Fidelity Assessment within the REVVA Generic Process

Manfred Roza
Martijn van Emmerik
Jeroen Voogd
TNO Defense, Safety and Security
P.O.Box 96864, 2509 JG
The Hague, The Netherlands
manfred.roza@tno.nl, martijn.vanemmerik@tno.nl, jeroen.voogd@tno.nl

Keywords: Validation, Fidelity Metrics, Fidelity MoP , Fidelity MoE, REVVA Generic Process

ABSTRACT: *The REVVA methodology, developed in the WEAG THALES Joint Program 11.20, provides a framework for verification and validation activities resulting eventually in an acceptance recommendation for data, models, and simulations. The core of the methodology is the decomposition of the 'intended purpose' into acceptability criteria (AC). The V&V effort focuses on determining whether these criteria are met. The AC are made quantifiable by Measures of Effectiveness (MoE), and through a further decomposition into Measures of Performance (MoP) and Dimensional Properties. It often turns out to be a highly subjective effort to quantify the items of evidence underlying these various measures.*

Around the turn of the century SISO sponsored two study groups on a common fidelity definition and implementation. The work of these study groups resulted in two significant reports stressing the importance of fidelity definition and measurements within any VV&A process. Based upon this work, amongst others, TNO Defense, Safety and Security developed a methodology to systematically define, specify and measure simulation fidelity.

This paper presents a practical approach how the TNO fidelity methodology can be used within the REVVA Generic Process to provide items of evidence that support the simulation validity and acceptability criteria. The approach is illustrated with some results and lessons learned from a practical case-study.

1 Introduction

In today's society, simulations are more and more used in crucial decision-making processes during engineering design, test and evaluation, acquisition of new systems, and in training and instruction. This increasing reliance on simulation results, the greater complexity and more frequent reuse of simulations have caused a great interest in comprehensive methods for verification, validation and accreditation (VV&A) [7] [18]. The availability of such comprehensive methods is essential in assessing the credibility of the outcomes of future simulations.

Since the early development of simulation technology, the notion of fidelity, i.e. how well a simulation relates to reality, has been an apparent, critical and recognized part of modeling and simulation (M&S) validation [18]. Despite this observation, the ability to systematically quantify and utilize fidelity characterizations within the VV&A process is even today a largely uncultivated area.

Both these concerns have been expressed by a wide variety of M&S user communities [6] [7] [8].

In 2003 this has been one of the reasons for the Western European Armament Group to initiate a research project to develop a common methodological framework for the VV&A of models and simulations [9]. This project named REVVA was funded by five European nations. On behalf of the Netherlands TNO Defense, Safety and Security participated in this project. The developed REVVA methodology is currently further improved and formalized in the follow-on project called REVVA2.

In the late nineties TNO also participated in the two SISO study groups on fidelity definition and implementation [8] [15]. Based upon the results from both these study groups, amongst others, TNO and the Delft University of Technology cooperated in de development of a common methodology to systematically define, specify and measure simulation fidelity [18]. All this was done from a general simulation system life-cycle perspective, not limited by any specific problem domain aspect, approach or process.

This paper presents a possible way to integrate the major concepts, metrics and recommended practices of the TNO fidelity methodology within the REVVA Generic Process. The objective of this paper is to illustrate how both methodologies complement each other and can enhance their practical applicability.

The paper starts with a summery of the major elements of the REVVA methodology (Section 2). Next in Section 3 a brief overview of the basic elements of the TNO fidelity methodology is given. Section 4 introduces the reader to the TNO fidelity methodology concepts, metrics and techniques, which can be used within the REVVA generic process model to enhance the definition and assessment of the acceptability and validity criteria along with the supporting items of evidence (Section 5). To make this fidelity assessment within the REVVA generic process more explicit the paper proceeds with a practical simulation VV&A case-study executed for the Royal Dutch Army (Section 6). The paper ends in Section 7 with a summary of the major results and directions for future research.

2 The REVVA Methodology

The REVVA methodology has been the result of a joint WEAG defense research program between France, Italy, Denmark, Sweden and the Netherlands. The objective of this program was to develop a common integrated set of VV&A methods, techniques and standards, which are used to determine and assess the confidence in the intended user purpose of a M&S product. In this section the three major building blocks, also known as the pillars, of the REVVA methodology will be discussed. For more background information about the underlying assumptions and key concepts of these pillars the reader is referred to [10].

2.1 Pillar One: The Organization

This pillar identifies all participating parties and their roles in the REVVA generic process (Paragraph 2.2). Parties can be an organization or divisions of organizations. REVVA utilizes four different categories of parties. The first one is the *customer* or the organization that uses a M&S solution. The second is the *supplier* or the developer of a M&S solution. The third category is a *third party VV&A agent*, an external and independent VV&A organization. The last party is the *acceptance authority* or an external and independent trustful organization.

Actors from the above parties can play various roles in the REVVA generic process. Each of these roles are described in terms of required knowledge and skills to complete the assigned tasks, the actor's authority and responsibility in the process, and the interaction with other roles. The major active VV&A roles identified by the REVVA methodology are the end-user, acceptance leader, V&V leader and V&V executioners. Besides these roles also affected roles are defined, which are those actors taking advantage of the REVVA methodology but are not directly involved in the implementation of the REVVA generic process.

2.2 Pillar Two: The Generic Process

The heart of the REVVA methodology is the generic process model. This VV&A process model consists of seven iterative phases that lead to official acceptance of a M&S solution for the intended purpose. Each phase is defined in terms of activities that have to be performed plus the involved roles and their type of involvement in these activities. The process also specifies both the input and output products for each phase (Paragraph 2.3).

