Human Factors and Ergonomics Society

NEWSLETTER

OF THE

TEST AND EVALUATION TECHNICAL GROUP

FALL 2006

FROM THE CHAIR

The Test and Evaluation Business Meeting will be in a new format – *Breakfast Meeting on Wednesday (07:00 – 08:30)*. Bring yourselves – meet with friends and colleagues – and enjoy food for the body, good cheer, and food for thought. Donald E. "Don" Farr, known for his broad and humorous views of our discipline, will be speaking on the "***NEXT 50 YEARS OF HFES T&E: THE FARR VIEW.*** We suspect that DON will be viewing Human Factors/Ergonomics (HF/E) as just becoming an "adolescent discipline" – after-all it is only 50 years old. We also understand Don – as he approaches this occasion – is also fully confident in making some 50 year prognostications based on his long (50 year) and broad experiences […and his "Farr view" that, "…after all, I likely will never have to answer to any agency, or client… fifty years from now".]. We encourage you to come and bravely experience the FARR-VIEW of T&E over the next 50 years!!

Alvah C. Bittner, PhD, CPE
Bittner & Associates
Kent WA, 98042

Special Feature:

Introducing a Method for Enabling Comparison of Research Results Between Vehicle Simulators

Selina Mård Berggren[1] and Laban Källgren[2]
Swedish National Road and Transport Research Institute
[1]Human, Vehicle, Transport System Interaction and [2]Vehicle Technology and Simulation

Abstract

Applying an interdisciplinary approach, this paper proposes a method that enables comparison of research results between platforms of different levels of fidelity. Technological and software

development over the last decades has given the research society new, complex tools for studying and understanding the interaction between human and vehicle transport systems. Today’s vehicle simulators range from low fidelity to extremely high fidelity. Simulator studies involving several partners make it possible to learn about cultural differences in driving behaviour, to share costs, to include a large number of study subjects, to generalize results and to increase research exchange. However, there are methodological problems in comparing data obtained in studies involving several different vehicle simulators. Differences in simulator complexity and fidelity give rise to uncertainty in the analysis of data. As collaboration worldwide increases it is evident that it is crucial to develop a tool for comparison of research results between vehicle simulators.

Introduction

The world of vehicle simulators has been evolving rapidly over the last decades. Today’s powerful computers give broad possibilities to high fidelity simulators and realistic test scenarios. However, the vehicle simulators worldwide range from low fidelity to high fidelity.

In research it is important to bridge the gap of fidelity to be able to compare results, generalize results or to merge data and at the same time maintain a clear methodological conscience. There are numerous examples of ambitious and well planned studies involving several research partners where, in the end, it is not possible to make general comparisons of results. Even if the study is carefully designed, it is not possible to estimate what effect simulator fidelity has had on the research results. Divergence in results tends to cause discussion on the reason for differences in results rather than focusing on the research question at hand. The purpose with the present paper is to introduce a method that enables comparison of research results between platforms of different levels of fidelity.

There is a recent example in a European study where the initial thought was to merge data but as the study proceeded it was evident that discrepancy in data could not be validly explained. This conclusion left the researchers with isolated groups of data, a much smaller subject sample than anticipated and the need for separate analysis of data at the respective platforms. One could argue that we should leave it at that and continue with our own separate research. However, it is of importance to learn about driver behaviour and traffic conditions regardless of cultural variation and different traffic rules between countries.

As long as the effects of differences in fidelity level on driver behaviour and driver performance is not fully understood, there is no way of knowing how shifts in research results should be explained. The differences in results between simulator platforms could be due to cultural differences, differences in instructing the participants as well as differences in simulator and fidelity. These circumstances call for a standardized method which enables comparison between vehicle simulators.

We propose a method that includes a Scenario Complexity Scale, SCS, and a behavioural measure. This will provide each research platform with a Simulator Specific Weight, SSW, making research results comparable regardless of simulator fidelity. An interdisciplinary approach has been employed where we combine the best of our knowledge in human factors with expertise in vehicle simulation engineering.

