European Commission

Information Society Technologies

HUMAN Factors

6th Framework Programme
Priority 2 "Information Society Technologies"
"**HUMAN** centred design for **I**nformation **S**ociety **T**echnologies"

Proposal n°507 420
Contract n°507420

Final proposal for common methodologies for analysing driver behaviour

Deliverable 6 of Task Force 2

Reference: 2TNO-070129-E1-DA(1)

CONFIDENTIALITY LEVEL : CONSORTIUM

Rev.	Issuing date	Pages	Written by	Visa	Veri-fied by	Visa	Approved by	Visa
A	January 29, 2007	27	W. Janssen, TNO Human Factors	V	Sophie Boets, BIVV Carlo Cacciabue, JRC			
Modifications : Creation of the document.								
Rev.	**Issuing date**	**Pages**	**Written by**	**Visa**	**Verified by**	**Visa**	**Approved by**	**Visa**
A(1)	November 29, 2007	26	W. Janssen, TNO Human Factors	V	Sophie Boets, BIVV Carlo Cacciabue, JRC	V		
Modifications : Accommodation of Audit comments and recommenations								

TABLE OF CONTENTS

1. Introduction

HUMANIST Task Force 2 works on the methods and tools in the NoE for measuring and analysing driving behaviour under the influence of Intelligent Transport Systems (ITS).

One part of the TF 2 effort is to develop the tools themselves to their maximum potential, and the other is to reach some form of agreement on a common methodology on the basis of the available tools. This Deliverable deals with the latter, though it obviously must take into account what is going on in the former.

The rationale for wanting to agree on some form of common methodology comes from the observation that there is often so little commonality in the research literature. This makes it hard, e.g., to compare findings from different studies or to draw general conclusions from them. This is why so many European Projects have taken precautions in actually spending a lot of their effort on developing and prescribing a common methodology for empirical behavioural studies to be performed *within* the particular Project. Repeatedly this has proven to be very difficult, so that an even more common framework *over* Projects has been slow to emerge naturally.

Apparently, this further step has to be taken explicitly, and HUMANIST presents an opportunity for doing it.

It is important to be explicit in stating what is the ‘common’ element we are aiming for. The aim is *not* to prescribe what NoE researchers should and should not do in a particular case, or to raise doubts on their craftsmanship when they design a driver behaviour study. Rather it is to formulate a set of recommendations, reflecting the state-of-the-art, for which agreement exists to be a *minimum* to be applied in ITS evaluation studies that have a driver behaviour component.

The Report is based on the earlier deliverable D 2.3, ‘Proposal for common methodologies for analysing driver behaviour’ (Janssen, 2006), that has been amended, and to which material has been added, after internal discussions within Task Force 2 and between this Task Force and, in particular, Task Forces E and D.

1.1 Elements to be covered by ITS behavioural studies

There usually is a large number of relevant behavioural aspects to be associated with driver information and support systems. A major distinction in this respect is, of course, between IVIS and ADAS. Whereas an IVIS is of an informative nature, and thus acts on driver behaviour in a somewhat indirect way, an ADAS is designed for the purpose of interacting directly with the driver’s behaviour. Furthermore, since an IVIS is meant for something else than supporting the driving task itself its primary benefits may lie outside the driving task proper, raising questions of how to assess trade-offs between benefits on one level and (possibly) increased risks on an other. Related to this the degree of self-pacing that IVIS permit is usually larger than for ADAS. For example, if the Collision Avoidance System gives an alarm your response had better be immediate, whereas you can decide to leave the consultation of the hotel list of your navigation system to a later time. Finally, an important negative effect associated with an IVIS may be distraction of the driver, while this is much less the case with ADAS.

As a consequence of these basic differences the design of behavioural studies may have to be slightly different for IVIS than for ADAS.

Taking this into account, the major aspects to be covered in driver behaviour studies with ITS appear to be the following:
(1) (For IVIS in particular): The possible trade-offs between benefits on one level and costs (risk) on the other.
(2) ITS effects on behavioural and vehicle parameters.
(3) ITS effects on driver distraction, inattention, workload and other driver state parameters.
(4) Usability and acceptance of ITS, in the context of behavioural studies
(5) Estimated overall ITS effects on accident risk, to be derived from some combination of behavioural and workload/state parameters.

A major choice to be made preceding all this is, of course, whether the study will be done in a simulator, on the road, in the lab (mini-sim), etc. This issue, although of extreme importance, cannot really be a subject of this Deliverable. What we must be prepared for, however, is that different experimental environments require different methodologies, parameters, and analysis tools. Therefore this differentiation has to be made whenever required.

1.2 General considerations of experimental design and conducting experiments

Issues of experimental design and the way in which experiments should be conducted in a practical sense are outside the scope of this Report. It should be mentioned, however, that a very important aspect of the latter – the ethical standards to be applied when experimenting with human subjects – is covered in another TF 2 activity (see Hanzlíková, 2004, and Macku, 2007).