The generic process starts with the development of the target of Acceptance (*ToA*) based upon the intended purpose of the M&S solution. In the second phase all available information about the system of interest and the simulation model is acquired. This knowledge is used and improved in all the other phases. From the ToA a Target of Validation and Verification (*ToVV*) is developed in phase three. Using the ToVV, the actual validation and verification is conducted in phase four, which followed in phase five by the assessment of the probative force of these V&V results. The results of this last phase are combined in phase six to assess whether the acceptance criteria in the ToA are passed. Based upon these outcomes a recommendation whether to accept or reject the M&S product is developed in phase seven. A full description of the generic process model phases can be found in [12].

2.3 Pillar Three: The Products

The third REVVA pillar is a description of all products developed in the above presented generic process model. In this paper only the four REVVA products are discussed that are relevant in the context of fidelity assessment.

The *Model Information and System Knowledge* is used as the basis to determine whether the simulation has a correct and valid representation of the system of interest. It is a collection of knowledge of both the real system and the simulation model identified and available during the VV&A process. So far the REVVA methodology only provides a very high level of classification of types and how information could be specified.

The Target of Acceptance (ToA) and the *Target of Validation & Verification (ToVV)* together with the *Items of Evidence* (*IoE*) form the core and most innovative part of the REVVA methodology. VV&A efforts address the question whether a simulation representation of reality fits its intended purpose [6] [8]. A ToA is a directed acyclic graph of (sub)objectives with its root being the intended purpose statement and its leaves being precise measurable acceptability criteria (*Figure 2-1*). In this sense the ToA specifies what must be assessed in terms of measures of effectiveness (MoE). Furthermore, the ToA states for each acceptability criterion the impact on its parent nodes up to the intended purpose if the criterion isn't met. A more detailed description of the ToA can be found in [4].

Like the ToA the ToVV is a directed acyclic graph with its roots being the acceptability criteria of the ToA. The ToVV specifies how for each acceptability criteria it is demonstrated whether a criteria is or isn't met. Compared to the ToA the focus of the ToVV is not on the user but on the internal model and simulation system structure, their characteristics and behavior. The leaf-nodes of the ToVV specify what V&V tasks have to be performed and what items of evidence, in terms of Measures of Performance (MoP), need to be collected.

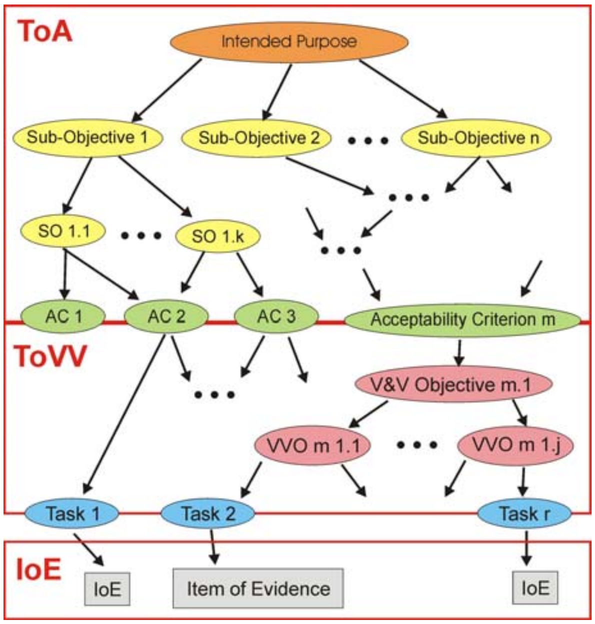

Figure 2-1 ToA, ToVV and IoE relationships

The IoE resulting from the executed ToVV tasks provide all the evidence for the claim that an acceptability criterion has passed. To substantiate such a claim it is necessary that for each item of evidence a judgment of it's probative force is stated. Based upon the integration of these IoEs and their probative force it can be assessed how convincingly the collected evidence for the acceptability criteria states that the criteria have passed. A more detailed presentation of the IoE in relationship with to the ToVV and ToA can found in references [4], [10] and [12].

3 The TNO Fidelity Methodology

From their collaborative research into distributed simulation technology, TNO and the Delft University of Technology (DUT) developed concerns regarding fidelity in relationship to the credibility, reuse and interoperability of simulations. Concerns similar to those expressed elsewhere in the M&S community. This resulted in two fidelity study groups sponsored by SISO [8] [15]. Based upon their participation in these study groups and their own experience, TNO together with DUT embarked on a new fundamental research project on simulation fidelity theory and practice [13] [14] [16] [17]. This research eventually resulted in a common fidelity methodology (TNO-FM) to systematically define, specify and measure simulation fidelity. This section gives a brief overview of the fidelity methodology. For a complete description the reader is referred to [10].

3.1 Fidelity Definition, Concepts and Principles

As shown by the SISO fidelity study group results there exist many connotations of what is considered to be the fidelity of a simulation. The common denominator in all these fidelity connotations is the ability to make a comparison between reality and the simulated reality and a specification of the degree of correspondence between both. It is this specification that is often called the degree of realism and this is considered within TNO-FM as the most conceptual right for the term fidelity. Based upon this premise fidelity is define by the TNO-FM as: '*the inverse difference between reality and simulated reality*'. A very compact and context free formulation for fidelity, which explicitly captures the essential principle underlying its measurement; which is the comparative analysis between reality and simulated reality. The inverse difference satisfies the normal apprehension that high fidelity implies a small difference.