Practice Innovation

Today it is close to impossible to compare research results or to merge data between vehicle simulators since the relation between fidelity level and driver behaviour is not fully understood. One can assume that perceived differences in fidelity gives an effect on driver behaviour and driver performance. There are some obvious ways of handling the gap in fidelity between simulators and the methodological problems that arise from this. One way would be to compare the results achieved in the simulator in a corresponding field study. In that case the same subjects would drive the same experiment both in a simulator and in real traffic. The circumstances would have to be as similar as possible. After having driven both versions (simulator and real traffic) the subjects would rate the difference between the "real" driving and the simulated driving. Every simulator platform involved in the study would do the same and then judge how much effect the differences between the simulated driving and the real driving has had on the results at respective platform. Another approach would be to have the same subjects travel to all the sites involved in a study and perform the same experiment everywhere. The differences in results between simulators would then be a measure on the differences of the simulator. Clearly there are concerns with both these approaches. In the first case it is problematic to find a fully comparable field experience. It is not possible to fully account for weather and natural variation in traffic flow in a field study, hence the prerequisites would not be exactly the same. Both methods would be time consuming and costly. Both methods are also dependent on subjective rating and subjective judgement of perceived differences. This paper proposes a constructive way of approaching future study comparison. In order to achieve this we propose a threefold method:

- The development of an objective Scenario Complexity Scale, SCS
- Finding the relation between SCS and a behavioural assessment, such as Mental Workload, MWL
- Use the relation between SCS and MWL to adjust research data according to a Simulator Specific Weight, SSW

In the following, we will sketch the general framework of the Scenario Complexity Scale and how to combine it with Mental Workload in order to create a Simulator Specific Weight.

Assumptions

- Mental workload increases with increasing scenario complexity
- If the scenario is held constant between simulator platforms, the difference in results concerning mental workload will be a measure on the difference between simulators.

The Scenario Complexity Scale, SCS

In pursue of a standardized measure for scenario construction, the Scenario Complexity Scale, SCS, is developed. While moving on the road, the scenario around the driver constantly changes. By imaging that each small road delta has its specific complexity, one can describe the scenario complexity as an information density function f(x), where x is travelled distance along the road. Firstly the density function depicts how different scenario elements relate to each other. On the other hand it gives a clue to what level the driver is exposed to accumulated mental workload. Example of scenario elements is road curvature and width, which may affect the mental workload of the driver. In each road delta the curvature complexity can be described by its bending radius. The complexity of the road width is described by the width itself. Another scenario element is meeting cars, whose density can be described as number of cars per meter. By assigning each scenario element a numeric density value, the density function is an objective measure of the scenario complexity in each road delta. One question that rises is how to assemble the density function. E.g. in order to combine curve bending radius and number of cars per meter, one has to take out constants which give meaning to the density function. One approach is to assume a linear relationship between the density function and its scenario elements, and that each element has its specific weight or constant. In order to achieve the scenario complexity between two points of the road, the density function f(x) is integrated. This makes it possible to calculate the scenario complexity for just a segment of the road (scenario) or to get a measure of the complexity for the whole scenario. In order to get a scenario complexity which is independent of distance travelled, the integrated density function can be normalized with distance, which gives us the Scenario Complexity Scale:

$$SCS = \frac{\int_{x0}^{x1} f(x)dx}{\Delta x}$$

SCS is independent of distance travelled, which makes sense: Driving one kilometre or 10 kilometres on a highway, does not change the scenario complexity. However, sleepiness or risk of disturbing events may increase over distance/time, but those variables are not part of the programmed scenario but influence on driver and such effects are outside the scope of this method.

The SCS measure can be used separately as a way of securing the same level of complexity between studies. Sometimes it could be desirable, for comparative reasons, to have the same scenario complexity between studies or between simulator platforms even if the scenarios are not identical. When using the SCS it is possible to describe the study scenario in a more refined way than just "…drove on a rural road and a motorway…" which is a common way of describing scenario setting today (p. 29 in Östlund, Nilsson, Törnros and Forsman, 2006).

In our proposal we take the SCS one step further and introduce a combination of the SCS and a behavioural measure (mental workload) to achieve a "Simulator Specific Weight", SSW. An assumption made, is that with increased scenario complexity (higher SCS) there is also an increase in subject rated mental workload, MWL.

Mental Workload, MWL

Mental workload is a concept used for describing the capacity of human information processing resources (Wickens, 1992). If an overuse of the resources should occur, the human performance deteriorates. Workload is defined "as the effort invested by the human operator into task performance" (Hart & Wickens, 1990, p. 258). Mental workload is responsive to changes in information load and can be measured by use of a subjective rating scale. There are several scales in use through different domains; one example is the Cooper-Harper Aircraft Handling Characteristic Scale (Cooper & Harper, 1969). Mental workload as a behavioural performance measure serves the purpose of the simulator comparison method well. The measure can be described as a static measure but when used at several different occasions during a scenario it gives the researcher a pseudo dynamic measure of the differentiating flow of the scenario. This provides the researcher with an overall picture of the changing flow in a scenario (Berggren, 2000).