2. Primary ITS effects

As said, a major difference between IVIS and ADAS systems is that the effects of ADAS are directly on driving behaviour itself, while an IVIS is not designed to do that, although it may have behavioural effects as well.

Thus, in case of an IVIS the benefits are on a different level than the driving task per se. While it may be relatively easy to assess these benefits by themselves, it may be very complicated to bring them together with driver behaviour effects and to come to a judgement on what the ‘net’ effects of the IVIS are.

An elementary example of this is navigation systems. These may save lives simply because they reduce excess mileage, but they may at the same time exert a negative effect on driving behaviour proper because they distract the driver or leave him/her in thought on the messages presented. Cell phones are another example, with an even more complex pattern of pros and cons on different levels.

It is hard to think of a general methodology incorporating effects on these different levels, and we have not been able to solve this problem within the HUMANIST Task Force 2 work. The solution for putting different things into one bin must come from other approaches, like cost-benefit analysis. Nevertheless, it is recommended to the behavioural researcher to show at least an awareness of the existence of trade-offs by trying to get hold of estimates of the magnitude of the primary beneficial effect. In case of a navigation system, for example, it would be highly recommendable to estimate how many miles have actually been saved by the system. Common estimates are in the order of 5 to 7 % of total mileage, but this should be verified in specific cases. In no case should the impression be created that we are always looking for negative effects of ITS systems without recognising that they have major beneficial effects on a higher level. The same applies to ADAS studies, though not to the same degree, because their benefits *are* directly on the behavioural level. Nevertheless, here too we should try to assess a net effect rather than focusing on the negative by-products the system may have.

3. Driving performance parameters

Driving performance measures reflect the execution of the driving task itself, on the so-called manoeuvring and operational levels.

3.1 Inventory

The following Table 1 is taken from the earlier Deliverable (Janssen, 2006).

In this Table, and in the remainder of the Report, the following is meant by the different experimental environments:
- ‘Field‘: use of a subject-driven instrumented vehicle either on a closed track or on the open road
- ‘Simulator’: a full-fledged driving simulator
- ‘Laboratory’: all types of mini-simulators, comprising possible combinations of screen + steering wheel + gas pedal + brake.

As far as driving performance studies go a distinction is usually made between parameters describing the actions of the driver on the vehicle and the consequent parameters of the vehicle’s position and motion, both in the longitudinal and the lateral dimensions. Also, parameters describing the interaction with preceding vehicles (car following) belong here as a category.

An ‘OK’ in the Table means that the parameter can sensibly be measured, from both the technical point of view and with respect to its meaning or interpretation.

Table 1: Commonly used driving performance measures for different experimental environments

Measure	Field	Simulator	Laboratory
DRIVER ACTIONS:			
Steering wheel angle	Limited	OK	OK
Reversal rate	OK	OK	
High freq component of steering wheel angle	OK	OK	
Steering entropy	OK	OK	
Abrupt onsets of brakes	OK	OK	
Brake force		OK	OK
Use of accelerator	OK	OK	OK
Use of secondary controls		OK	OK
LONGITUDINAL:			
Travel time	OK	OK	OK
Distance travelled	OK	OK	OK
Average speed	OK	OK	OK

SD Speed	OK	OK	OK
Longitudinal acceleration	OK	OK	OK
LATERAL:			
Lateral position	OK (not urban)	OK	OK
SD Lateral position	OK (not urban)	OK	OK
Lateral speed	OK	OK	Difficult; accurate lateral position data required
Absolute lateral acceleration	OK	OK	OK
Road relative lateral acceleration	OK	OK	Difficult, accurate lateral speed data required
Time-to-line crossing (TLC)	OK (not urban)	OK	OK
Amount of time outside lane (LANEX)	OK (not urban)	OK	OK
INTERACTIONS:			
Time headway	OK	OK	OK
Distance headway	OK	OK	OK
Time-to-collision (TTC)	OK	OK	OK
Post-Encroachment Time (PET)		OK	OK

3.2 Recommended minimum set

The first thing to note is that many of these parameters must be related to each other: one cannot get out of one's lane without performing a noticeable action on the steering wheel, etc. This said, it is also true that all parameters are, so to say, individuals whose peculiarities should be known and respected before deciding whether to include or reject them. However, this is no plea to measure them all, just to be sure. Experience has taught us that problems of interpretation increase exponentially the more parameters are included. In such cases it often happens that there are, at best, arbitrary reasons for attributing more weight to one parameter than to an other in the final judgement.

Therefore a dimensional or factor analysis should be applied that – on the basis of a comprehensive data set - sorts out the existing correlations between parameters and yields a number of ‘pure’, underlying dimensions that exhaustively describe driving behaviour. What has been done in this area (Janssen, 1994; Bekiaris et al., 2006) suggests that the driving task has the following underlying, statistically independent, dimensions:

- Speed choice
- Speed variability
- Lateral positioning
- Car following.

