STATE-OF-THE-ART OF THE SNRA/JARI/BAST JOINT RESEARCH ON DRIVER WORKLOAD MEASUREMENT WITHIN THE FRAMEWORK OF IHRA-ITS

Gelau, Christhard
Bundesanstalt fuer Strassenwesen/Federal Highway Research Institute (BASt),
Jahn, Georg; Krems, Josef F.
Chemnitz University of Technology, Germany
Uno, Hiroshi
Japanese Automotive Research Institute (JARI), Japan
Kircher, Albert; Östlund, Joakim& Nilsson, Lena
VTI, Swedish National Road and Transport Research Institute, Sweden
Paper 242

ABSTRACT

The Swedish National Road Administration (SNRA), the Japanese Automobile Research Institute (JARI) and the Federal Highway Research Institute (BASt) are co-operating in the International Harmonized Research Activities on Intelligent Transportation Systems (IHRA-ITS). Under this umbrella a joint study was conducted. The overall objective of this study was to contribute to the definition and validation of a "battery of tools" which enables a prediction and an assessment of changes in driver workload due to the use of in-vehicle information systems (IVIS) while driving. In this sense "validation" means to produce empirical evidence from which it can be concluded that these methods reliably discriminate between IVIS which differ in terms of relevant features of the HMI-design. Additionally these methods should also be sensitive to the task demands imposed on the driver by the traffic situation and their interactions with HMI-design. To achieve these goals experimental validation studies (on-road and in the simulator) were performed in Sweden, Germany and Japan. As a common element these studies focused on the secondary task methodology as an approach to the study of driver workload. In a joint German-Swedish on-road study the Peripheral Detection Task (PDT) was assessed with respect to its sensitivity to the complexity of traffic situations and effects of different types of navigation systems. Results show that the PDT performance of both the German and the Swedish subjects reflects the task demands of the traffic situations better than those of the IVIS. However, alternative explanations are possible which will be examined by further analyses. Results of this study are supplemented by the Japanese study where informational demands induced by various traffic situations were analysed by using a simple arithmetic task as a secondary task. Results of this study show that relatively large task demands can be expected even from simple traffic situations.

INTRODUCTION

In-vehicle information systems (IVIS) are becoming a more and more common equipment of modern cars. Despite of their obvious benefits there are concerns that their use while driving may cause safety problems due to distraction and increased workload (e.g. Sprenger, 2000). Thereby it is assumed that it depends mainly on the user-friendly design of the Human-Machine-Interface (HMI) if the use of IVIS is compatible with the primary task of driving or if interferences have to be expected (e.g. Haller, 1999). For instance, in Europe recent efforts to develop and evaluate a catalogue of design goals for the in-vehicle HMI of IVIS, the so called "European Statement of Principles" (ESoP), reflect this approach. Nevertheless, the character of the ESoP is generic, i.e. defining no criteria to assess if the design goals have been achieved by a certain HMI solution (e.g. Gail, Nicklisch, Gelau et al., 2002). From all this it follows that there is a need for methods to evaluate the effects of IVIS on driver workload and behaviour in order to assess problems for traffic safety and to improve HMI design. Thus, both authorities and manufacturers have an interest in the development and standardization of tools and methods for the HMI evaluation of IVIS.

For these reasons the Swedish National Road Administration (SNRA), the Federal Highway Research Institute (BASt) and the Japanese Automobile Research Institute (JARI) started to co-operate under the umbrella of the International Harmonized Research Activities on Intelligent Transportation Systems (IHRA-ITS). Parallel studies were conducted at VTI (Sweden) and Chemnitz University of Technology (Germany) to contribute to the research on promising methods, with special focus on the Peripheral Detection Task (PDT, see next paragraph) which could become part of a standardised set of evaluation methods. Among the advantages of parallel studies are the outcome of a large data set and evidence on the reproducability of results at different sites which is of special importance for standardised methods. To ensure the advantages of common studies an important goal in both studies was to arrive at directly comparable data. The joint German-Swedish study is supplemented by Japanese research performed in the laboratory and in the simulator which also aimed at further exploring and improving measures of driver workload. In this paper we summarise the results and discuss the state-of-the-art of this ongoing joint research activity.

Measurement of driver workload and distraction

The assessment of distraction and workload effects of an IVIS requires the consideration of HMI characteristics as well as of situational factors and strategical use by drivers. Interindividual differences in drivers' skills and abilities also have to be considered. Three broad classes of safety-relevant distraction effects have been identified in the literature (Tijerina, 2001). These are:

- General withdrawal of attention,
- selective withdrawal of attention, and
- biomechanical interference.