The Simulator Specific Weight, SSW

The final piece in the development of a method for research result comparison between platforms is the Simulator Specific Weight, SSW. The specification of the weight is easiest demonstrated though an example of an initial weighting pilot study. Suppose two simulators with different levels of fidelity are involved in the study, simulator A and simulator B. A scenario is developed comprising four road segments with different degrees of complexity according to the SCS (see figure 1). Ten subjects drive the entire scenario, assessing their MWL at each level of complexity (at each road segment). The combination of the two measures, SCS and MWL, will provide each simulator platform with a regression line (linear relation assumed) as seen in figure 2. Depending of level of fidelity there will be a shift in achieved data (fig 2). Each platform engaged in the network of simulator comparison will achieve its specific regression line as a function of SCS and MWL. A weight will be calculated for each SCS level. This will make it possible to adjust for shift in future research results depending on fidelity differences between simulators.

Discussion

It is evident that there is a need for a standardized method for comparison of research results from simulators with different degree of fidelity. Of course there are many variables apart from fidelity that can account for differences in results between vehicle simulator studies but it seems that the question on differences in fidelity is a major dilemma. It is desirable to have an open discussion on how to approach this topic. There are still specific issues to consider regarding the development of the SCS, the relation between SCS and MWL, and the calculation of the SSW from the regression line. As of this paper we have given our view of what we think is a constructive approach in order to come to a solution.

Figures

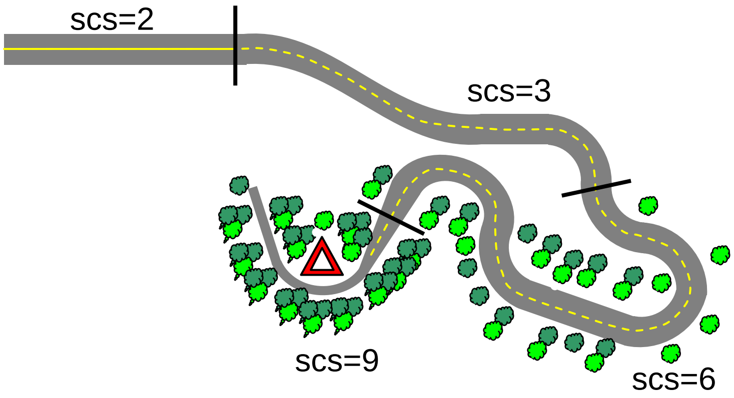

Figure 1. Example of road segments with different degrees of scenario complexity according to the SCS.

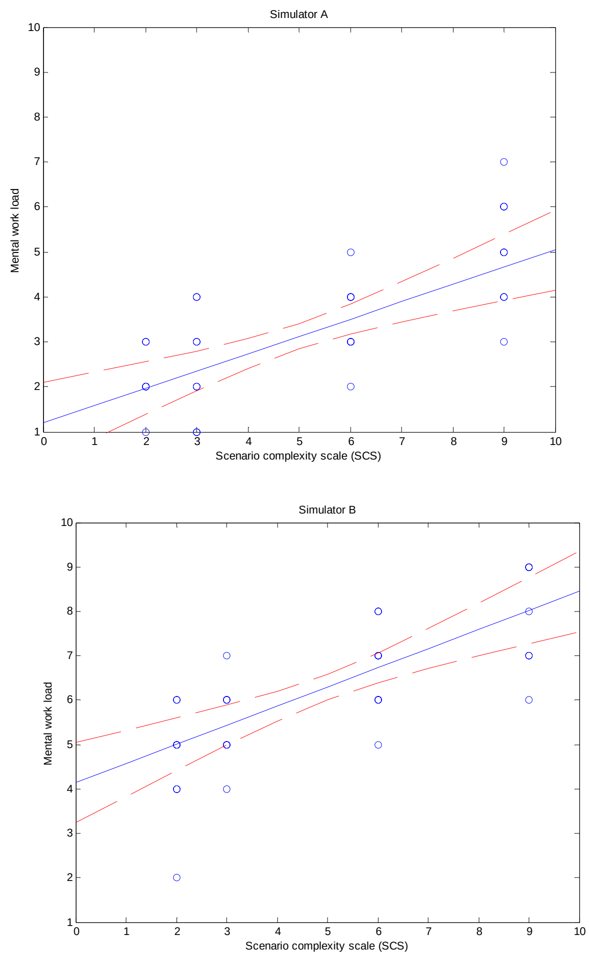

Figure 2. Simulated regression line for SCS and MWL at the two simulators involved in the initial pilot study.