What this says is that driving behavior can be described by just four variables, and that our task has become to find the functions relating each of those to accident risk. Note that there is still a choice to be made about what exact metric will be used in order to index each of the underlying dimensions. This is a separate problem, on which we have by now collected sufficient evidence to be able to come to specific recommendations.

On that basis, the proposed recommendation for the minimum set comprises the following, where the idea is that there should be at least one metric from each of the dimensions into which the measures can be divided:

- High frequency content of steering wheel signal: Probably the best representative of this category. A more or less equivalent measure is steering wheel reversal rate, but here a lot depends upon the exact efinition of ‘reversal’, and this has not yet been sorted out completely (Markkula & Engström, 2006).
- Average speed: Very directly related to accident risk.
- SD speed: also. Should always be considered together with the pattern in average speed to come to a sensible interpretation of speed effects.
- SD lateral position: this is an oldie, but still to be considered as a sensible and sensitive indicator of position keeping performance in the lane.
- TLC: takes it all together as a composite index of lateral performance. NB: Calculation and analysis may be tricky.
- LANEX: the result of it all, i.e., the nearest- to- real-risk index of lateral performance that may be derived from driving performance.
- Time headway: the primary index of car-following behaviour, and strongly related to the probability of a rear-end collision.

3.3 Observational methods

The above are all continuous measurements of driving behaviour per se, and one thing that they do not cover very well is the overall quality of handling of discrete events, in particular, interactions with other road users. In cases where this is relevant methods of an observational nature could be applied, such as the Wiener Fahrprobe (Risser and Brandstätter, 1985). This method is a behaviour observation method with the help of which one can get an over-all impression of the driving performance of a person. Two trained observers register, along a standardised route, speed adaptation in front of crossings or obstacles, over-taking manoeuvres, the distance to the car in front, the use of the indicators, lane keeping and lane changing behaviour, or the behaviour in the frame of communication with other road users, including vulnerable ones, among other things. The method gives the possibility to register not only just technical driving details, but also the drivers' interaction both with the physical and the social environment. One of the advantages of this holistic approach is that the types of behaviour that would be registered as identical ones with the help of an objective form of measurement can be differentiated according to their context. One example is speed in relation to the situation.

3.4 Conclusion

There is an abundance of parameters describing driver performance. A choice for a minimum set from these should ultimately be dictated by empirical arguments, i.e., a dimensional analysis of the smallest number of underlying dimensions that is sufficient to describe behaviour. The above set is the result of that. In cases in which a more general meaning should be attached to the interpretation of driving behaviour it should be considered to apply a more subjective, but broader, form of behavioural observation.

4. Driver state measures

An ITS will of necessity affect the state a driver is in: this is the other side of the coin in behavioural studies. Thus, while it is relatively straightforward to measure observable performance, these measurements need to be complemented by parameters reflecting relevant driver states. Workload and distraction is what occupies us in this respect.

4.1 Indicators for workload and distraction

The Table 2 below contains the most prominent parameters that are currently being used in driver behaviour research, i.e., that are available within the HUMANIST NoE.

Table 2: Commonly used driver workload and distraction measures for different experimental environments

Measure	Field	Simulator	Laboratory
Primary task performance	OK	OK	
Secondary task performance	OK	OK	
Peripheral Detection Task (PDT)	OK	OK	OK
Response time (RT) to unexpected events		OK	OK
Subjective workload measures	OK	OK	OK
For visual ITS: Glance frequency and duration	OK	OK	OK
Pupil parameters	OK	OK	OK
Occlusion	OK	OK	
Heart rate and variability	OK	OK	OK
Respiration	OK	OK	OK
Hormones	OK	OK	
EEG	OK	OK	OK

It is true that workload and distraction measures are particularly sensitive to system characteristics, even more so than performance parameters. A recommended minimum from the list would, nevertheless, comprise the following.

Recommendation:
- PDT: this is a ‘neutral’ and minimally intrusive secondary task, indexing primary task load in a sensitive way. RT to unexpected events (covering the quality of

event detection) is probably strongly related to PDT performance, so would not have to be considered separately, i.e., would not be needed to be included as an extra parameter.

- Subjective measures: these should better be included because they sometimes have been found to be discriminative while ‘objective’ parameters weren’t (e.g., in the HASTE Project: Johansson et al., 2005). A specific recommendation is to apply already widely-used scales like the NASA -TLX and the RSME scale for invested mental effort (Zijlstra, 1993).

4.2 Indicators for driver situation awareness

Since the driver behaviour researcher may be unfamiliar with some of these methods – at least more so than with the ones discussed so far - the details of the different methods are listed in Table 3.