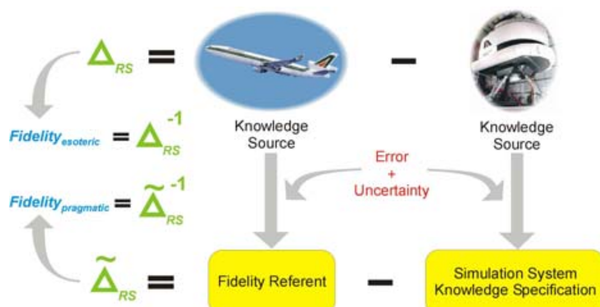

Figure 3-1 Esoteric versus pragmatic fidelity

A practical implementation of this fidelity definition is never attainable in real-life. The main reason for this is the limited human capability to directly and objectively compare reality against the simulated reality through his own sensory systems or some technical equipment. This means that fidelity can only be measured indirectly by a comparative analysis of knowledge specifications for both reality and simulated reality. A knowledge specification is always the result of a process that comprises observation, experimentation, processing, abstraction, interpretation and specification. Such a process is always error prone and will contain uncertainties. Uncertainty in this regard is characterized as incomplete or lacking information, which limits the exact correctness with which any kind of real-world knowledge can be known [11] [20]. Reality and simulated reality are thus knowledge sources. This yields a weaker but pragmatic definition for fidelity being, '*The formal specification of the inverse difference between the referent and the simulation knowledge specification*'. It is this definition of fidelity that is used as the basis of the TNO-FM. The difference between both fidelity formulations is visualized in *Figure 3-1*.

In this formulation of fidelity the referent is a codified, structured and formal specification of the real entity that is simulated. This concept of the fidelity referent is similar to the one developed by the SISO study groups [8]. The TNO-FM provides a generic template for a referent, which builds upon Zeigler's hierarchy of system knowledge specification and complemented with additional elements for managing its contents [17] [24]. The simulation system knowledge specification is concept similar to the fidelity referent except it contains knowledge about the simulation system itself and its representation of reality. A description of the associated TNO-FM template for this knowledge specification can be found in [18]. Fidelity measurement now becomes an activity of quantifying and qualifying the inverse difference between both knowledge specifications and the assessment of the errors and uncertainties related to these inverse differences.

This formulation of fidelity (F) is formally expressed in the TNO-FM by the following sextuple:

$$F = \left\langle R_{ref}, S_{spec}, C_{\Delta_{RS}}, \Delta_{RS}^{-1}, \delta\Delta_{RS}, U\left(\Delta_{RS}^{-1}\right)\right\rangle$$

In here $C_{\Delta_{RS}}$ is a set containing all practical available and applicable fidelity evaluator functions that quantify and/or qualify the difference between the referent (R_{ref}) and simulation knowledge specification (S_{spec}). Fidelity evaluator functions are the implementation of one of TNO-FM's twelve basic theorems which state that fidelity is not a singular value but an enumeration of various multidimensional and multifaceted metrics. Δ_{RS}^{-1} is the set of inverse proportional scaled differences resulting from executing all fidelity evaluator functions in $C_{\Delta_{RS}}$, $\delta\Delta_{RS}$ and $U\left(\Delta_{RS}^{-1}\right)$ are respectively the error and uncertainty of each measured (inverse) difference.

3.2 Fidelity Characterization and Taxonomy

The application of Zeigler's hierarchal system knowledge specification in the TNO-FM enables a uniform and coherent characterization of fidelity at different levels of abstraction or (de)composition. This results in a fidelity characterization in terms of the next eight descriptive concepts: *detail, resolution, accuracy, interaction, temporality, causality, precision* and *sensitivity* [8] [15] [17]. Together, these concepts outline the basis for the development of practical fidelity evaluator functions or in short fidelity metrics. The TNO-FM comprises a basic taxonomy with the most common and elementary fidelity metrics. Top down the taxonomy classifies fidelity metrics into quantitative and qualitative metrics (*Figure 3-2*). Each class is subdivided in structural and behavioral metrics.

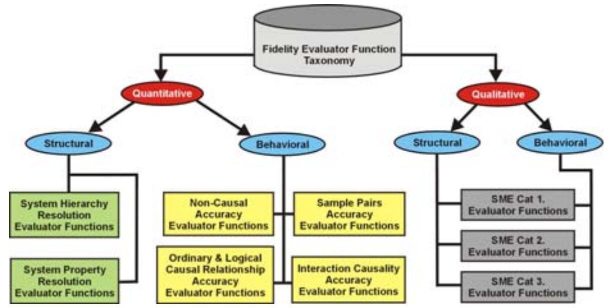

Figure 3-2 The TNO-FM fidelity metrics taxonomy

Structural fidelity metrics objectively characterize and quantify the simulation fidelity in terms of resolution or the level of detail. These metrics are the most course grained metrics to quantify the difference between reality and simulated reality in terms of whether (sub)systems and/or their characteristics properties and interactions are represented by the model or simulation.

Behavioral oriented fidelity metrics focus on the pair-wise comparison between the specified behavioral knowledge found in the fidelity referent and the simulation system specification knowledge base. These metrics and methods objectively characterize and quantify the simulation fidelity in terms of accuracy of the behavioral replication of the complete real-world system and/or its subsystems. Following a similar approach as proposed by Birta [3], the behavioral fidelity metrics are further catalogued by the taxonomy in the following four categories: non-causal system behavior, behavior samples, ordinary causal and logical relationships, and interaction causality metrics.

