Author's note: This document is the extended abstract originally submitted to HCIC. The full paper presented at the consortium is being revised for submission to a journal and is not available at this time.

Interfaces on the Road: Rapid Evaluation of In-Vehicle Devices

Dario D. Salvucci
Dept. of Computer Science
Drexel University
Philadelphia, PA 19104
+1 215 895 2674
salvucci@cs.drexel.edu

Extended Abstract

As user interfaces move off the desktop, they have become increasingly prevalent in our everyday lives and in one environment in particular: our cars. Private vehicles are used for 91% of all personal travel in the United States (Federal Highway Administration, 2001), with the average American spending about 80 minutes in their car every day — up from about 60 minutes just over a decade ago in 1990 (Polzin et al., 2003). Meanwhile, coinciding with the burgeoning use of secondary devices — cellular phones to navigation devices to PDAs — in vehicles, "driver inattention" has become the primary cause of vehicle crashes (22.7%), ahead of even excessive speed (18.7%) and alcohol impairment (18.2%) (Hendricks et al., 2001).

Given the importance of in-vehicle interfaces and their effect on our lives, effective evaluation methods for such devices are sorely needed to determine how they might affect driver behavior and performance. The field of human-computer interaction has enjoyed an enormous body of research devoted to techniques for evaluating, and thus refining and improving, user interfaces. From collection of verbal protocols (e.g., Ericsson & Simon, 1984) to heuristic evaluation methods (see, e.g., Nielsen & Phillips, 1993) to predictive computational modeling (e.g., Card, Moran, & Newell, 1983), the literature contains a rich variety of evaluation methods suited for a variety of purposes. However, the vast majority of these techniques have been developed for and tested in the context of desktop interfaces, making them inadequate for evaluating in-vehicle interfaces in at least two significant ways. First, use of in-vehicle interfaces requires multitasking between the primary driving task and the secondary interface task, and thus evaluation methods must understand the requirements of driver behavior and how interface use can be interleaved with driving. Second, the most common evaluation measures for desktop interfaces, such as time on task, errors, and user preference, are often not ideal for in-vehicle interfaces, where safety and driver performance are the critical measures for evaluation.

We are currently working on a new approach to evaluating in-vehicle interfaces quickly and easily. The approach is embodied by a multi-stage process as follows:

1. **Specification**. Given some description of a new device interface (whether implemented, prototyped, or simply verbally described), a designer thinks up sample tasks and model these tasks using the ACT-Simple modeling language (Salvucci & Lee, 2003), which closely resembles the well-known KLM framework (Card, Moran, & Newell, 1983). ACT-Simple provides a variety of operators that specify keystrokes, mouse movements and clicks, etc. by simply writing a sequence of commands such as (press-key), (move-mouse), (click-mouse), etc.

2. **Compilation**. While ACT-Simple resembles KLM in spirit and functionality, the framework has the advantage that it can be translated, or "compiled," down to a lower-level architecture

— namely, the ACT-R cognitive architecture (Anderson & Lebiere, 1998). The compilation process results in an ACT-R model that utilizes all the built-in predictions of the ACT-R architecture, and also has the benefit of being amenable to integration with other ACT-R models (unlike a typical KLM or related framework).

3. **Integration**. The compiled ACT-R model can now be integrated with a model of driver behavior with the goal of predicting driver behavior when performing both the driving and secondary (interface) task. For this purpose, we use an existing ACT-R model of driver behavior (Salvucci, Boer, & Liu, 2001) that has been validated for basic driving behavior as well as integration with secondary-task models (e.g., Salvucci, 2001).

4. **Simulation**. The final integrated ACT-R driving/task model can now be run in simulation, given a full simulated environment (e.g., a two-lane highway with traffic) and producing a full complement of predicted behavior; in fact, the model produces identical protocols as human drivers navigating the environment in a driving simulator, including steering, acceleration, even eye-movement data. The behavioral data can be analyzed in a host of ways to examine not only task time but actual measures of driving performance — for instance, lateral deviations of the vehicle, speed deviations with respect to a lead car, braking reaction time and distance, etc.

