

**OBJECTIVE METHODS FOR ASSESSMENTS OF INFLUENCE OF IVIS
(IN-VEHICLE INFORMATION SYSTEMS) ON SAFE DRIVING**

Petr Bouchner, Stanislav Novotný, Roman Piekník
Driving Simulation Research Group
Laboratory of Systems Reliability
Czech Technical University in Prague
Faculty of Transportation Sciences
Czech Republic
E-mail: bouchner@lss.fd.cvut.cz

Summary: The main objective of experiments focused on assessments of in-car devices (IVIS in this case) is to evaluate objectively their influence on safety and comfort of driving. This paper presents a big set of experiments performed on our driving simulators. Besides the experiment setup, it describes several different methods used for the classification of driver behavior based either on analysis of technical outputs from simulated driving or subjective and psycho-physiological measures. The aim of the whole project is to find out the objective methodology (a complex set of methods) for assessment of IVIS (or further any HMI devices used in the cars) with respect to drivers' skills and personalities.

INTRODUCTION

The problem of sufficient attention of a human operator is still a very important topic. There is a strong need for investigation into the safety and comfort of use of IVIS. Its importance is proven by many national and international projects interested in driving and HMI—the European E-safety for instance (ESoP, 2005). If the evaluation of this influence is to be objective, we need to measure lots of experiments under safe, laboratory conditions. Advanced driving simulators have proved to be very suitable for that task. Common evaluations of suitability of IVIS are in the majority of cases done by experts. The assessment of the rate, how each particular device influences safety of driving, are mostly based on subjective measures and expert scoring. Those are always loaded by certain error caused usually by the small number of tested subjects, not covering statistically all possible groups of further users. Because of this fact, there is still strong need for objectification of such evaluation, as can be found for example in Carryl & Baldwin (2006), Gärtner et al. (2001), or Zoldan et al. (2006).

Our approach is based on the idea that improper behavior, if measured as an output from the drive, should appear similarly regardless of the fact if it is primarily caused by drivers' drowsiness or by distraction by unsuitably designed IVIS. We had done a lot of experiments aiming at detection and classification of driver drowsiness either on the simulator or in simplified laboratory environments. A complex of objective methods approved as suitable for drowsiness detection (Vysoký, 2004; Bouchner et al., 2006) was implemented on IVIS assessment measurements (Novák et al., 2003; Novák et al., 2006).

EXPERIMENT

Navigation systems are among the most used IVIS devices in contemporary cars. The majority of mid- and high-class cars contain those devices in the standard package. The use of such navigation systems consists of two activities; entering the information and obtaining the information. From the point of view of navigation systems, obtaining information is usually either *watching the map or listening to the audio guidance*.

These activities can of course distract the driver from primary driving tasks. Entering the targets into the navigation systems while driving appears to be the most demanding and distracting task. Sometimes it requires so much time that the driver can completely lose track of the situation on the road. The aim of our investigation was to detect (and compare if possible) the influence of such devices on comfort and safety of driving.

The simulator

The simulator we used for this experiment is steady based and fully interactive, composed of a fully equipped cockpit of a higher-middle class European car with automatic gear-shifting. It is equipped with a system of measuring devices and an in-car video recording system. The field of view is 100° with no active back mirrors. A description of the simulator and its usage can be found in Bouchner & Novotny (2005) or Bouchner (2004).

Testing track

The testing track is divided into two parts: an easy and a demanding one. The driver was asked to keep the speed at 90 km/h for the easy track and 50 km/h for the demanding one. From the perspective of the driver, the easy track seems to be almost straight. The radii of curve for the easy part were set at about 2000m, and for the demanding track it was decreased up to 300m. Figure 1 shows a track top view (left) and a screenshot from our scenery (right).