Although quantitative metrics are preferred, the use of qualitative or subject matter expert (SME) based methods and metrics are usually inevitable. These qualitative metrics are structured in the taxonomy according to three categories (*Figure 3-2*) in which both real-world referent and simulation system knowledge is elicited and interrelated by an SME. Furthermore, the TNO-FM provides a formal and systematic approach, consisting of several statistical analysis methods and expertise level ratings, to increase the reliability and repeatability of subjective SME evaluations.

4 Useful TNO-FM Results for REVVA

Both the TNO-FM and the REVVA do propose full methodologies for VV&A. Since REVVA is still under development and targeted for larger user community, the concepts and results of the TNO-FM can be used to fill some of the current blanks in REVVA. Hereby, improving the REVVA methodology. The next paragraphs discuss how the TNO-FM results can address the following open issues in REVVA. There exist in REVVA:

- No formal definition for validity.
- No precise definition for all REVVA products.
- No aids to integrate evidence in the ToA/ToVV.

4.1 Validity in terms of Fidelity Requirements

Fidelity and validity are closely related but are no synonyms. Comparison of the TNO-FM definition for fidelity and the commonly known definitions for simulation validity [2] [6] [7], allows for expressing validity in terms of fidelity as follows: *'A simulation is considered to be valid if its level of fidelity is sufficient for its intended purpose'*. This implies that a model or simulation is said to be valid when the actual achieved level of fidelity meets the required level of fidelity. In this section the formal TNO-FM definition for fidelity requirements is presented along with two other VV&A related fidelity concepts: fidelity measure of performance and effectiveness.

The development of a simulation system, like any other system, starts with the identification of the user needs that arise from a problem, question or deficiency (Section 2.3). Besides the regular (non)functional requirements a simulation has a special class of requirements, which specify the required real-world representational and behavioral capabilities to achieve the users needs. These kind of requirements are know as fidelity requirements [8]. Based upon the formal definition of fidelity presented in paragraph 3.1, fidelity requirements ($F_{required}$)in the TNO-FM are formally defined by the following triplet:

$$F_{required} = \left\langle R_{ref}^{req}, \hat{C}_{\Delta_{RS}}, T_{\hat{C}_{\Delta_{RS}}} \right\rangle$$

Where R_{ref}^{req} is the required fidelity referent. This is an essential element since it specifies how and from what sources the real-world reference knowledge is elicited, and what the quality of this real-world knowledge should be in terms of allowed error and uncertainty. The second element, $\hat{C}_{\Delta_{RS}}$, is the set of required fidelity evaluator functions, which covers those real-world aspects that have to be evaluated or measured during the simulation fidelity specification of the target simulation system. The last element, $T_{\hat{C}_{\Delta_{RS}}}$, is the set of fidelity tolerances, which specify for each fidelity evaluator function the tolerated deviations of the simulated representations and behaviors from its fidelity referent counterparts. These tolerances are specified in terms of an upper u_{bound} and lower bound l_{bound}.

Together the fidelity tolerances enclose an n-dimensional space inside which all measured deviations from the fidelity referent must reside to properly fulfil the simulation application purpose. Hereby validity in the context of simulation becomes a transparent Boolean proposition in terms of the fidelity adjectives *required* and *available,* which is either true or false. For a fictive one-dimensional fidelity case validity reduces to the following Boolean proposition:

$$l_{bound_1} \leq c_{\Delta_1}\left(R_{ref}^{req}, S_{spec}^{available}\right) \leq u_{bound_1}$$

This Boolean proposition is true when the measured difference between the required fidelity referent and the available simulated reality $S_{spec}^{available}$ resides within the required fidelity tolerances.

4.2 Fidelity Templates for REVVA products

As discussed in paragraph 2.3. the model and system knowledge specification is one of the essential REVVA products. Except for a top-level description its is up to the REVVA user to define how and what knowledge should be specified. The TNO-FM real-world reference and simulation system knowledge specification together serve a similar purpose and do contain similar types of knowledge. Within TNO-FM both these products are defined by means of advanced and interrelated template structures [17][18]. Compared to REVVA these templates provide the user much more details on what and how knowledge must be collected and structured. Furthermore, these templates are easily translated in equivalent spread sheets and XML-DTD's for support tool implementations.

At the terminal nodes of the ToA/ToVV a clear task description is needed of what and how IoE's must be elicited and evaluated. So far REVVA lacks guidelines on how to formulate these tasks. These tasks are similar to defining fidelity requirements in the TNO-FM as was described in the previous paragraph. The TNO-FM also provides a detailed template structure for this purpose along with a set of guidelines to populated this structure [14]. To this extend, the TNO-FM furthermore provides a taxonomy of fidelity metrics (Paragraph 3.2) that serves as a repository of readily to be used measures. Measures which provide well defined fidelity IoE's in a quantified fashion for the ToA/ToVV.

4.3 Fidelity Effectiveness for Integrating Evidence

In M&S it is common practice that one often has to evaluate several models or simulation system for their validity. Usually certain parts of the fidelity requirements are not fully met by a single model or simulation. In other words there isn't a perfect alternative that meets the validity proposition outlined in paragraph 4.1. Therefore, one has to select the simulation alternative that best fits the fidelity requirements instead. This selection not only involves the integration of fidelity measurement outcomes at different aggregation levels of interest but also evaluation of the relative importance of each fidelity requirement in achieving the intended simulation purpose. Although REVVA is focused at the front-end validity and acceptability of simulations, the way in which evidence is aggregated through the ToA/ToVV is comparable to the TNO-FM approach for aggregating fidelity scores.