A number of issues and challenges arise in the application of this approach to real-world problems. We have been grappling with two issues in particular. First, unlike models for one-task desktop interfaces, the secondary task models require specification of when driving may be interleaved with the secondary task. Do drivers performing a secondary task look back at driving after every step in the task, at the completion of every subgoal, or only after the entire task is complete? We are exploring ways both of automating the placement of such interleaving events and of allowing the user to specify these manually. Second, standard methods for treating "mental operators" ('M's) in secondary-task models do not seem completely adequate for application to driver distraction. For instance, does the cognitive processor stall for the entire time of such an operator (1.20 – 1.35)? Can this time be interleaved with driving? If so, when/where can the interleaving happen? Initial explorations have led us to one straightforward solution of carving the mental operator time into multiple segments, in between which some driving updates can take place, and we are attempting to estimate the number of segments empirically using empirical results in a cell-phone dialing task.

We believe that the proposed approach can help "close the loop" for the development of in-vehicle interfaces by providing a rapid process (approximately several hours) to move from interface ideas to recognized real-world measures that quantify driver distraction. Also, the approach could potentially be augmented with a prototyping front-end such that the designer can, for instance, quickly mock-up the desired interface and generate models automatically by demonstration (e.g., John et al., submitted). While we would not expect such a system to single out the one best idea from hundreds of possible ideas, we hope that it might be used to narrow down the field of ideas to a much smaller set (3-5), thus quickly eliminating bad possibilities and allowing more time to fully explore the few best possibilities.

Acknowledgments

The author was supported by a grant from the Office of Naval Research (N00014-03-1-0036) and by a gift from Ford Motor Company.

References

Anderson, J. R., & Lebiere, C. (1998). *The atomic components of thought*. Hillsdale, NJ: Erlbaum.

Card, S., Moran, T., & Newell, A. (1983). *The psychology of human-computer interaction*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Ericsson, K. A., & Simon, H. A. (1984). *Protocol analysis: Verbal reports as data*. Cambridge, MA: MIT Press.

Federal Highway Administration (2001). Our nation's highways 2000 (Tech. Rep. No. FHWA-PL-01-1012). Washington, DC: U. S. Department of Transportation.

Gray, W. D., John, B. E., & Atwood, M. E. (1993). Project Ernestine: A validation of GOMS for prediction and explanation of real-world task performance. *Human-Computer Interaction*, *8*, 237-309.

Hendricks, D. L., Freedman, M., Zador, P. L., & Fell, J. C. (2001). The relative frequency of unsafe driving acts in serious traffic crashes. Report for Contract No. DTNH22-94-C-05020, National Highway Traffic Safety Administration, U.S. Department of Transportation.

John, B. E., & Kieras, D. E. (1996). The GOMS family of user interface analysis techniques: Comparison and contrast. *ACM Transactions on Computer-Human Interaction*, *3*, 320-351.

John, B. E., Prevas, K., Salvucci, D. D., & Koedinger, K. (submitted). Predictive human performance modeling made easy. Submitted to CHI 2004.

Nielsen, J., & Phillips, V. L. (1993). Estimating the relative usability of two interfaces: Heuristic, formal, and empirical methods compared. In *Proceedings of INTERCHI 1993* (pp. 214-221). New York: ACM Press.

Polzin, S. E., Chu, X., & Toole-Holt, L. (2003). The case for moderate growth in vehicle miles of travel: A critical juncture in U.S. travel behavior trends. 2001 National Household Travel Survey, http://nhts.ornl.gov/2001/articles/index.shtml.

Salvucci, D. D. (2001). Predicting the effects of in-car interface use on driver performance: An integrated model approach. *International Journal of Human-Computer Studies*, *55*, 85-107.

Salvucci, D. D., Boer, E. R., & Liu, A. (2001). Toward an integrated model of driver behavior in a cognitive architecture. *Transportation Research Record*, 1779.

Salvucci, D. D., & Lee, F. J. (2003). Simple cognitive modeling in a complex cognitive architecture. In *Human Factors in Computing Systems: CHI 2003 Conference Proceedings* (pp. 265-272). New York: ACM Press.