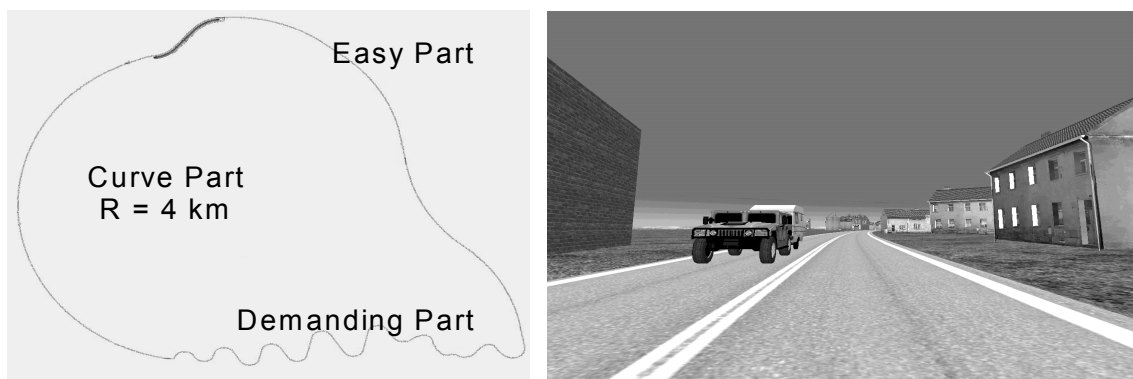


Figure 1. Testing track - track top view (left) and a screenshot from our scenery (right)

Tested devices and cohort

We performed 9 different experiments with 9 different setups; 24 experimentees participated in each experiment. All those experiments had to examine the influence of several different interfaces (devices) that allow the driver to input the target city into the navigation system. The interfaces were of different design, size, position, input/output type (detailed descriptions cannot be published since the devices were commercial prototypes). It was required that each experimental group would include 30% older people (>58 years) and 30% women. All of them were non-professional but skilled and active drivers familiar with simulator driving. All also had to be healthy and fresh during the experiment.

Testing procedure

The primary task for the driver was to drive safely following common rules. The secondary task was to insert the name of the target city into the navigation system. The stimuli were audio-visual, composed of a ring signal and text with the target city projected on the screen. The driver's task consisted of three actions:

- Keeping the appropriate position within their lane
- Keeping the speed defined by traffic signs
- Inserting the name of a given target city into the navigation system correctly and as fast as possible

EVALUATION OF EXPERIMENTS

Measured data

During the experiments, technical and psycho-physiological data were collected. All data were synchronized so that it is possible to do further correlation analysis on them. Before the experiment, case histories of the experimentees (including driving skills, etc.) were collected and after, they were asked to do a subjective evaluation. (See Table 1 for reference.)

Table 1. Data measured during experiments

Technical data	Biological data	Subjective data
Trajectory and geometrically ideal path	EEG	Questionnaires
Actual velocity vector	EOG	TLX
Steering wheel and pedals	ECG	Self rating
Video recording of the scenery and experimentees	Head movements	Anamnesis

Although the complex sets of outputs were measured, the analysis was focused mainly on the technical outputs. This was done mainly because these data are considered less “personality dependent” and their analysis is much more straightforward.

Trajectory analysis

The rate of deviation from the geometrically ideal path was chosen as the most important measure. The greater the driving deviations, the more dangerous the situation in real traffic. The

deviations from the reference curve are measured as the distance from a rigid point (we chose a middle point between the front wheels) from the geometrically ideal path. Our virtual road is constructed from polygons, which approximate real curvature of the testing track. We used a spline interpolation of the middle lane points, which are in between each couple of diametric vertices. This creates an appropriate “ideal path.” Figure 2 (left) shows the exemplary case of driving with and without the load.