In TNO-FM the found fidelity outcomes are aggregated by means of the so-called fidelity effectiveness measure $F_{\Delta_{RS}}^{effective}$. This dimensionless score characterizes the degree of validity of the alternatives, i.e. how effectively the available fidelity of a simulation and/or its constituent components suit the given specific set of fidelity requirements. The basis for this fidelity effectiveness is a rating function, $f_{\Delta}^{effective}$, which specifies how well each fidelity evaluator function result meets its associated required tolerances. The range of any fidelity effectiveness rating function must vary from one to zero. A value of one indicates that the requirement is fully achieved. The higher the effectiveness rating the better fulfillment of the requirement. Therefore a Fidelity effectiveness rating function provides a means to rate the impact or severity of the relaxation of fidelity tolerances in the simulation application purpose context. Some typical examples of such rating functions are shown below.

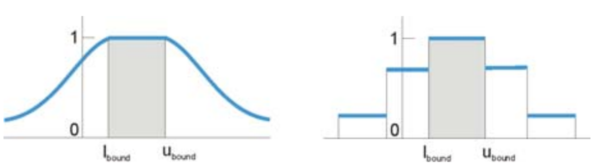

Figure 4-1 Fidelity effectiveness rating function Samples

Using the fidelity rating function the fidelity effectiveness is formally specified as follows:

$$F_{\Delta_{RS}}^{effective} = \sum_{j=1}^{k} w_j \cdot f_{\Delta_j}^{effective}$$

$$with$$

$$k = |\Delta_{RS}|$$

In here w_j is defined as the j^{th} weight of the normalized relative importance weight vector for the i^{th} fidelity evaluator function output vector:

$$W_{\Delta_{RS}} = (w_1, w_2, \ldots, w_k)$$

These weights are assigned by means of SME judgments using the Saaty's pair-wise comparison scale [1]. Fidelity effectiveness measures can be constructed on various aggregation levels and for different sub-areas of interest within the context of a specific simulation purpose. Therefore, fidelity effectiveness measures provide a useful means to assist in integrating evidence within the ToA/ToVV structure.

5 Applying TNO-FM Products in REVVA

The TNO-FM provides a management process model, which specifies a series of practical guidelines to assist in developing and applying the previously discussed TNO-FM results [13] [18]. Base-upon this model this section discusses how the TNO-FM products can be applied in the REVVA process.

5.1 Develop a ToA, ToVV and Acquire Information

From a fidelity perspective the first three phases of the generic REVVA process primarily comprise the development and formal specification of the simulation fidelity requirements. The REVVA process starts with the ToA development by identifying the objectives from the intended purpose of the simulation. These objectives can be expressed in terms of abstract fidelity acceptability criteria, which give an initial textual specification of the fidelity that's required. The recommended practice from the TNO-FM for this phase is the specification of the simulation experimental frame in terms of the simulation execution scenarios with the help of the customer party. This is because scenario's identify the major and coarse-grained structural, behavioral and other functional aspects of the desired real-world from a situational and operational perspective. These scenario's form the starting point for acquiring all required information and knowledge about the reality to be simulated to populate the fidelity referent. Acquiring information about reality is a separate phase in the current version of the REVVA process. Recommended practices and experiences with the TNO-FM show that elicitation of real-world knowledge for a fidelity referent is not a sequential activity. Instead it is an iterative process which is best conducted concurrently throughout the whole ToA/ToVV hierarchy development. The type and quality of the fidelity referent knowledge is selected according to the prioritized objectives in the ToA. This also involves assessing the nature, magnitude and impact of possible uncertainties $U(\Delta_{RS}^{-1})$ and errors $\delta\Delta_{RS}$.

With the use of the first version of the fidelity referent the fidelity effectiveness measures can be identified or developed from the fidelity acceptability criteria at the bottom of the ToA. Depending on the type and complexity of such a fidelity acceptability can be either directly specified as a ToVV fidelity effectiveness node with the underlying fidelity evaluator functions (IoE) or decomposed into a hierarchy of derived fidelity effectiveness nodes. In the latter case a relative importance weight allocation must be made to properly aggregate the result of each separate fidelity effectiveness node into a parent node. Furthermore, suitable fidelity tolerance must be allocated along fidelity effectiveness rating functions that specify the impact and severity of not fully meeting the tolerances. Both activities depend on the evaluation and associated risks analysis with respect to the parent objectives and intended simulation purpose.

Next for each fidelity evaluator function the exact measurement activities are specified. Which comprises aspects such as what referent data is used, how the simulated data is acquired, initial and post conditions and metric algorithm description. If this activity is completed a full set of fidelity requirements has been developed in the form of a REVVA ToA/ToVV structure (*Figure 5-1*).

