cue Correlation and Synchronization of Networked Simulators", AIAA Flight Simulation Technologies Conference, New Orleans, LA, August, 1991.

[11]Universal Threat System for Simulators (UTSS) Front End Analysis (FEA) Volume I, Information Spectrum Incorporated, Contract Number N00600-89-D-0262, Falls Church, VA, September, 1990.

[12]Cardullo, F.M., and Brown,Y.J., "Visual System Lags: The Problem, The Cause, The Cure", IMAGE V Conference, Phoenix, AZ, June, 1990.

[13]Roscoe, A.H., and Ellis, G.A., "A Subjective Rating Scale for Assessing Pilot Workload in Flight: A Decade of Practical Use", Royal Aerospace Establishment Technical Report, March, 1990.

[14]Bruce, R.C., "Simulation Fidelity: A Rational Process For Its Identification and Implementation", *Interservice/Industry Training Systems Conference*, Fort Worth, November, 1989.

[15]Cross, K.D., and Gainer, C.A., "An Enumeration of Research to Determine the Optimal Design and Use of Army Flight Training Simulators", U.S. Army Research Institute for the Behavioral and Social Sciences Technical Report 763, October, 1987.

[16]Anton, J., "Fidelity Validation Issues for BDS-D (SIMNET/AIRNET)", *Selective Fidelity and Validation Issues for Rotorcraft Flight Simulation*, Atlanta, GA, April, 1992.

[17]Thomas, B.W., and Gainor, C.A., "Simulation Networking: Low Fidelity Simulation in U.S. Army Aviation", Royal Aeronautical Society Progress in Helicopter and V/STOL Aircraft Simulation Conference, London, UK, May, 1990.

[18]*Shepherd, R., and Thompson, P.*, "Lies, Damned Lies, and Benchmarks", Inmos Corporation, Technical Note 27, July, 1987.

[19]Federal Aviation Administration, "Advisory Circular 120-40B: Airplane Simulator Qualification", U.S. Department of Transportation, Washington, DC, November, 1990.

acceptable cue correlation error is not introduced to a higher fidelity player because it is networked with a lower fidelity player. The fidelity differential can introduce correlation errors in three ways:[10]

1. Fidelity of network data is insufficient to support the required cue correlation.
2. Update rate of network traffic is insufficient to support the required cue correlation.
3. Fidelity differential between representations of environment models (navigation, terrain, etc.) exceeds the tolerance of the required cue correlation.

MEASURING FIDELITY AND FIDELITY DIFFERENTIAL

Measurement of the fidelity of a particular simulation depends upon which definition of fidelity is used and how it is applied. A general perception is associated with fidelity, such that we assume a high-fidelity solution is somehow better than a low-fidelity solution. Fidelity is a relative measurement, and must be determined in reference to a particular task or set of tasks. An engineering simulator is generally a high-fidelity replication of a real-world system, yet might be considered a low fidelity training medium. Often, it is not fidelity that we are trying to measure, but the realism or accuracy of the simulation, of which fidelity plays a part.

The Universal Threat Systems for Simulators (UTSS) program commissioned a front end analysis that studied the issues of fidelity and validation.[11] This study found that fidelity may be defined in several ways:

1. How well the appearance and behavior of an object are represented.
2. Which modeling approach (analytic, dynamic, emulation) was chosen.
3. How well the model matches real-world data.
4. How realistic the user perceives the simulation to be.

These measurements, however, do not always produce consistent results. An analytical approach to modeling (providing a pre-defined response to a given input, possibly by means of lookup tables) was described by the UTSS study team as low-fidelity (using the second UTSS definition of fidelity) while a dynamic approach to the simulation, in which modeling of subassemblies and internal interfaces is performed, was considered high-fidelity. However, for cases where limited data is available, the analytical approach will most likely be the most accurate model, and therefore high-fidelity if one chooses the third UTSS definition of fidelity.

One might consider determining simulator fidelity by examining the effect of specific cues on the performance of a particular task. These same cues would then be compared with their real-world counterparts to determine correlation. A composite score could then be calculated to determine overall fidelity. There are generally accepted methods of measuring the correlation between a simulated system's performance and the performance of its real-world counterpart.[12] The more controversial issue is concerned with measuring the subjective quality of the simulation — perceptual fidelity.

One of the most widely used measures of perceptual fidelity in flight simulation is the Cooper-Harper rating. This rating scale assigns a value of one to ten based on aircraft characteristics and pilot workload for the selected task. The pilot performing the flying task uses a decision tree and rating descriptors to assign the rating value.

The Bedford scale, a modified Cooper-Harper type rating (in conjunction with heart-rate monitoring), has been proven effective for measurement of pilot workload levels.[13] Such subjective ratings may be effective for simulation fidelity measurements if they are modified to pertain not to vehicle handling qualities or workload assessment, but to the realism and accuracy of certain qualities of a simulated system. One suggested method of subjective measurement is the rating of two factors: the physical characteristics, and the informational or stimulus and response options of the equipment.[5] These two measurements, for a given task, provide a fidelity factor. AAI's fidelity verification model[14] uses ratings in three categories for each training task: task commonality, physical similarity, and functional similarity. A device fidelity index can then be derived from the individual ratings.

Ratings for specific training tasks have been used to assess aircraft qualities, and a similar approach could be effective for simulation fidelity measurement. The compilation of a list of training tasks and conditions was considered essential for a study of simulation fidelity required for effective training transfer.[15] In general, the tasks to be trained in a simulated environment are known by the end-user and drive requirements and specifications for procurement purposes.