Table 3: Commonly used driver situation awareness measures for different experimental environments

Measure	Field	Simulator	Laboratory
SAGAT		OK	OK
Post event probes	OK	OK	OK
Subjective ratings	OK	OK	OK
SART	OK	OK	OK
SA-SWORD			OK
Observer ratings	OK	OK	OK
Indirect/performance measures	OK	OK	OK
Physiological measures	OK	OK	OK

SAGAT

In the Situation Awareness Global Assessment Technique (SAGAT) a trial is suspended and all displays are blanked while the participant answers the relevant questions. Once the participant has responded to the questions the trial is continued. There are a number of beneficial aspects to this procedure. It has the benefit of asking the questions when the relevant information is still fresh for the participant. Probes can be made at various times during the trial. The questions can address a wide variety of aspects of the situation addressing all three levels of SA, i.e., perception, interpretation, and extrapolation. The one potential drawback is that the ongoing event must be stopped for the questions to be asked.

Post event probes

In this approach, subjects are probed for recall after the event. Although convenient to administer, there are a number of concerns with this method. First, if the test session is long, subjects may have forgotten the relevant information. It is also possible that responses will be affected by inferences that the subject has made in the intervening period between the event and when the response is collected. Second, if the nature of the task environment is highly dynamic (as it often is in driving), then it is unlikely that post event probes will capture this complexity.

Subjective ratings

In contrast with performance or probe-based measures, subjects in this approach are required to generate numerical or scale ratings based on their subjective evaluation of certain aspects of their situation awareness. While this approach is relatively easy to administer, it relies on the person's subjective evaluation of SA. A major drawback is that subjective ratings assume that the subject knows what SA is and can accurately assess its level. The ratings are also subject to influences from other factors such as performance.

SART

The Situation Awareness Rating Technique (SART) was devised for use in the aviation domain. It provides measurement of three basic dimensions of performance: demand (D), supply (S), and understanding (U) of the situation. Each of these factors is in turn broken down into contributing dimensions as follows:
Attention demand: complexity, variability, and instability
Attentional supply: arousal, concentration, division of attention, spare mental capacity
Understanding: information quality, information quantity, and familiarity

The major limitation of this approach is that situation awareness is confounded with workload in the use of measures of attentional demand and supply.

SA-SWORD

This approach uses a modified version of the Subjective Workload Dominance (SWORD) technique and was developed to obtain subjective ratings of SA of the information provided by displays. This method requires subjects to make comparative evaluations of the relative amount of SA provided by different display formats and consequently has limited applicability in assessing a driver's situation awareness of the larger driving environment.

Observer ratings

In this type of procedure, trained observers rate the degree of the driver's situation awareness. The Wiener Fahrprobe is a well-known case in point. The primary limitation of this type of technique is that it is assumed that the observer would have knowledge of the driver's SA and that this could be inferred from the driver's behaviour. Nevertheless, this approach may be among the more practicable that could be used (Carsten et al., 2005).

Indirect / performance measures for SA

It is possible to introduce changes in the environment and then observe drivers' responses to those changes. This approach has the benefit of not interrupting the ongoing task in that the manipulations and responses can be made online. These sort of indirect or performance measures can be assessed through measures of hazard detection, accuracy of detection, or latency to respond measures. It is, however, difficult to interpret non-responses when these types of manipulations are made as one is unsure whether the event was not detected or detected and a response not given.

Physiological measures of SA

A number of physiological procedures have been proposed to assess SA such as EEG and eye tracking. All of these are quite intrusive and require considerable instrumentation to carry out. The results can be difficult to relate to SA and to interpret.

Recommendation:

In Projects that have investigated the role of SA in driving, the HASTE Project in particular (Carsten et al., 2005), the results have demonstrated the applicability of observer rating methods, like the Wiener Fahrprobe. Other methods are not to be recommended because they are very hard to incorporate in a driver behaviour study, particularly in the field, since they will interfere with the performance of the study itself..

4.3 Indicators for driver mental models

SA is best regarded as a component of driver mental models in a wider sense. In order to get hold of mental models in the wider sense the methods of Table 4 are available.

Table 4: Commonly used methods for assessing drivers' mental models

Measure	Field	Simululator	Laboratory
Focus groups	OK	OK	OK
In-depth interviews	OK	OK	OK
Questionnaires	OK	OK	OK
Diaries	OK		

Focus groups

Focus Groups (FG) are a qualitative method of research for inspecting a system, at any phase of its development, focused on specific topics and in which a group discussion takes place.

A Focus Group is essentially a discussion made by prospective users of a system who are asked to talk (almost) freely about the system, its abilities, and its issues.

A professional moderator is often present, with the purpose of guiding the discussion onto a pre-designed schema, and maintaining it as transparently as possible. Users' opinions are tracked, so that it is possible to understand the needs of a user who is willing to use the system.

Focus Groups present the problem that users might be imprecise about their statements, and that there might be a difference between what users say and what and how they really do it.

In-depth interviews

In-Depth Interviews are done with one person at a time.

The purpose of interviews is to let interviewees freely talk about important topics, and the interviewer has the purpose to guide the interview, collect data and analyse it. Well-conducted interviews last long, and therefore are expensive. On the other hand, they provide individuals' perceptions, opinions, facts and forecasts, and their reactions to initial findings and potential solutions.

