Driving Simulation Scenario Definition Based on Performance Measures

Yiannis Papelis  
Omar Ahmad  
Ginger Watson  

NADS & Simulation Center  
The University of Iowa  
2401 Oakdale Blvd.  
Iowa City, IA 52242-5003  
(319) 335-4786  
Fax: 335-4658  
Contact e-mail: yiannis@nads-sc.uiowa.edu

For many reasons, the development of a quantitative methodology for describing driving simulation scenarios has been elusive. Without a quantitative technique for describing scenarios, portability suffers and as a result, the ability of independently verifying results is diminished, something that sheds doubt to all driving simulation research activities. The challenge in creating a universal and quantitative technique for defining scenarios can be partially attributed to the different software packages utilized for creating and executing scenarios. However, even for scenarios produced by identical software codes, differences in simulator fidelity can have a profound effect on the actual scenario rendering, and in turn on the equivalence of such scenarios when run on different simulators. This paper describes some research towards a unified specification methodology that addresses scenario definition starting at the performance measures that a scenario is intended to sample. Once the performance measures have been identified, differences between software fidelity and rendering can be analyzed to ultimately determine effect on measuring the required performance. Whereas it may not always be possible to define scenarios that provide identical results even across simulators, having an understanding on how differences affect performance is a first step towards scenario portability.
Introduction

Each year in the United States, over 42,000 people lose their lives in traffic accidents [1]. It is estimated that the root cause of the majority of these accidents involves some degree of human error. Simulator-based research holds great promise when it comes to studying and mitigating the human error component of traffic accidents. As in all disciplines, repeatability of research experiments is necessary in order to validate the results and to provide confidence in the findings. Unfortunately, repeatability of driving simulator experiments is a challenging process because of the wide range of simulator technology that is currently available. Another contributing factor hampering repeatability is the nature of the research protocols typically used in simulator studies. Depending on the simulator and the specific technology being used to implement the necessary scenes and scenarios, the effort needed to recreate an equivalent virtual testing environment can vary widely. Recent efforts aiming to address scene and scenario reuse typically focus on the technical approach used to replicate scenes and scenarios within the same simulator platform [2]. Other efforts have concentrated on identifying key characteristics and studying the possibility of comparing, reusing, and/or replicating scenarios across different simulator platforms, something that is more promising and if achieved, would increase confidence in results obtained in simulator studies. Unfortunately, several challenges exist when attempting to recreate scenarios across different simulators. Lack of formal specification methods and the variation of fidelity of simulator subsystems are the two most prominent ones.

In this paper, we discuss some fundamental issues pertaining to the feasibility of reusing or replicating scenarios across different simulator platforms. The approach is based on focusing on the eventual outcome of a study in terms of performance measures and quantifying a simulator’s fidelity in rendering cues used to excite a response and eventually measure human performance. This is in contrast to comparing the implementation or similarity of virtual testing environments. Based on this logic, portability of scenarios across simulators whose fidelity varies greatly is inherently problematic, even if the remaining details of the implementation remain the same. Note that this does not imply that a specific level of fidelity is necessary in order to conduct valid research. On the other hand, comparing scenarios implemented on simulators of different fidelity cannot be done simply by comparing the virtual environment implementation (scenes, scenarios, cues), but must take into account how rendering this environment can be inherently so different between simulators that comparing performance measures is like comparing apples and oranges.

Background and Motivation

The typical lifecycle of a research study involves a hypothesis that is proven or unproven based on the data obtained during the execution of the experiment. An experimental design is constructed around the research objectives, questions, and hypotheses. From this, performance measures are defined in terms of dependent and independent
experimental variables. Scenarios and events are defined that will expose research participants to situations during which their performance is measured. This involves creating the virtual testing environment, collecting raw data, and then reducing that data into meaningful measures. The final analysis phase involves going back to performance measures to see if the data supports the original hypotheses.

![Figure 1 – Typical study lifecycle.](image)

To meaningfully compare experiments performed on different simulators, one must be able to independently recreate each of these steps. A problem that transcends any technical aspect of recreating an experiment is that of concise and precise documentation. Lack of detailed or accurate documentation on any of the steps involved in this process will prevent any legitimate replication of the research. We consider this issue to be beyond the scope of this paper, and hereto assume that adequate documentation exists for all components of the study.

Figure 1 illustrates a typical research study lifecycle. Even though scene and scenario definition and implementation is just one aspect in the process of running a study, it is the one that depends the most on simulator technology. Any effort whose ultimate goal includes research repeatability or reuse of scenes and scenarios must concentrate on the portability as well as the meaningful equivalence of scenes and scenarios across simulators.