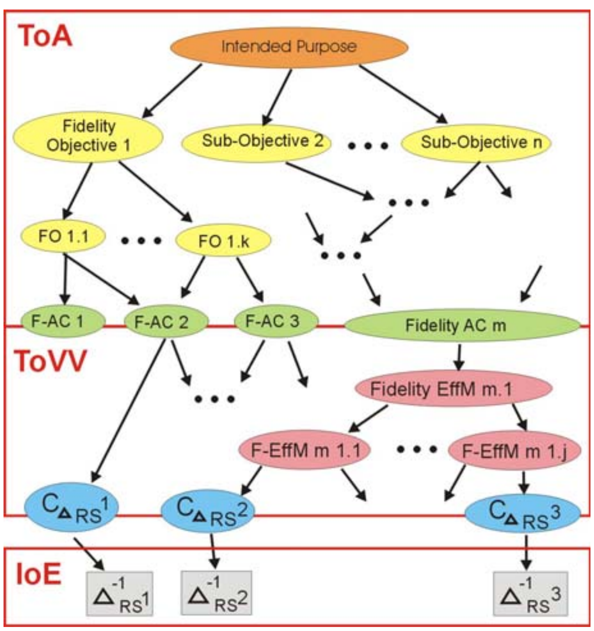

Figure 5-1 A Fidelity based ToA/ToVV

5.2 Conduct V&V, Asses and Integrate Evidence

The first three phases of the generic REVVA process develop a fidelity based ToA/ToVV structure in a more or less top-down fashion. Summarized this ToA/ToVV structure specifies the all fidelity requirements needed to obtain a valid and acceptable simulation for the costumer's intended purpose. Phases four to seven of the REVVA process traverse the ToA/ToVV structure backwards and starts at the bottom with the actual fidelity measurement itself, which is know in RVVA as conducting V&V. This activity comprises the simulation system execution according to the test description of each fidelity evaluator function $C_{\Delta_{RS}}$ in the ToA/ToVV.

The resulting simulation data are then compared to the fidelity referent counterparts by means of the metric associated with each fidelity evaluator function. The outcome of these metrics Δ_{RS}^{-1} provides the fidelity items of evidence (IoE) for how well a simulation replicates a real entity or system i.e. the level of available fidelity. Before fidelity measurements are accepted as items of evidence, the actual errors $\delta\Delta_{RS}$ and uncertainties $U\left(\Delta_{RS}^{-1}\right)$ encountered in the measurement process must be assessed. This provides essential information necessary to evaluate the probative force of each fidelity IoE. Next the individual items of evidence are aggregated and expressed by the fidelity effectiveness measures in the ToVV. In here a value of one indicates that the fidelity acceptability criteria is met, i.e. fidelity is within the tolerance. The closer the fidelity effectiveness outcome approaches zero the less effective the fidelity acceptability criteria is met. If all fidelity MoE's in the ToVV have a value of one the simulation is said to have a sufficient level of fidelity for its application purpose. How strong this validity statement is depends upon the convincing force of the ToVV. The convincing force of each fidelity effectiveness outcome is based upon the aggregated evaluation of the probative force of each contributing fidelity IoE. It is yet hard to provide a mature and formal technique for this purpose. Therefore, the TNO-FM uses a formal SME review process to evaluate the credibility of the validity proposition based upon the relative importance weights of each fidelity effectiveness measure together with the fidelity uncertainty and error characterizations [18].

6 A Practical Simulation Case-Study

In the north-east of the Netherlands a training village "Marnehuizen" has been constructed for live simulation based training of MOUT operations of army personnel. The training involves group coordination at platoon and company level with various organic vehicles and weapon systems. Currently it is difficult to follow and quantify the performance of trainees during live practice. Observers in the training area are necessary for recording events during the course of the training. These observers may have an influence on the development of the simulation scenario, and moreover, the events during the training take place in different locations, complicating the work of the observer. To address these issues TNO Defense, Security and Safety has built a highly detailed digital terrain database of the training village areas. This terrain database is going to be used as a 3D synthetic environment and stealth viewing federate to observe and score the performance of the trainees during the simulation exercise. Furthermore, the federate will also be used for after action review of simulation exercises.

The development of the ToA/ToVV for the Marnehuizen federate started with interviews of the employees (the users) of this training facility [23]. From these interviews and a scenario(s) review session the following user-purpose or top goal has been formulated: *The Marnehuizen federate terrain database is suited for performance evaluation of training exercises in the Marnehuizen facility*. Next in several brainstorm sessions with a SME and the TNO project team a complete ToA hierarchy and a first fidelity referent was developed. In here a rather subjective and heuristic approach has been used to assign rather coarse grained fidelity tolerances and relative importance weights to each ToA node. Similarly, the allowed fidelity uncertainties for each node claim were established. The complete resulting ToA consists of multiple nodes on four aggregation levels and eleven fidelity related acceptability criteria (AC). For brevity only the following fidelity AC is considered here: *To facilitate easy recognition of the situation during the training exercise the Marnehuizen federate data-base must resemble the real Marnehuizen terrain very well.*

Before a measurable fidelity effectiveness measures could be assigned to this and the other fidelity AC's, each AC had to be expanded into a ToVV with high-level and compound fidelity measures, fidelity evaluator functions (IoE's) and evaluation task descriptions. This required substantial help of several SME's, who are very well know with the real terrain. During this effort the fidelity referent was improved with newly identified data. Furthermore, since the impact of not meeting these criteria are very significant for the intended purpose the fidelity effectiveness measure of each AC should score at least 0.9 with a maximum residual uncertainty of 0.1 to be accepted as valid i.e. have sufficient level of fidelity.