Questionnaires

Questionnaires are one of the most used tools to assess a system. Because of their generality of application, they are suitable to assess almost everything of a system, at any phase of its development.

Questionnaires have some drawbacks:
1) Subjects may tend to give personal interpretations of events, by introducing biases due to previous experience and knowledge; this may lead the test results to be misleading;
2) Questionnaires are to be done as soon as possible, if they deal with system particular topics, as subjects tend to forget events;
3) Subjects are often difficult to find, and might mostly have a technical background. This could bring misleading results, due to the incorrect mirroring of a pure statistical sample.

Diaries

Self-Reported Diaries are a technique in which users are requested to keep a diary during a period of time, recording personal information about their behaviour, when interacting with a system. The purpose is to assess the impact of the system in everyday life. Diaries are totally subjective, and therefore collected information must be verified in some other ways, due to possible biases and personal background. After data have been collected they must be aggregated in order to identify topics to refine and adjust.

Recommendation:

These methods have a more qualitative and naturalistic flavour than the ones discussed so far. For this reason, it is hard to make recommendations as to which one could be applied in what type of driver behaviour studies, and we will abstain from that.

4.4 Conclusion

Reasonable recommendations can be made for workload and distraction measures to be included in ITS studies. For Situation Awareness the most practicable method would be observer ratings. For driver mental models we have not been able to recommend specific procedures. Here more research is clearly needed (see Bellet et al., 2005).

5. Usability and acceptance measures

5.1 Usability

Usability is the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use. Evaluating an IVIS /ADAS according to its Usability therefore means to identify how user-friendly and easy-learning the system is, how much it influences the driving experience, and how it is liked and used in a satisfying way by drivers.

Methods to assess usability can be divided in methods with experts and methods with users. However, experts can assess systems only up to a certain point: due to their expertise, they do not reflect "average user" behaviour. Due to this fact, it is mandatory to use "normal people", those who are likely to use the system in-car, when driving.

A thing to note is that it doesn't make much sense to judge the appropriateness of usability methods for different experimental environments, as they are usually not applied in driver behaviour studies as we know them.

Table 5: Available methods for drivers' usability assessment

Measure
Brooke questionnaire (SUS)
Thinking aloud
Co-Discovery learning
User monitoring

The Brooke Questionnaire

This questionnaire, also known as System Usability Scale (SUS), was developed by John Brooke in 1986 (see Brooke, 1992), as part of the introduction of usability engineering to Digital's integrated office systems programme. Its objectives were to provide an easy test for subjects to complete (i.e. minimal number of questions), easy scoring, and to allow cross-product comparisons. It has been used extensively in evaluations of projects in Digital World (office systems, system management, technical tools and hardware systems) and has been estimated simple and reliable.

The SUS scale is generally used after a user has tried a system but before any debriefing or discussion takes place. Users should be asked to record their immediate response to each item, rather than thinking about items for a long time. All items must be checked. If users feel they cannot respond to a particular item, they should mark the centre point of the scale.

Thinking aloud

Thinking-Aloud is a self-explanatory technique: when a subject interacts with the system, he/she is asked to think aloud, so as to externalise every thought he/she has about the system. With this method, it is possible to understand which parts of the system are more difficult to reach, and what is the mental process the subject runs to find a way out to the task completion. This method allows capturing very personal and subjective impressions, hardly collectable in a different way, because they are told as they happen.

Thinking-Aloud is easily done with low costs, there is no extreme need of videotaping (notebooks work well), and it allows a great number of subjects to attend it. A drawback is the constraint that only one subject at a time can be observed, in order to watch correctly his/her behaviour, and to avoid influences by other people in the task completion as well as to avoid simple distraction.

Co-Discovery Learning

Co-Discovery Learning (CDL) uses a collaborative task completion procedure. Subjects, in pairs, test the system through accomplishing a pre-defined task, and help each other in doing that. A scenario is described, and the subjects are asked to perform a list of tasks pertinent to that specific scenario. During the task accomplishment, subjects are asked to externalise their thoughts.

User monitoring

User Monitoring is carried out by experts who look at users interacting with a system, either directly or by means of recorded media.

Observation by means of recorded media is useful to find more clues which would have gone unseen otherwise, while direct observation allows experts to focus on precise areas, at the price of sacrificing other topics. Traditional methods are less complete, but are fast to analyse. Conversely, recording data is more precise, but more time-consuming in the analysis phase.

Recommendation:
There is no sufficient available material to make a recommendation on these rather informal methods.

5.2 Acceptance

Acceptance is the measure according to which a user expresses the will to use a system and the level of appreciation resulted after use, in terms of safety improvement, reduction of vehicle operation costs, saving in travel time, improvement in driving comfort, HMI friendliness, etc.

Acceptance is a subjective measurement. It is not possible to assess it in objective terms, for several factors such as age, expertise, knowledge as well as personal needs and likes are strictly tied to evaluation results. Therefore, the measurement can be assessed through subjective means.