General withdrawal of attention occurs, for example, if drivers move their eyes away from the road scene to the HMI of an IVIS. The resulting impairment of vehicle control and object and event detection depends on the frequency and duration of glances away from the road. It also depends on the direction of glances which varies according to the location of the in-vehicle display. The intentions and activities causing glances away from the road also contribute to overall distraction. For example, the driver may just confirm compliance with the current speed limit by a glance to the speedometer, but will be much more distracted when glancing to an IVIS trying to understand a cryptic e-mail message.

The last mentioned example of drivers' activities within the vehicle can also be used to characterise the second class of distraction termed *selective withdrawal of attention*, which is a result of "attention to thoughts" resulting in cognitive workload. Even hands-free use of a mobile phone may cause this kind of distraction effects (e.g. Nunes & Recartes, 2002). Vehicle control may remain unaffected, if eyes are on the road, but object and event detection suffer. The useful field of view narrows because of reduced and less guided visual scanning and because of a reduced ability to detect stimuli in the peripheral field of view ("tunnel vision", Williams, 1985).

The third class of distraction effects, *biomechanical interference*, includes body shifts out of the neutral seated position and hands off the steering wheel. This might occur because the driver manipulates objects with one or both hands or reaches for objects inside the car, e.g. the remote control of a route guidance system. Biomechanical interference can impede the fast and effective execution of manoeuvres.

The safety implications of distraction effects grow with the demands traffic situations put on the driver. Small lapses of vehicle control may be less critical on a free highway at moderate speed than in dense traffic, and events and objects that have to be noticed instantly are more probable in more complex traffic situations like turning left on inner city four-way crossings.
Drivers are aware of the demands of the primary task of driving and can – within limits - strategically manage their workload e.g. by adapting their glance behaviour or by lowering speed. But the potential of strategic workload management and time-sharing is reduced if timing is not under the control of the driver or if in-vehicle tasks take a long time. Furthermore workload management can break down, e.g. caused by emotional involvement during a phone conversation or because of high cognitive workload. Workload management itself requires cognitive resources, that may be unavailable. Difficult to prevent, workload management also can be performed unsafely on purpose in response to time pressure.

Strategic management of workload bears relevance to the evaluation of safety effects of in-vehicle information and communication systems. Specifically it would be invalid to assume that drivers usually operate at their limits and sacrifice safety when integrating driving and interacting with in-vehicle information and communication systems. On the other hand events that occur unexpectedly can instantly demand full attention. Then the outcome may be the worse, the more workload was present in addition to the workload caused by driving, even if overall workload was well below limits before the critical event occurred.

To summarise, the assessment of the in-vehicle HMI of IVIS should consider the frequency and duration of general withdrawal of attention, the amount of selective withdrawal of attention over time and biomechanical interference which might differ between traffic situations and between subgroups of the driver population. The potential of strategical management of workload should be kept in mind, including the probability of failures in workload management.

Thus, for HMI assessment the focus should be on visual attention, overall cognitive workload and overall response execution workload. Visual attention is of special importance for safe driving, for vehicle control as well as for event detection. Overall cognitive workload may impair visual object and event detection and degrade response selection. Response execution may suffer from overall response execution workload. Among methods for workload measurement those aiming at visual attention and overall workload are of special interest for HMI assessment.

General withdrawal of visual attention can be quantified by observing and recording drivers' viewing behaviour. Although this can be very useful the disadvantages are the need for either expensive eye tracking equipment or many working hours of video coding. The presence of selective withdrawal of attention is less obvious from overt behaviour and better diagnosed by workload measurement. Techniques for workload measurement often are subdivided in primary-task measures, secondary-task measures, physiological measures, and subjective rating techniques (e.g. O'Donnell & Eggemeier, 1986; Wickens & Hollands, 2000).

A primary-task measure of workload during driving could be, for example, counting lane exceedences. Obviously levels of workload that do not impair driving cannot be differentiated by primary-task measures of driving. Secondary-task measures of spare capacity ideally do just this. If the secondary task is well suited to the primary task it is assumed that secondary-task performance is inversely proportional to primary-task performance. If the driver is instructed to allocate enough resources to the primary-task to conserve primary-task performance, the secondary-task is a "subsidiary task" and secondary-task performance reflects changes in primary-task resource demand (c.f. O'Donnell & Eggemeier, 1986; Wickens & Hollands, 2000). A possible drawback of secondary-task techniques is the occurrence of interference with the primary task, i.e. they may be obtrusive. Unobtrusiveness is a main advantage of physiological workload measures. But physiological data require interpretation to infer workload and spare capacities, which are more directly captured by secondary-task techniques.