Several efforts focusing on scene and scenario reuse by the authors and their colleagues have concentrated on the technical aspects involved in recreating or reusing scenes and scenarios across the same simulator or different simulators. The challenges there are primarily of a software engineering nature. Typical issues include converting visual environment representations between different formats, ensuring that different visual databases render at compatible frame rates, automatically converting road network representations between different proprietary formats, automatically documenting scenarios, and automatically or semi-automatically implementing scenario descriptions across different software environments. One thing, however, that to the best of our knowledge has not been directly addressed or explicitly acknowledged is the detailed comparability of how different simulators render the cues necessary to implement the virtual testing environment. Human performance within a virtual environment depends on the cues produced by the simulator. When the fidelity of rendered cues varies greatly, or worse, when cues are missing altogether, human performance will vary independent of how similar the scenes and scenarios are. As an extreme example, consider field of view. One can very quickly come up with numerous scenarios that depend so heavily on having
a certain field of view (i.e., a right incursion) that implementing them on a simulator with narrow field of view would induce no response whatsoever.

Whereas extreme examples can easily be dismissed as unrealistic because any competent researcher would not utilize a simulator that grossly lacks the cueing fidelity necessary to implement a scenario, we believe that the cueing fidelity affects the obtained performance measures in ways that may not be apparent. Let us consider a more complicated example in order to demonstrate how numerous aspects of simulator fidelity can affect scenarios in unobvious ways.

Consider a scenario involving a vehicle pull-out event. The goal is to measure the participant’s reaction time to a vehicle pulling out into the participant’s lane as shown in Figure 2. As the participant approaches the intersection from the south, a vehicle pulls out in front of the driver. The critical component of this scenario is the timing of the interaction. Let us assume that the intent is to ensure a consistent time-to-collision between the incursion vehicle and the participant’s vehicle. Using the instantaneous speed of the participant’s vehicle, the incuring vehicle can adjust its speed so that when it initially enters the participant’s lane, time-to-collision will be independent of any other factors. A second consideration may be that the incursion vehicle is not to start moving too early, as this can act as a warning to the driver, who may decide to slow down based on early anticipation, which in turn will eliminate the rogue reaction behavior that the researcher is seeking to study. Let us now consider the ways in which the participant performance can vary, solely because of the cue fidelity provided by the simulator.

Incomplete specification of ambient visual elements

In addition to time-to-collision, one must take into account the time that the participant can first observe the incursion vehicle move. If that happens too early or too late, then the reaction time will act as a dependent, rather than an independent, variable. The presence of buildings, signs, or other obstructions that may occlude visibility of the vehicle can greatly change the outcome of this scenario. Similarly, the field of view should be such that it does not affect the visibility of the existence or motion of the incurring vehicle.

Visual rendering quality

One aspect that is rarely mentioned in the description of such scenarios is the visual rendering quality as it relates to the specific performance measure. For example, one can argue that choosing a bright color for the incurring vehicle and a dark color for the background buildings will provide a visual cue that is significantly more noticeable to a participant when compared to a low-contrast choice of colors. Simulator specifications often include technical details such as contrast ratio, but the selection of colors and the specific tuning of the visual display system can introduce great variability in the actual contrast ratio for the scenario-relevant features. The effect of such choices can greatly affect scenarios that involve visual processing occurring near the peripheral field of view, yet is rarely documented or explicitly taken into account when discussing scenario portability.
A similar factor is that of overall scene complexity. The identical scenario implemented on a barren scene is more likely to draw the participant’s attention than a scene that is full of colorful buildings, banners, and other similar background artifacts.
Simulator disorientation

It is a well known fact [3] that simulator exposure can lead to various levels of disorientation or discomfort. The levels of the symptoms vary depending on the fidelity of the simulator and the degree to which the rendered cues match the expectation of reality. Whereas extreme symptoms that cause a participant to interrupt a drive can be easily measured, of great concern are participants who experience symptoms that are significant enough to modify their behavior, yet are not strong enough to force quitting the drive. If a participant compensates for their symptoms by adjusting their behavior, then the measured performance is biased by the simulator cueing fidelity, something that is rarely captured when comparing similar scenarios across different simulators. In the aforementioned example, if due to poor implementation or complete lack of motion cues, the participant is hesitant to apply the brakes because of fear of discomfort, then the performance is dominated by the simulator characteristic and not the scenario itself.

Driving task workload related to lack of cues

Even though driving is generally characterized as largely a visual task [4], the vestibular system is largely responsible for sensing acceleration and serves as a significant factor in reducing the workload on a driver in lane tracking and steering. Lack of motion cues, or motion cues that are not representative of actual conditions, necessitate that the driver depend more on other senses in lieu of lateral acceleration. This unfamiliar mode of driving changes the baseline workload, which in turn renders any comparison with the same scenario implemented on a different simulator suspect.