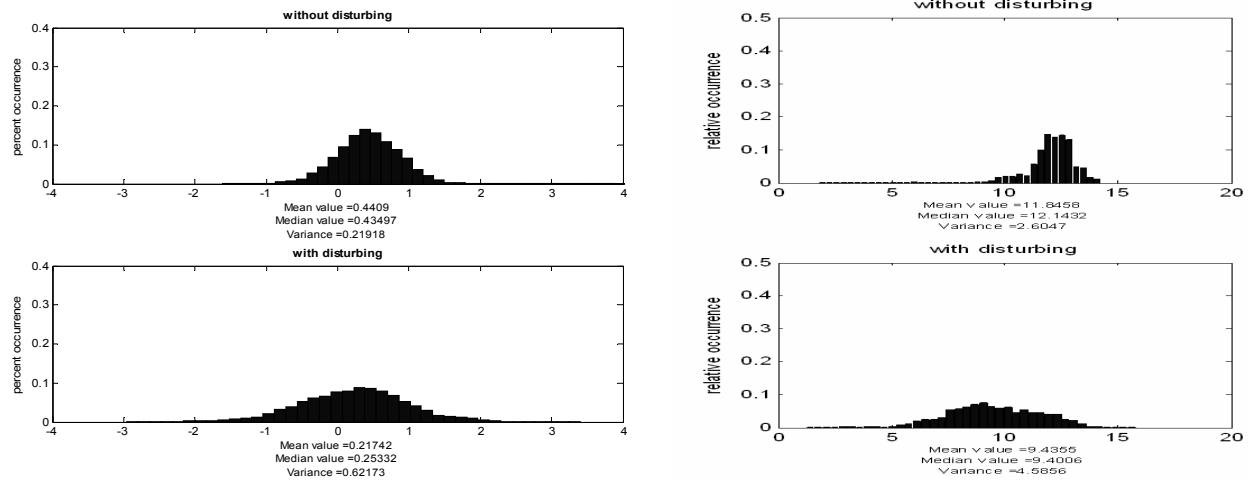


Figure 2. Normalized histograms of speed fluctuation upper without load (left) and normalized histograms of path deviations upper without load (right)

Speed fluctuation analysis

We supposed drivers’ corrections of the speed not only during secondary task but also after it was completed (when the driver again fully concentrates on driving). It required additional standardizing rounds to be performed. The resulting graphs (Figure 2, right) show the exemplary case of different driver’s ability to keep a constant speed; they describe the variance and average values of a car velocity measured for various drivers driving on the simulator. The differences in speed variations are evident in both the demanding and easy parts of the testing track. The differences in average velocities are negligible.

Response time

Both the reaction time and the time needed to fulfill the task (in our case enter target city into the navigation system) are loaded with an error caused by drivers’ habits and the responsibility of safe driving. They also depend on the situation on the road; whether the driver decides in each particular situation to focus more on the primary or on the secondary task. All the experimentees were instructed primarily to drive safely, but we cannot be sure if they kept this in mind during the whole experiment. In reality, the response time is composed of several cognitive and transport delays. For our purposes, a rough approximation (which sums them up) suffices. The placement of each particular tested device is approximately the same, so the reaction time is expected to give evidence of the complexity of the driver’s intended action. The time needed to fulfill the task should also testify about the quality of a design and the ergonomics of the tested devices, since all the experimentees enter the same sequence of targets.

Psycho-physiological measures

The EEG analysis was the only one that we have included in the set of evaluation methods so far. We focused mainly on frequency changes in alpha and delta bands on T and O electrodes, since these should testify about the cognitive demands of the task. Unfortunately, the heart activity (ECG) data were full of artifacts and noisy and could not be brought into the final evaluation. The EOG was not expected to give results because of its very noisy character—in future experiments it will be replaced by an eye-tracker.

Subjective measures

TLX. Workload is defined as the physiological and mental demands that occur while performing a task or a combination of tasks. The NASA Task Load Index (TLX) (Rubio S., et al., 2004) incorporates six subscales: mental demand, physical demand, temporal demand, performance, effort level, and frustration level. Because of the fact that the experimentees did not understand the real difference between demands, we did not use it for further experiments. For reference the group of experimentees had to pass independently experiments with primary task only (no additional load but simple driving) and afterwards they were examined for the workload (using TLX) caused by the simulator itself. Figure 3-left shows the statistics from 2 sets of experiments and one set with reference driving without any secondary task.