Table 6: Commonly used methods for drivers' acceptance assessment

Measure
Van der Laan scale
Willingness to pay
Importance ranking

Standardised Attitude Scale Calculus Algorithm

This method (developed by van der Laan et al., 1997) is based on a Semantic Differential Scale that measures people's reactions to stimulus world and concepts, and ratings are given on bipolar 7-point scales defined with opposite adjectives at each end. In particular, this algorithm tries to represent the acceptance of a system according to whether the system is:

Useful	Pleasant	Good
Nice	Effective	Likeable
Assisting	Desirable	Raising alertness

The items are, after subjects mark them, collapsed into two dimensions (Usability and Pleasantness), describing the system as lying in one of four quadrants.

Willingness to Pay / Use / Purchase

Willingness to Pay expresses the measure by which a user is prone to pay to have a particular system installed on-car. It is a very important measure for marketing purposes, to see whether or not the new system could be produced and sold with profit.

This measure is assessed with precise questions in questionnaires or Interviews. It can be influenced by several personal factors, such as age and expertise and / or localisation and climatic means - the geographical position and usual weather conditions -. A slightly different pattern in answers will probably be observed when systems are to be bought separately from the car, in comparison to when systems are already part of a new car.

Importance Ranking

Importance Ranking can be used to identify the way users perceive the relative importance of a set of pre-defined items. They are asked to rank a number of items according to a specific criterion. This can be useful when choosing among different possibilities to realise a system (different mock-ups to elaborate, different interfaces to be presented, different sets of icons, terms etc.).

Recommendation:
The Van Der Laan scales appear to be easily applicable, and easily interpretable, in driver behaviour studies of ITS effects.

5.3 Conclusion

A reasonable recommendation can be made for including the 'acceptance' aspect in driver behaviour studies. The case is less clear for 'usability', which is harder to incorporate in behavioural studies as they are currently done.

6. Getting to a risk (reduction?) estimate for IVIS and ADAS

A question to which research has devoted relatively little attention until now is how driver behaviour parameters ‘translate’ into safety effects. One approach, advocated by Brookhuis et al. (2003), is to develop pass-fail criteria for behavioural parameters, in terms of what risks are to be considered acceptable or not. The other approach is to derive continuous functions, which produce risk (reduction) estimates which are then handed down to policymakers for decisions (e.g., by comparing them to risks that are already accepted or rejected societally).

It should be noted that until today both approaches have only considered driving performance parameters. The question of what the relationship might be between, for example, workload and accident risk is almost completely unstudied, except in the very general sense that both high and low workload levels are more risky than an average one. Some pragmatic rules-of-thumb for estimating the trade-off effects on accident risk between behavior and driver state are, however, under development in the AIDE Project (Janssen et al., 2006; also Jamson et al., 2006, and Bekiaris et al., 2006). For example, one rule-of-thumb says that, if general driver alertness level is shown to drop from ‘excellent’ to ‘poor’, accident risk for that reason alone will increase by a factor of 2. Crude as they may be, this type of rule is useful at least until we will have derived more accurate ways of estimation.

6.1 The Brookhuis et al. (2003) criteria

The following tables (proposed by Brookhuis et al., 2003) represent the state of affairs. They contain proposals for describing when certain parameters should be considered to have turned into unsafe (i.e., unacceptable) values.

Table 7: Cut-off criteria for a number of driving performance measures; expressed as either an absolute value or a relative change from one condition to another (e.g., without and with system), whichever is reached first

	Absolute value	Relative change
time headway to lead vehicle	< 0.7s	– 0.3s
steering SD	> 1.5°	+ 0.5°
lateral deviation (SD) of vehicle	> 0.25m	+ 0.04m
Min time-to-line crossing (TLC) right lane	< 1.3s	– 0.3s
Min time-to-line crossing (TLC) left lane	< 1.7s	– 0.2s
median TLC (right lane)	< 3.1s	– 0.7s
median TLC (left lane)	< 4.0s	– 1.4s
vehicle speed	Limit + 10%	+ / - 20%

Table 8: Further refinement of (some of) the Brookhuis et al. (2003) cut-off criteria

Measure	Speed	Absolute change
Standard deviation of lateral position (SDLP)	>50km/h+	0.25
SD Steer	At 60 km/h >80 km/h - 120 km/h	1.7° 1.5°
Median TLC	60km/h 80km/h 100km/h 120km/h	6.0s 5.7s 5.0s 4.2s
15% TLC	60km/h 80km/h 100km/h 120km/h	3.8s 3.5s 3.1s 2.9s
Min TLC at different speeds		1.1s

6.2 Continuous behaviour-risk functions

Quantitative relations between certain behavioural parameters and ensuing accident risk have also been considered in the literature. They are shown in the following graphs for speed, speed variability, and headway, respectively.