References

Berggren, P. (2000). Situational awareness, mental workload, and pilot performance – relationships and conceptual aspects. *FOA: Scientific Report*. FOA-R--00-01438-706--SE Linköping: Defence Research Establishment.

Cooper, G., E., & Harper, R., Jr. (1969). The use of pilot rating in the evaluation of aircraft handling qualities. Moffett Field, CA: *NASA. Report* No. NASA TN-D-5153.

Hart, S., G., & Wickens, C., D. (1990). Workload Assessment and Prediction. In H. Booher, R. (Ed.), *Manprint: An approach to systems integration.* (pp. 257-296). New York: Van Nostrand Reinhold.

Wickens, C., D. (1992). *Engineering Psychology and Human Performance.* (2nd ed.). New York: Harper Collins Publishers Inc.

Östlund, J., Nilsson, L., Törnros, J., & Forsman, Å. (2006). Effects of cognitive and visual load in real and simulated driving. VTI publication.

Test and Evaluation Technical Group Program for 2006 Annual Meeting

Based on information in the online program posted at the Conference website (http://www.hfes.org/web/HFESMeetings/06annualmeeting.html), the Test and Evaluation Technical Group will sponsor three lecture sessions and one symposium. In addition, the conference will have poster sessions on Tuesday afternoon, Wednesday afternoon, Thursday morning and afternoon, and Friday morning. In addition, based on feedback received at the last HFES Annual Meeting, the business meeting of the Test and Evaluation Technical Group will be held as a breakfast meeting (7:00 am – 8:30 am) on Wednesday morning, September 18. Please check-out the TE Technical Program and the entire Conference Program, and plan to participate.

SESSION TE1:
Context, Complexity, and Changing Goals: Testing and Evaluation Challenges (5 lectures)
Tuesday, October 17, 2006
13:30 – 15:00

1. Context, Complexity, and Changing Goals: Testing and Evaluation Challenges
   - Joyce Cameron
   - Alvah Bittner (Bittner & Assoc.)
   - Brian Gore (NASA Ames Research Ctr./SJSU Foundation)
2. Adventures in Usability Testing: User-Centered Design in a Large-Scale Naval Ship Design Program
   - Larry Hettinger (Northrop Grumman Information Technology)
   - Vince Quintana (Bath Iron Works)
   - Robert Bennett (Northrop Grumman Mission Systems)
   - Robert Howells (Northrop Grumman Ship Systems)
   - Dan Donohoo (Raytheon Corporation)

3. What Kinds of Usability Problem Descriptions Are Useful to Developers?
    - Kasper Hornbæk (U. of Copenhagen)
    - Erik Frøkjær (U. of Copenhagen)

4. Thinking About Human Factors Testing and Evaluation: Perspectives from a Piano Bench
    - Joyce Cameron

5. Intraindividual Ergonomics (I2E): Framework and Future
    - Alvah Bittner (Bittner & Assoc.)
    - Yoshihide Sakuraga (Pacific Lotus Trading Co.)

SESSION TE2:
Test and Evaluation in the Real World – I
(5 lectures)
Wednesday, October 18, 2006
8:30 – 10:00

1. Real-Time Collaborative Heuristic Review: Meeting the Production Schedule Without Sacrificing Quality
    - Jesse Walker (U. of Dayton)
    - Drew Bowers (U. of Dayton)
    - Stephen Karth (U. of Dayton)
    -Carlton Donahoo (U. of Dayton)

2. Multiple Criteria to Evaluate Decision Support Outcomes
    - Holly Handley (Pacific Science & Engineering Group, Inc.)
    - Nancy Heacox (Pacific Science & Engineering Group, Inc.)

3. Measuring the Usability of Paper Ballots: Efficiency, Effectiveness, and Satisfaction
    - Sarah Everett (Rice U.)
    - Michael Byrne (Rice U.)
    - Kristen Greene (Rice U.)

4. Field Evaluation of Blue Force Tracking with a 2-D/3-D Command and Control System
    - Patrik Lif (FOI - Swedish Defence Research Agency)
    - Birgitta Kylesten (FOI - Man-System-Interaction)
    - Jenny Lindoff (FOI - Man-System-Interaction)

5. Tactical Evaluation of an Unmanned Ground Vehicle During a MOUT Exercise
    - Patrik Lif (FOI - Man-System-Interaction)
    - Hans Jander (FOI - Man-System-Interaction)
    - Jonathan Borgvall (FOI - Man-System-Interaction)

SESSION TE3:
Test and Evaluation Tools and Techniques
(5 lectures)
Thursday, October 19, 2006
10:30 – 12:00

1. Escape from Designers' Dilemma on Creeping Featurism
    - Dong-Seok Lee (Ohio State U.)
    - David Woods (Ohio State U.)
    - Daniel Kidwell (Ohio State U.)