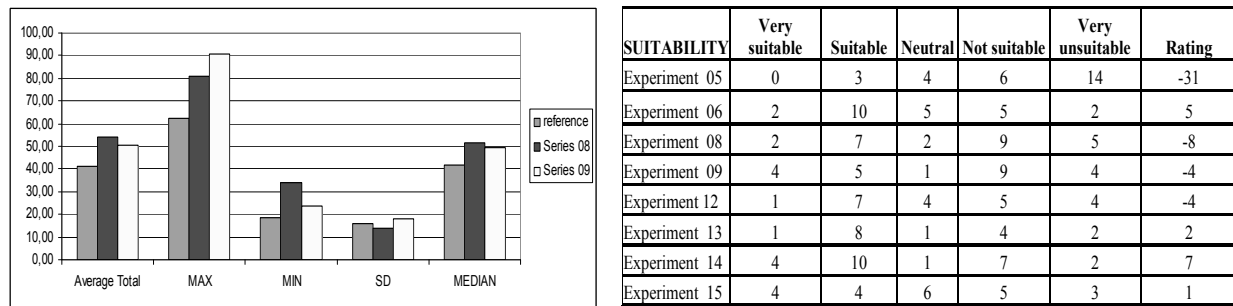


Figure 3. Subjective workload evaluation - TLX (left) and Likert scale (right)

Questionnaires. All experimentees had to complete a complex questionnaire asking for specific features of each device (placement of input, output interface, legibility of particular visual output, etc.), which cannot be applied for the overall evaluation. For comparison of methods, only a final evaluation question is discussed here. This is: “*Would you recommend the tested setup for usage in real cars?*” For that purpose, the five degree Likert scale was used (Barnett, 1991), with neutral opinion in the middle (Figure 3, right).

COMPARISON OF THE ANALYSIS APPROACHES

The next table (Table 2) shows the average values per each set of experiments. Columns represent measures and their analysis. Rows represent the experiments (consisting of 24 experimentees). Results per experiment are divided into four groups: all experimentees, women only, men only, and elderly drivers (>58 years old) of mixed gender. Technical data from the experiments 5 and 6 were not analyzed due to their incompatibility.

Table 2. Data measured during experiments. VAR- variation, DEM- demanding part of the circuit, NL – not loaded (driving only) L – loaded (driving with secondary task)