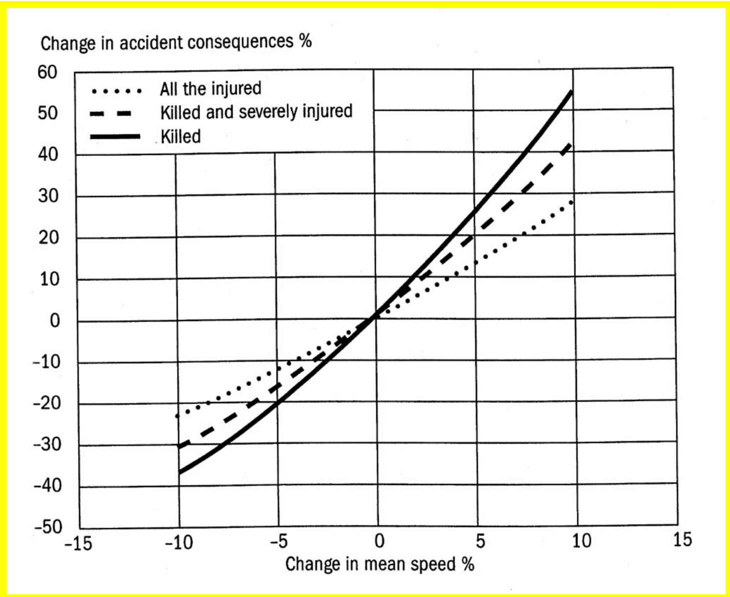

Fig. 1: Empirical function relating average driving speed (change in % relative to existing average) and its effect on accident risk (according to Nilsson, 1984).

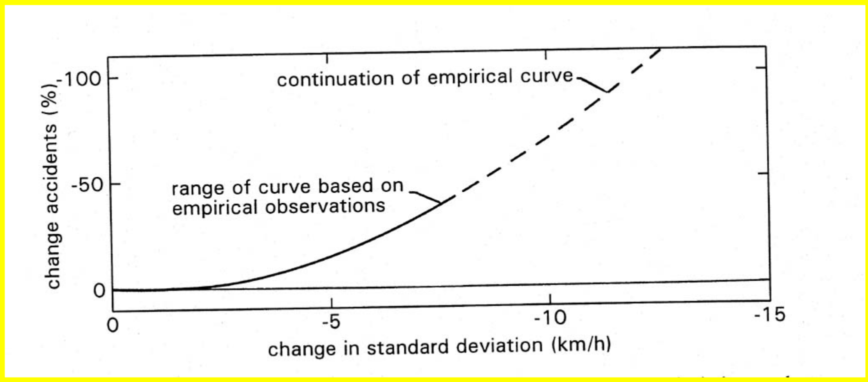

Fig. 2: Empirical function relating (change in) speed variability, relative to existing level, and its effect on accident risk (according to Salusjärvi, 1990).

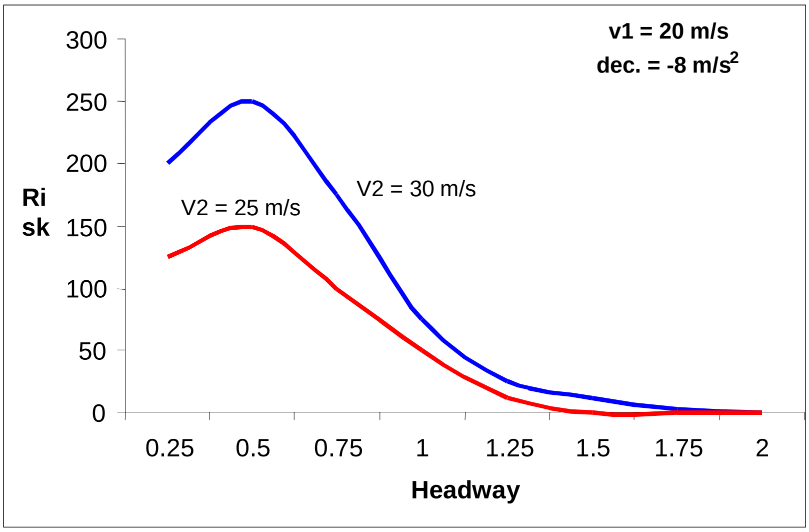

Fig. 3: Headway and relative accident risk when following a lead vehicle at different speeds (after Farber, 1993, 1994; and Janssen, 2000). Lead vehicle drives at 20 m/s and brakes suddenly at full braking power. Two cases are shown: following vehicle travels at 25, respectively 30 m/s. The case for following at 0.75 s with a speed difference of 25 m/s has been defined as a risk level of 100.

6.3 Conclusion

While knowledge is still far from complete there is some useful material available that can help us to extrapolate from behavioural data (that is, driving performance measures) to accident risk. It is recommended that this extrapolation becomes a standard part of behavioural studies on ITS effects.