The final ToVV for the fidelity AC considered here consists of eight necessary fidelity IoE's to be checked. In this paper two of those required fidelity IoE's are considered:

1. *All relevant objects in Marnehuizen terrain must be represented in the Marnehuizen federate data-base.*
2. *The height profile of the federate data-base must match that of the Marnehuizen terrain.*

Based upon a list of sixteen referent object types that are present in the real Marnehuizen terrain, two types of system level resolution evaluator functions were assigned to fidelity IoE one (Paragraph 3.2). The first one is the total system scope difference which is formally written as:

$$(d_1)^{-1} = 1 - \frac{n\left(R_{ref} \setminus S_{db}\right)}{n(R_{ref})}$$

This fidelity metric specifies on a scale axis from zero to one how many objects in total are represented by the terrain data-base (S_{db}). The second one is the partial object scope difference which has the same mathematical structure but instead evaluates how many of each type of terrain objects are represented.
Next the SME was asked to assign relative importance weights to each type of object with respect to their contribution in the recognizability of the terrain. This outcome was then used to assign the tolerated deviation in presence of the number of objects and to construct the fidelity effectiveness functions. The SME judged that if the evaluator function results were outside the fidelity tolerance band its effectiveness rating must be zero. Together with the importance weights the evaluator functions were combined in a single fidelity effectiveness measure for fidelity IoE one (Paragraph 4.2). According the SME the maximum residual uncertainty for this fidelity effectiveness must not exceed the value of 0.1

For the second fidelity IoE two types of evaluator functions from the sample pair accuracy category in the TNO-FM fidelity metrics taxonomy (Paragraph 3.2) were chosen and combined in a fidelity effectiveness measure for the terrain profile fidelity. The first one is the absolute maximum elevation difference evaluation function:

$$(d_{2.1})^{-1} = \frac{\max\left|z_{ref}(gp)\right|}{\max\left|z_{ref}(gp) - z_{db}(gp)\right| + \max\left|z_{ref}(gp)\right|}$$

In this case the maximum difference between the real and the terrain data-base elevation profile grid point (*gp*) on a grids-size of one meter is measured. The second fidelity metric is the integrated absolute difference of the terrain slope at each grid point:

$$(d_{2.2})^{-1} = \frac{\sum\left|sl_{ref}(gp)\right|}{\sum\left|sl_{ref} - sl_{db}\right| + \sum\left|sl_{ref}(gp)\right|}$$

Both metrics are chosen such that they both return a dimensionless value between zero and one. A value of one indicates perfect replication of the terrain properties.

Again a SME was asked to specify the fidelity tolerances for both accuracy metrics being respectively absolute maximum different of 0.2 meter and integrated slope deviation of 1%. Furthermore, the SME assigned the following importance weights respectively: 0.8 and 0.2. The same residual uncertainty for this fidelity effectiveness measure was chosen as for the previous fidelity effectiveness measure. However, for both fidelity metrics a fidelity effectiveness rating function with a block shape was chosen (*Figure 4-1*) with the following rating scheme:

d2.1>0.7	d2.1 <0.8	d2.1 <0.8
0.0	0.5	1
d2.2 >0.33	d2.2 >0.66	d2.2 <0.66
0.0	0.8	1

Table 1 Fidelity Effectiveness Ratings

By executing the previously described fidelity evaluation task and aggregating the results into the fidelity effectiveness measures a final score could be calculated for each of the two fidelity acceptability criteria. The first fidelity IoE scored 0.75 and the second fidelity IoE scored 0.95, both with an aggregated uncertainty of 0.05. The aggregation was based upon SME and TNO project team estimations and educated guesses. These results were then aggregated with the other six IoE to make a validity statement for the stated complete "recognizability" acceptability criteria. Although one fidelity tolerance wasn't met it was judged by the SME's and the users that the impact on the total recognizability was acceptable for the application purpose.

7 Conclusions and Future Work

This paper presented a brief overview of the current version of the WEAG THALES Joint Program 11.20 VV&A methodology, or in short REVVA. The REVVA methodology provides a common integrated set of VV&A methods, techniques and standards to assess the validity of simulation systems. Next an introduction to the TNO Fidelity Methodology (TNO-FM) has been discussed, which is a methodology to systematically define, specify and measure simulation fidelity. This discussion showed that it is possible to specify simulation validity in terms of fidelity. The concept of fidelity requirements has been introduced as the linking pin between fidelity and validity. Based upon a formal definition of fidelity requirements and the presented taxonomy of fidelity evaluator functions the concept of fidelity effectiveness measure was introduced. Then it was shown how these TNO-FM concepts and products can be used to more precisely define the REVVA's main elements (ToA/ToVV) and be integrated within the REVVA generic process. A small case-study was presented to illustrate and investigate the benefits of combining both methodologies.

The major lessons learned from this integration exercise and case-study are the following:

- The REVVA itself lacks precise definitions for the elements in the ToA and ToVV. The templates of the TNO-FM products like fidelity referent, requirements and fidelity effectiveness measures provide a possible and useful means to fill this void.
- (De)Aggregation of all information in the ToA/ToVV structure is not a sinecure. Although the TNO-FM fidelity products improve this, it still remains a highly subjective and informal process. (particularly, in the area of aggregating the probative force of each item of evidence underlying the ToA acceptance criteria.)
- Developing this fidelity-based ToA/ToVV proved not to be a water-fall process but a spiral development in which step-by-step fidelity tolerances, uncertainties, referent, MoE's, MoP's and evaluator functions are found and refined in conjunction to each other.
- Building a ToA/ToVV tree, even for the presented small case-study, becomes rapidly very large and complex. To keep the Marnehuizen fidelity-based ToA/ToVV well-ordered and manageable the TNO project team used the Assurance and Safety Case Environment (ASCE) tool of Adelard.
- The TNO-FM fidelity case-study showed that the REVVA process proves to be a generic framework that allows for tailoring and extension of its products to improve the VV&A process. However, this might also a drawback for future users.

Currently, TNO Defense, Safety and Security participates in the REVVA follow-up project, REVVA2, to further improve the REVVA with the lessons-learned from this TNO case-study. In this context also the TNO-FM will be reviewed and its integration within the REVVA generic process will be studied in more detail.