2. Effect of Level of Problem Description on Problem Discovery Rate: Two Case Studies
    - James Lewis (IBM Corp.)

3. Development of an Operational Setting-Specific Field Cognitive Assessment Procedure
    - Wayne Harris (U. of Central Florida)
    - Dennis Reeves (Clinvest, Inc.)
    - Timothy Elsmore (Activity Research Service)
    - Peter Hancock (U. of Central Florida)

4. Evaluation of the FSA Hand Force Measurement System

- Kihyo Jung (Pohang U. of Science & Technology)
- Heecheon You (Pohang U. of Science & Technology)
- Ochae Kwon (Pohang U. of Science & Technology)

5. Is Less More When Using and Creating Checklists?
    - Janeen Sharma (Northrop Grumman Corp.)
    - Jennifer Rousey (Northrop Grumman Corp.)

SESSION TE4:
Test and Evaluation in the Real World – II (5 lectures)
Friday, October 20, 2006
10:30 -12:00

1. Dynamic Measures for Performance Assessment in Complex Environments
    - Erland Svensson (FOI - Man-System-Interaction)
    - Carin Rencrantz (FOI - Man-System-Interaction)
    - Jenny Lindoff (FOI - Man-System-Interaction)
    - Peter Berggren (FOI - Man-System-Interaction)
    - Arne Norlander (FOI - Man-System-Interaction)

2. Supporting the Navy Sailor Through Console Design
    - Melissa Weaver (Basic Commerce and Industries, Inc.)
    - Nancy Burton (Naval Surface Warfare Ctr.)
    - Katie Hall (Naval Surface Warfare Ctr.)
    - Jeanette Smith (Naval Surface Warfare Ctr.)

3. Lessons Learned from a Comparative High-Fidelity Usability Evaluation of Anesthesia Information Systems
    - Anjum Chagpar (University Health Network)
    - Joe Cafazzo (University Health Network)
    - Tony Easty (University Health Network)

4. Evaluation of a Tactile Navigation Cueing System and Real-Time Assessment of Cognitive State
    - Michael Dorneich (Honeywell Labs)
    - Patricia Ververs (Honeywell Labs)
    - Stephen Whitlow (Honeywell Labs)
    - Santosh Mathan (Honeywell Labs)

5. An Example of Objective Warfighter Performance Measurement: The Joint Distributed Free-Play Event
    - James Fielder (U.S. Army Aberdeen Test Ctr.)
    - Jay Winters (Trideum Corp.)
    - Katrina Baker (U.S. Army Aberdeen Test Ctr.)

CALL FOR PARTICIPATION IN SPECIAL SESSION ON PRODUCT DESIGN

Lois Smith forwarded the following e-mail to Newsletter Editors and Webmasters:

This year marks the beginning of celebrations in honor of the 50th Anniversary of the Human Factors and Ergonomics Society. It is also the 40th anniversary of the Product Design Technical Group.

In celebration of these events, we have scheduled a session, "Forty Years of Product Design," on Thursday, October 19th from 8:30 until 10AM. The purpose of this session is to highlight superior – indeed some of the best - commercial, personal, or military designs of the last 40-50 years.

For this session, we are soliciting members of the Society to present a vignette or "shorts" - each lasting 3-5 minutes - highlighting the contributions made by HFES to product design during the past 40 years.

One such example will be presented by Andy LeCocq. Andy worked for Texas Instruments and has several interesting stories to share surrounding the development of the Speak and Spell - a revolutionary toy when it was released in the late 1970's.

Presentation slots are limited, and if some of the early entries are indicative of what is to come, this promises to be a very enjoyable session.

Please FORWARD your response to Steven Belz (stevenbelz@gmail.com) by September 15 with the product you would like to present. DO NOT REPLY TO THIS LIST. And feel free to distribute this message to your TG members.

It is our goal to include as many of these as possible; however, we also need to ensure that the same product is not represented multiple times. Presenters will be notified the first week in October as to the order of presentation.

Thanks,
Steven M. Belz
Program Chair, Product Design Technical Group

Test and Evaluation Technical Group Business Meeting

!NEW TIME!
Wednesday, October 18, 2006
7:00 am – 8:30 am
Breakfast will be served

SPEAKER

Donald E. Farr

***NEXT 50 YEARS* OF HFES T&E:**
"THE FARR VIEW"

RSVP
(acceptances only)
drbittner@comcast.net
joyce.cameron@worldnet.att.net