The secondary task evaluated in the German-Swedish study is the so called "Peripheral Detection Task" (PDT) that was used in simulator studies and in driving studies in recent years to assess changes in workload during driving, and to assess workload and distraction caused by in-vehicle systems. The PDT is considered as a candidate for a standard set of methods or tools for HMI assessment. The PDT requires a simple manual responses to stimuli usually presented left to the drivers' normal line of sight. Stimuli are visible for 1 to 2 sec. and are presented with intervals of a few seconds, e.g. 3 sec. to 5 sec. or 3 sec. to 6 sec. Van Winsum, Martens, and Herland (1999) at TNO developed the task mainly based on studies of Miura (1986) and Williams (1985; 1995). Miura (1986) found that response times to spots of light presented at different horizontal eccentricities on the windscreen during driving increased with traffic density and by this reflected demands of the driving task. Williams (1985; 1995) showed that with increasing foveal load the accuracy of responses to stimuli presented peripherally decreased.

A subjective measure which has been widely used for the HMI assessment of IVIS in recent years and which will also be considered (in a simplified version) by the German-Swedish study is the "NASA Task Load Index" (NASA -TLX). The NASA –TLX was developed by Hart and Staveland (1988) is a subjective workload assessment tool. NASA -TLX requires users to perform subjective workload assessments on operator(s) working with various human-machine systems. It is a multi-dimensional rating procedure that derives an overall workload score based on a weighted average of ratings on six subscales. These subscales include *Mental Demands, Physical Demands, Temporal Demands, Own Performance, Effort,* and *Frustration*. Three dimensions relate to the demands imposed on the subject (Mental, Physical. and Temporal Demands) and three to the interaction of a subject with the task (Effort, Frustration and Own Performance). Besides the six scales, an overall weighted measure of task load can also calculated on the basis of the scales. The NASA-TLX has been developed to assess workload in various human-machine environments such as aircraft cockpits; command, control, and communication (C3) workstations; supervisory and process control environments, simulations, and laboratory tests. This method has been tested in a variety of experimental tasks that range from simulated flight to supervisory control simulations and laboratory tasks (e.g., the Sternberg memory task, choice reaction time, critical instability tracking, compensatory tracking, mental arithmetic, mental rotation, target acquisition, grammatical reasoning, etc.). The derived workload scores have been found to have substantially less between-rater variability than unidimensional workload ratings, and the subscales provide diagnostic information about the sources of load.

The German-Swedish study: Evaluating the PDT and the NASA-TLX as tools for HMI assessment

As already mentioned it was the main objective of the German-Swedish study to contribute to the definition and validation of a standardised "battery of tools" for the HMI assessment of IVIS. In order to achieve this goal two field studies were designed and conducted in Germany and Sweden. Both studies replicated each other with respect to a number of "common elements" which are crucial in terms of comparability of results (i.e. definition and length of test routes, subject samples, experimental design, systems used, workload measures). Apart from these "common elements" both studies also had their "specific elements" (e.g. laboratory experiments with the occlusion technique in Germany) will be reported elsewhere. In the present paper we focus on the field study as a major "common element" and results on the PDT and the NASA-TLX as methods evaluated in both the Swedish and the German study.

Methods

Participants

For both, the German and the Swedish study, taxi drivers were recruited as subjects. This was done for safety reasons, i.e. taxi drivers are experienced drivers who are used to driving with IVIS. Furthermore, this group is homogeneous with regard to knowledge about the roadmap of the respective city.

German study: In this study 49 subjects participated. 48 participants were aged between 26 and 55, one was aged 65, 47 were male, 2 were female. They held their driving license for at least 9 years and reported at least 130.000 km driving experience. Subjects were paid 50 for their participation.

Swedish study: 41 participants, 33 men and eight women, were selected and 40 completed the experiment. They were aged from 21 to 55 years and mileage of more than 15000 km in the last year. All participants were Swedes or fluent in the Swedish language. After completing the experiment data from all the participants but one were usable for analyses; in this case a vehicle problem occurred.

Procedure

For both the German and the Swedish study the *test routes* were defined according to a taxonomy of traffic situations suggested by Fastenmeier (1995). Traffic situations are classified with regard to the demands they put on the driver in terms of information processing and vehicle handling. In selecting the route the following descriptions of traffic situations were used to compose experimental sections with differing demands:

- *High demands on information processing and high demands on vehicle handling (HH):* Typical examples of this group of situations are “driving within city centres”, complex intersections with road signs where the driver has to give right of way.
- *Low demands on information processing and low demands on vehicle handling (LL):* Low demands result from all those situations in urban and rural areas and on motorways where “free driving”, i.e. without interactions with other traffic participants is possible.