Audio/tactile/vibration feedback

Similarly to the changes in baseline workload introduced by the lack of motion cues, lack or differences in the rendering fidelity of audio or tactile cues can drastically affect baseline driver behavior. In addition to affecting the overall fidelity and the level of immersion, tactile and audio/vibration feedback may change the overall perception of vehicle velocity, among other things. Such changes in baseline behaviors can drastically affect the outcome of scenarios and make meaningful comparisons between scenarios implemented on different simulators very difficult.

Scenario implementation details

Lastly, one aspect of scenario portability that is rarely addressed is the numerous details involved in programming the autonomous behaviors of ambient traffic. Typically, scenario control systems utilized in simulators produce various volumes of ambient traffic that behaves in a semi or fully autonomous manner. Because there is wide latitude in what constitutes autonomous behavior as well as the sheer breadth of micro-behaviors involved in traffic modeling, such details are rarely documented. It is conceivable, however, that such differences could easily affect the participant’s performance. Consider as an example the default following policy used by autonomous traffic. If it is set at an aggressive level, it is more likely to create a distraction to the participant than a less aggressive policy that ensures that autonomous traffic follows at a further distance. Other details that can affect the participants’ behavior include but are not limited to: default acceleration after a stop, lane deviation variation, route selection, speed selection relative to speed limit, passing behavior, and speed variation.
An Alternative to Scenario Portability

As explained earlier, numerous aspects of how a virtual testing environment is rendered can affect the performance measures obtained in the course of a simulator study. Variation in simulator fidelity or even incomplete specification of tunable parameters on simulators with the same fidelity can be equal culprits in complicating the issue of scenario portability. Performance measures must be part of any approach that tries to implement portability across simulators.

The alternative approach proposed in this paper involves the incorporation of the performance measures as the key element in guiding the essence of a scenario. In lieu of describing a series of events or scenarios and their implementation, this approach utilizes a list of performance variables and the type of driving interactions used to measure them. By focusing on the performance measures, as opposed to focusing on scenario events, it is easier to document and understand how differences in simulator fidelity affect the performance measures. This in turn will allow researchers to compare performance measures across simulators and eventually quantify the differences.

Figure 3 illustrates the current approach to building simulator experiments. Scenarios are defined by researchers in conjunction with simulation engineers. Prior experience and expertise is used to design scenarios that are likely to produce performance measures that are sensitive to the hypothesis. These scenarios consist of several events, and for each event, several performance variables are obtained. After the appropriate processing, the performance measures are used for analysis. Implicit in this process is the effect that simulator fidelity has on the performance measures.

When viewing the process in the manner depicted in Figure 3, one can see how focusing on scenario portability or reuse is not the most effective way of ensuring that the research is replicated. Understanding the effects that simulator fidelity has on the performance measures is as critical as ensuring that the scenario and event implementation are comparable.
Figure 4 provides an alternative approach to thinking about scenario definition, one that we believe is much more amenable to allowing portability across simulators. In this approach, instead of defining scenarios first, one starts with the high-level constructs to be measured. This in turn defines the desired performance measures. Prior expertise is still necessary to ensure appropriate selection of performance measures; however, the focus is first on the measures themselves, as opposed to the events or scenarios. Explicit in this approach is the careful study of how these performance measures are affected by simulator fidelity. Again, it is important that this step is done before any mapping to actual events or scenarios is performed. This is also the time at which careful analysis must be made to determine if the simulator fidelity is adequate for measuring a particular performance variable.

![Diagram](image)

Figure 4 – Alternative way of thinking about simulator experiments.

Only after the effect of simulator fidelity on the performance measures has been studied and analyzed can the specific events be defined. Note that there is a one-too-many relationship between performance measures and events. In effect, it is possible to define several events that will measure the same performance. As a simple example, consider a high-level construct of driver distraction measured through reaction time. There can be numerous events that measure reaction time as a dependent variable under distraction. This is the point when intelligent decisions can be made about which events are more applicable based on simulator fidelity and capabilities in general. Additional factors that affect the mapping from performance measures to events and eventually to scenarios include the experimental design and the degree of desired redundancy in measuring performance. Table 1 provides some examples of constructs and possible mappings into performance measures and associated events.
Table 1 – Performance measures grouped by constructs