Measurement #	Trajectory VAR		Trajectory mean [m/s]		Velocity fluctuation				Time of first response [s]	Time needed to fulfill the task [s]	Likert scale	TLX	EEG [μ V eff]	
	NL	L	NL	L	NL		L						delta	alpha
					easy	DEM	easy	DEM						
Experiment No. 5														
ALL experimentees	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	3,0690377	40,59593	-31	N/A	5,1773	3,1988
Eldery drivers	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	3,0111463	46,11832	N/A	N/A	6,2808	3,4808
Women (Mixed ages)	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	3,2221683	38,29584	N/A	N/A	4,8594	2,9156
Men (Mixed Ages)	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	2,9737985	37,37361	N/A	N/A	4,3917	3,2
Experiment No. 13														
ALL experimentees	0,240029	0,397248	0,277705	0,179384	41,16102	13,35731	14,30627	1,700045	3,138028	44,42879	2	N/A	4,042133	2,380133
Eldery drivers	0,25249	0,728277	0,35124	0,118413	44,72894	13,86024	17,36587	2,034637	3,9035745	53,59529	N/A	N/A	4,9286	2,4071
Women (Mixed ages)	0,279678	0,37883	0,30601	0,258396	41,23443	13,17319	11,42047	2,213428	2,4991479	39,32333	N/A	N/A	4,0786	2,8179
Men (Mixed Ages)	0,200704	0,271166	0,221928	0,13779	39,56899	13,29959	15,46856	1,116606	3,0113616	40,36774	N/A	N/A	3,1192	1,9154
Experiment No. 9														
ALL experimentees	0,327108	0,292029	0,262171	0,235732	23,88438	5,973124	16,56861	4,887857	3,1997376	37,35482	-4	50,38095	5,144767	2,494167
Eldery drivers	0,556548	0,47264	0,160273	0,155174	18,10861	4,825039	15,68076	4,024204	3,5700381	45,92225	N/A	45,33333	5,0382	2,8647
Women (Mixed ages)	0,281928	0,272326	0,213616	0,176711	20,6896	5,608961	20,0854	4,424019	3,2714412	36,65707	N/A	N/A	57,71429	4,985
Men (Mixed Ages)	0,234977	0,214116	0,337397	0,305522	28,36965	6,729249	15,25415	5,497346	2,7577335	29,48513	N/A	47,26667	5,4111	2,3278
Experiment No. 6														
ALL experimentees	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	2,5503397	38,89353	5	N/A	5,411633	2,743867
Eldery drivers	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	2,7214295	47,97962	N/A	N/A	7,34	3,07
Women (Mixed ages)	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	2,1794165	31,88104	N/A	N/A	5,3722	3,3889
Men (Mixed Ages)	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	2,7501732	36,81992	N/A	N/A	3,5227	1,7727
Experiment No. 8														
ALL experimentees	0,282024	0,266617	0,277586	0,277853	17,41384	8,446899	12,10956	5,430432	3,8681583	17,93462	-8	54	6,692933	3,152167
Eldery drivers	0,38249	0,370735	0,181433	0,175243	11,25535	9,65327	11,53354	6,195279	5,1284025	26,24926	N/A	54,53334	3,41	2,31
Women (Mixed ages)	0,253307	0,208679	0,291076	0,277421	14,94995	7,378129	9,571274	6,584934	3,2538684	13,1389	N/A	54,85714	7,515	3,135
Men (Mixed Ages)	0,266574	0,265778	0,299909	0,309657	20,63546	8,651201	13,65357	4,573439	3,222204	14,4157	N/A	53,13334	9,1538	4,0115
Experiment No. 12														
ALL experimentees	0,235692	0,389481	0,261871	0,177087	25,39942	11,20155	26,78843	1,561899	4,035299	16,92607	-4	N/A	5,4885	2,456833
Eldery drivers	0,295008	0,636362	0,267827	0,000108	22,55466	11,92891	29,07572	2,677541	5,309566	22,94839	N/A	N/A	5,3944	2,1444
Women (Mixed ages)	0,20018	0,305583	0,187802	0,207689	23,27243	10,2189	18,80621	0,726291	3,6207017	15,65355	N/A	N/A	6,2611	3,0111
Men (Mixed Ages)	0,228768	0,33753	0,285826	0,235488	28,20553	10,89524	29,95894	1,489546	3,1756295	12,17627	N/A	N/A	4,81	2,215
Experiment No. 14														
ALL experimentees	0,212335	0,340116	0,243023	0,165635	37,22245	15,11968	10,98951	1,726759	2,5108946	37,46681	7	N/A	4,438146	2,468622
Eldery drivers	0,25095	0,400955	0,10366	-0,16834	21,03027	15,12514	13,05681	2,83677	3,4867869	50,40908	N/A	N/A	5,316667	2,313889
Women (Mixed ages)	0,22517	0,334834	0,198816	0,062509	43,38034	15,05498	11,13227	1,623256	2,1876487	33,52998	N/A	N/A	4,890909	2,777273
Men (Mixed Ages)	0,195628	0,331646	0,301841	0,304619	36,15035	15,16388	10,47611	1,577208	1,8582481	28,46137	N/A	N/A	3,106863	2,314706
Experiment No. 15														
ALL experimentees	0,213623	0,34654	0,292319	0,189088	46,91776	15,8592	10,35566	2,205668	2,2541471	40,87823	1	N/A	4,195667	2,620202
Eldery drivers	0,22123	0,3071	0,401265	0,165353	61,94083	15,91511	10,49186	2,342239	2,9298854	56,55859	N/A	N/A	3,642647	1,672059
Women (Mixed ages)	0,245555	0,43656	0,32459	0,177628	41,12882	15,93255	10,34593	2,408863	1,9430253	34,37211	N/A	N/A	4,275	3,995
Men (Mixed Ages)	0,1898	0,299673	0,23449	0,20464	45,76935	15,79166	10,31674	2,024681	1,8895306	31,70397	N/A	N/A	4,669355	2,193548