8 References

[1] Balci O., 'A Methodology for Certification of Modeling and Simulation Applications'. *ACM Transactions on Modeling and Simulation.* Vol. 6., No. 1, January 1996, pp. 76-98

[2] Balci O., R Nance, J. Arthur, '*Expanding our horizons in verification, validation and accreditation research and practice'*. In proceedings of the 2002 Winter Simulation Conference: San-Diego, December 2002.

[3] Birta L.G., F.N. Ozmizrak, 'A Knowledge-Based Approach for the Validation of Simulation Models: The Foundation'. *ACM Transactions on Modeling and Computer Simulation*. Vol. 11., No. 4, October 2001, pp. 352 – 377.

[4] Brade D., J. Voogd, 'The Target of Acceptance: Towards a Precise Definition of Intended Purpose'. *Proceedings 05 Spring Simulation Interoperability Workshop*, March 2005, Paper 041.

[5] Defense Modeling and Simulation Office (DMSO), 'Federation Development and Execution Process (FEDEP) Model, version 1.5'. Department of Defense: December 1999.

[6] Defense Modeling and Simulation Office (DMSO), 'Verification, Validation and Accreditation (VV&A) Recommended Practice Guide, build 2.5'. Department of Defense: June 2000.

[7] Defense Modeling and Simulation Office (DMSO), ‘Foundations Workshop V&V Workshops’. Proceedings 2002 and 2004.

[8] Gross D.C., et al., ‘Report from the Fidelity Definition Implementation Study Group’. 99 Spring Simulation Interoperability Workshop, March 1999, Paper 176.

[9] Jaquart R., et. al., ‘WEAG THALES JP11.20 (REVVA) Results and Perspectives’. 05 European Simulation Interoperability Workshop, June 2005, Paper 021.

[10] METHGU2 2004. VV&A Methodological Guidelines Reference Manual. THALES JP11.20 Technical Report. JP1120-WE5100-D5101-METHGU2-V1.2.

[11] Oberkampf W.L., et. al. ‘*Mathematical Representation of Uncertainty’*. Proceedings AIAA Non-Deterministic Approaches Forum. Paper AIAA-2001-1645. Seattle, WA: April 2001.

[12] PROSPEC. 2004. VV&A Process Specification: User’s Manual. THALES JP11.20 Technical Report. JP1120-WE5200-D5201-PROSPEC-V1.3.

[13] Roza M., P. van Gool and H. Jense, ‘A Fidelity Management Process Overlay onto the FEDEP Model’. *Proceedings 98 Fall Simulation Interoperability Workshop*, September 98, Paper 083.

[14] Roza M., J. Voogd, H. Jense and P. van Gool, ‘Fidelity Requirements Specification: A Process Oriented View’. *99 Fall Simulation Interoperability Workshop*, September 99, Paper 032.

[15] Roza Z.C., et al., ‘Report out the Fidelity Experimentation Implementation Study Group’. *Proceedings 00 Spring Simulation Interoperability Workshop*, March 2000, Paper 151.

[16] Roza M. J. Voogd, P van Gool, ‘Fidelity Considerations for Civil Aviation Distributed Simulations’, *AIAA Modeling and Simulation Technologies Conference*, August 2000, Paper AIAA-2000-4397.

[17] Roza Z.C., J. Voogd, H. Jense, ‘Defining, Specifying and Developing Fidelity Referents. *Proceedings 01 European Simulation Interoperability Workshop*, June 2001, Paper 008.

[18] Roza Z.C., Simulation Fidelity Theory and Practice: A Unified Approach to Defining, Specifying and Measuring the Realism of Simulations. Delft University Press. Delft: December 2004.

[19] Saaty T., ‘The Seven Pillars of the Analytic Hierarchy Process’. *ISAHP Proceedings*. Kobe 1999.

[20] Stevenson, Steve, “A Critical Look at Design, Verification, and Validation of Large Scale Simulations”. *IEEE Computing in Science and Engineering Journal*, 1(3): 53-63, May/June 1999.

[21] TECH2. 2004. VV&A Techniques. THALES JP11.20 Technical Report JP1120-WE3100-D3102-TECH2-V2.0. TOAGUID. 2004.

[22] TOA Guidelines. THALES JP11.20 Technical Report JP1120-WE2200-D2201-TOAGUID-V1.1.

[23] Voogd J.M. M.L. van Emmerik. ‘A practical case-study of the VV&A JP11.20 Methodology: Marnehuizen’. TNO-Report, 2006. (In progress).

[24] Zeigler B.P., H. Praehofer, T.G. Kim. Theory of Modeling and Simulation. Academic Press: January 2000.

Author Biographies

Dr. Manfred Roza is a member of the scientific staff at the Training and Instruction Department of TNO Defense, Security and Safety. He holds a M.Sc. and Ph.D. in Aerospace M&S Technology from the Delft University of Technology. Besides VV&A his research interests include modeling and (distributed) simulation, human-factors and real-time systems software engineering.

Dr. Martijn van Emmerik is a member of the scientific staff at the Training and Instruction Department of TNO Defense, Security and Safety. He holds a Ph.D. in Behavioral Sciences from the Twente University of Technology. The relation between fidelity and validity in training simulations has been one of his main research interests over the past years.

Dr. Jeroen Voogd is a member of the scientific staff at the Modeling and Simulation Division Department of TNO Defense, Security and Safety. He holds a Ph.D. in Computational Physics from the University of Amsterdam in the field of modeling and simulation biophysical systems on parallel distributed computation platforms. Recurring themes in his current work are issue around simulation fidelity and validity.