<table>
<thead>
<tr>
<th>Construct</th>
<th>Performance Measure</th>
<th>Scenario Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drowsiness/drug impairment</td>
<td>Velocity variance</td>
<td>Car following</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Maintain speed task</td>
</tr>
<tr>
<td></td>
<td>Steering entropy</td>
<td>Any</td>
</tr>
<tr>
<td></td>
<td>Lane deviation</td>
<td>Any</td>
</tr>
<tr>
<td></td>
<td>Lane excursion</td>
<td>Curve negotiation</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Interaction with vehicle</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Long drive w/ fatigue</td>
</tr>
<tr>
<td></td>
<td>Eye perclos.</td>
<td>Any</td>
</tr>
<tr>
<td></td>
<td>Blink rate</td>
<td>Any</td>
</tr>
<tr>
<td>Driver distraction</td>
<td>Rote RT</td>
<td>Traffic light change</td>
</tr>
<tr>
<td></td>
<td>Decision RT</td>
<td>Yellow light</td>
</tr>
<tr>
<td></td>
<td>Horizontal scanning field</td>
<td>Target detection</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(tunnel vision)</td>
</tr>
<tr>
<td></td>
<td>Saccades</td>
<td>Any</td>
</tr>
<tr>
<td>Effectiveness of loss of</td>
<td>Loss of/maintain control</td>
<td>Emergency maneuver</td>
</tr>
<tr>
<td>control countermeasure</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Technical Errors</td>
<td>Incorrect road geometry</td>
</tr>
</tbody>
</table>

It is important to realize that this approach does not directly address gross differences in fidelity between simulators. Certain performance measures may be impossible to meaningfully obtain on a simulator that lacks the necessary cues. For example, it is well known that simulators with no motion feedback cannot be used for research that involves vehicle handling. On the hand, this approach of developing scenarios helps amplify certain deficiencies and, to the degree possible, allows meaningful comparisons between similar scenarios when implemented on simulators of different fidelity.

**Ensuring Scenario Portability**

Given this alternative strategy of developing scenarios, portability can now be addressed by reframing the problem from scenario equivalence to that of performance measurement equivalence. Instead of focusing on how to create similar virtual testing environments, as well as how to specify the detailed interactions between the research participant and the virtual environment, one must focus on how to port the desired performance measurements from one simulator to another.

Consider the simplistic example of a research study focusing on the effectiveness of a warning system in preventing rear-end collisions. The high-level construct of system effectiveness is translated into a series of performance measures that may include reaction time when faced with an unexpected lead vehicle deceleration.
The old approach would involve a description of the scenarios: “A lead vehicle travels at 50 mph, the participant follows at 1.5 seconds, then the lead vehicle brakes at 0.6 G”. Porting that scenario to a different simulator would then concentrate on re-implementing this set of interactions, disregarding all the details that may affect the performance measures; aspects such as the contrast of the brake lights, the color of the vehicle, and the complexity of the visual scene are just a few of the issues that would be lost in the “porting” effort.

The strategy presented here would approach the porting effort by looking at how to measure reaction time when faced with an unexpected lead vehicle deceleration. Instead of describing the event, one would consider all cueing aspects and how they affect reaction time. Reaction time can be broken down into creating a stimuli and measuring the response. The stimuli and response can then be selected based on the simulator capabilities. If the decision is made to utilize the brake light change as the stimuli, then attention is automatically directed to maintaining the validity of the stimuli as scenarios are ported to a different simulator. The size of the brake lights on the screen and the contrast between the brake lights on versus off can be made to appear the same, even if the visuals systems vary. When measuring the response, issues such as the sampling rate can be taken into account to ensure portability. Ultimately, porting the scenario involves not recreating the same scenario, but recreating the same set of measurements, something that may even lead to a different scenario altogether.

Note that this approach does not automate or necessarily simplify the problem of scenario portability. It does, however, bring to the foreground all the issues that are currently unaddressed. When properly addressed, the differences in fidelity can provide a foundation upon which meaningful comparisons can be made. Eventually, this will allow portability of scenarios across simulators.

**Conclusion and Future Work**

This paper presented the challenges involved with considering the issue of porting scenarios across simulators of different fidelity. Examples were presented to show how performance variables measured on different simulators cannot be compared while explicitly addressing the effect that cue rendering has on human performance. An alternative strategy for developing scenarios was presented; this strategy involves using the research constructs to identify performance measures, which are then used to select the appropriate events, while taking into account how simulator fidelity affects measurement. When using this strategy, scenario portability is a non-issue; instead, the issue is framed in the context of porting performance measurements.

A large volume of future work is necessary to refine this approach and reach an implementable strategy. This work involves understanding how cue rendering affects performance. While there is a large body of research on how variations in motion subsystems, drive algorithms, visual displays, and haptic, tactile, and audio feedback all affect performance, the level of quantification is not enough to allow meaningful
comparison. Only as these effects are studied and quantified can portability of scenarios be effectively addressed.

References