DISCUSSION AND CONCLUSION

Experts’ and experimentees’ evaluation are not objective—because of their subjective nature. The drivers who passed the experiment sometimes positively evaluated the tested system, but their behavior on the road (simulated road) was dangerous. On the other hand, those devices that were not positively classified by the experimentees did not significantly influence safety of driving. For that reason, objective classification is needed to accompany the subjective measures. Questionnaires for detection of driver’s load (TLX) gave us ambiguous results, since the experimentees did not understand them enough. The analysis of deviation variance confirmed the expectation that a considerable increase of variance of deviations from the ideal path was noticed during experiments where the driver spent long time periods doing the secondary task. The expected connection between the drivers’ task-load and speed fluctuation was in the opposite direction to that we had expected: driving while performing a secondary task resulted in less speed fluctuation. This might result from the fact that the driver anticipates the situation. He estimates the time he can spend on the secondary task, and then tries to keep a constant speed while performing the task in order to meet his expectations about the time he could spend.

Good and reliable classification tools based on psycho-physiological measures require good

knowledge about each one's personality and "standard" behavior; also the appearance of brain waves is very individual. For that reason the use of psycho-physiological measures only is not suitable, since it gives very individual results and would require the testing of a large number of persons covering all the possible target groups. Unfortunately we were not able to create scale and order the tested devices depending on their impact on safe driving. To create a really objective evaluation, which is able to order and decide which of the devices influence the drivers more, it is necessary to apply a combination of several different methods.

REFERENCES

- Barnett, V. (1991). *Sample Survey principles and methods*. Edward Arnold, London.
- Bouchner, P. (2004). VR Simulation Device as a Support for Research in Driver's Micro-sleeps and Attention Decrease Detection. Athens: TRANSTEC 2004, 65-72.
- Bouchner, P., Novotný, S. (2005). System with driving simulation device for HMI measurements. 9th WSEAS Circuits, Systems, Communications and Computers Multiconference, Athens.
- Bouchner, P., Hajný, M., Novotný, S., Piekník, R., Pěkný, J. & Valtrová, K. (2006). Analysis of technical and biological outputs from simulated driving, focused on driver's fatigue detection. Driving Simulation Conference, Tsukuba, Japan.
- Baldwin, Carryl L. (2006). User-centred design of in-vehicle route guidance system. Yvonne B. (Eds.), *Developments in human factors in transportation, design and evaluation*. Maastricht, Netherlands: Shaker Publishing, 43-49.
- European Commission, Information Society (ESoP). (2005). *Final report of the eSafety Working Group on HMI*. Available at http://ec.europa.eu/information_society
- Gärtner, U., König, W., & Wittig, T. (2001). Evaluation of Manual vs. Speech Input When Using a Driver Information System in Real Traffic. International driving symposium on human factors in driver assessment, training and vehicle design, Colorado, USA.
- Novák, M., Fáber, J., Votruba, Z., Přenosil, V., Tichý, T., Svoboda, P. & Tatarinov, V. (2003). Reliability of Human Subject – Artificial System Interactions. *Acta Polytechnica*, 42, 31-40.
- Novák, M., Votruba, Z., Svítek, M. & Bouchner, P. (2006). Improvement of Bus and Truck Driving Safety. IEEE International Conference on Systems, Man, and Cybernetics. Taipei, Taiwan.
- Rubio, S., Díaz, E., Martín, J., Puente, J.M. (2004). Evaluation of Subjective Mental Workload: A Comparison of SWAT, NASA-TLX, and WP Methods. *Applied Psychology*, 53(1), 61–86.
- Vysoký, P. (2004). Changes in Car Driver Dynamics Caused by Fatigue 1, 109-117. SSGRR - 2004, NNW 14. Piacenta: ICS UNIDO.
- Zoldan, C., Toffetti, A., & Marendo, E. (2006). Usability and ergonomic evaluation of in-car multifunction system: a tool for Human-Factors Professional. Yvonne B. (Eds.), *Developments in human factors in transportation, design and evaluation*. Maastricht, Netherlands: Shaker Publishing, 63-72.