7. General Conclusion

This Deliverable contains a proposal for a minimum methodology to assess driver behaviour in ITS design and evaluation studies. Its rationale is that a minimum set of parameters should be selected so as to cover all relevant dimensions describing that behaviour. We conclude that for some categories, like direct driver performance parameters, there are indeed firm recommendations to make. For other categories, however, there is a variety of methods available but less evidence to recommend using one rather than the other. In these cases, practicable recommendations have been made whenever possible. However, for a number of aspects there is presently insufficient evidence to do even that.

8. References

Bekiaris, E., Portouli, V., Papakostopoulos, V. & Janssen, W.H. (2006). Combining workload and behavioural effects into overall risk reduction estimate. AIDE Deliverable D 2.3.3.

Bellet, Th., Bailly, B., Mayenobe, P., Georgeon, O., Tattegrain-Veste, H., Martin, R., Mille, A. & Trassoudaine, L. (2005). Cognitive modelling and simulation of drivers mental models. In Proc. of International Workshop on Modelling Driver Behaviour in Automotive Environments. JRC, Ispra.

Brooke, J. (1992). SUS – A quick and dirty usablity scale. Redhatch Consulting Ltd., UK.

Brookhuis, K., de Waard, D. & Fairclough, S. (2003). Criteria for driver impairment. Ergonomics, 46, 433-445

Carsten, O., Merat, N., Janssen, W., Johansson, E., Fowkes, M. & Brookhuis, K. (2005). HASTE Final Report.

Farber, E. (1993). Using freeway traffic data to estimate the effectiveness of rear-end collision countermeasures. Washington, DC: Proc. Third Annual Meeting of the Intelligent Vehicle Society of America.

Farber, E. (1994). Using the Reamacs model to compare the effectiveness of alternative rear end collision warning algorithms. Paper No. 94 S3 O 03., XIVth International Technical Conference on Enhanced Safety of Vehicles, München.

Hanzlíková, I. (2004). Inventory of ethical laws which apply (at the national and EC levels) during behavioural experimental studies. HUMANIST Deliverable 2.1

Jamson, S., Batley, R., Janssen, W., Portouli, V., Papakostopoulos, V., Tapani, A., Lundgren, J., Huang, Y. & Hollnagel, E. (2006). Obtaining the functions describing the relations between behaviour and risk. AIDE Deliverable D 2.3.1.

Janssen, W.H. (1994). Seat-belt wearing and driving behaviour: an instrumented - vehicle study. Accident Analysis and Prevention, 26, 249-261.

Janssen, W.H. (2000). Functions relating driver behaviour and accident risk. Soesterberg: TNO Human Factors, Report TM-00-D004.

Janssen, W.H. (2006). Proposal for common methodologies for analysing driver behaviour . HUMANIST Deliverable 2.6.

Janssen, W.H., Brouwer, R.F.T. & Huang, Y. (2006). Risk trade-offs between driving behaviour and driver state. AIDE Deliverable D 2.3.2.

Johansson, E., O. Carsten, W. Janssen, S. Jamson, H. Jamson, J. Östlund, , R. Brouwer, S. Mouta, J. Harbluk, V. Anttila, H. Sandberg, J. Luoma. (2005). Validation of the HASTE protocol specification. HASTE Deliverable 3.

Macku, I. (2007). Report on the conditions of applications of ethical rules during jointly executed experimentations. HUMANIST Deliverable D 2.4.

Markkula, G. & Engström, J. (2006). A steering wheel reversal rate metric for assessing effects of visual and cognitive secondary task load. Proceedings of the ITS World Congress, London 2006.

Nilsson, G. (1984). Speeds, accident rates and personal injury consequences for different road types. Linköping: VTI Rept. 277.

Risser, R. & Brandstätter Ch. (1985). Die Wiener Fahrprobe. Freie Beobachtung, Kleine Fachbuchreihe des Kuratoriums für Verkehrssicherheit

Salusjärvi, M. (1990). Finland. In: G. Nilsson (Ed.), Speed and safety: research results from the Nordic Countries. Linköping: VTI

Van der Laan, J.D., Heino, A., & De Waard, D. (1997). A simple procedure for the assessment of acceptance of advanced transport telematics. Transportation Research - Part C: Emerging Technologies, 5, 1-10.

Zijlstra, F.R.H. (1993). Efficiency in work behavior. A design approach for modern tools. Ph.D. Thesis, Delft University of Technology.

9. List of abbreviations

CDL	Co-Discovery Learning
EEG	Electro Encephalogram
FG	Focus Groups
HMI	Human Machine Interaction
ITS	Intelligent Transport Systems
LANEX	Amount of time outside lane
NASA-TLX	Task load index developed within NASA
PDT	Peripheral Detection Task
PET	Post Encroachment Time
RSME	Rating Scale Mental Effort
RT	Response Time
SA	Situation Awareness
SAGAT	Situation Awareness Global Assessment Technique
SART	Situation Awareness Rating Technique
SA-SWORD	Situation Awareness -Subjective Workload Dominance
SD	Standard Deviation
SUS	System Usability Scale
TLC	Time-to-Line Crossing
TTC	Time to Collision