Participants in each study drove the same route in the city of Chemnitz (260.000 inhabitants) and Linköping (130.000 inhabitants) respectively. Routes were selected to resemble each other as much as possible. In both studies the test routes covered each two experimental sections with high demands (HH1 and HH2) and low demands (LL1 and LL2). The length of the complete route in Chemnitz was 11.2 km, the part that contained the experimental sections was 8.2 km long. In Linköping the experimental part of the route had a length of 8.6 km.

In both studies participants performed their test drives with one of two navigation systems which mainly differed in features of the HMI which were expected to have an impact on task load. More precisely, the VDO Dayton MS 4200 (“small”) and the VDO Dayton MS 5000 (“large”) were used which differ mainly in display size, display organization and functionality (see Table 1). A main reason for choosing the VDO Dayton systems was that they provided audible information in Swedish and German and that CD-ROMs were available for Sweden as well as for Germany. A further requirement was a remote control. Both navigation systems provided verbal and symbolic guidance. The MS 4200 was mounted in the radio slot and has a small monochromatic display that shows arrow symbols, street names and distances for route guidance. The MS 5000 has a larger colour display that was mounted on a flexible holding device. The displays of both systems were located approximately 40 cm to the right measured from the centre of the steering wheel to the display centre. The large display was located vertically 5 cm below the centre of the speedometer, the small display was located 9 cm further down. The eccentricity from the forward line of sight was around 30° for both displays.

Settings for the MS 5000 were chosen so that during route guidance the symbolic and distance information

Table 1.
Main differences between the navigation systems

	VDO Dayton MS 4200	VDO Dayton MS 5000
Display	“small” 6.7 cm diagonal (5.9 cm x 3.1 cm) monochrome	“large” 14.6 cm diagonal (12.7 cm x 7.2 cm) 256 colours
Route guidance	Basic arrows No information about the actual road name Simple sketch of the next intersection	Different pointer forms Actual road name is shown Well defined diagram of the next intersection

was comparable to that displayed by the MS 4200 (no map was shown). In addition to the information displayed by the MS 4200 the MS 5000 displayed actual street names and more detailed diagrams of the next intersection. Both navigation systems were programmed with 5 destinations in advance. The destinations were changed during 5 short stops by the experimenter in the back seat using a remote control.

Workload measures

Peripheral Detection Task (PDT): The PDT used in both studies was provided by VOLVO. It required responses to LED signals projected in the left part of the windscreen. It consists of a main unit that controls signal presentation, a LED board with 6 red high-intensity LEDs arranged in two rows and a pushbutton to be attached to the left index finger. A mounting device was constructed for the LED board that allowed to adjust the board relative to the windscreen. The LEDs were shielded from direct view of the driver by black cardboard.

The LED board was mounted on the left side below the windscreen. LEDs projected in the area recommended by van Winsum et al. (1999): at a horizontal angle of 11° to 23° to the left of the line of sight of the driver and at a vertical angle between 2° to 4° above the horizon. The location of the signal varied randomly. The signal rate has been adjusted so that the interval between two presentations was 3 to 5 sec. A LED was on for maximally 2 sec. Within 2 sec. it went off as soon as the driver made a response. The LED projection area and the presentation parameters were identical for the German and the Swedish study. Participants responded with the pushbutton on their index fingers either by pushing with the thumb or by pressing the pushbutton against the steering wheel. The responses were collected on a PC in the back of the car.

NASA –TLX: In both the German and the Swedish study a simplified version of the NASA-TLX was administered which does not require the weighting procedure originally proposed by Hart and Staveland (1988). This was done in order to make the filling-in and scoring procedure faster and easier for the subjects. As there is evidence from the literature that the correlation between the original and simplified version is about r=0,95 (Byers, Bittner & Hill, 1989) we concluded that the more complicated scaling procedure virtually does not provide substantially more information. The modified form of the NASA-TLX in German and Swedish language was used to evaluate the subjective workload of the participants during driving. The participants filled in the form after the second complex route section (HH2), and after the second simple section (LL2).

Results

PDT

German study: On average 60 PDT signals were presented to each of the 49 participants within section HH1 (stops excluded), 73 PDT signals were presented within section HH2, 44 PDT signals within section LL1 and 23 PDT signals within section LL2. The mean hit rates and response times are presented in Table 2. Few signals were missed on the LL sections, the mean hit rates for both systems on these sections are around 93.5%. Detection performance on the HH sections was lower than on LL sections as expected with mean hit rates on the HH sections varying from 85.6% to 88.7%. As can also be seen from Table 2, differences between HMI types were negligible, independent from the demands of the traffic situations (HH vs. LL).

Responses with latencies above 2000 ms were classified as misses. Another criterion would not have changed the results considerably because only a small percentage of all responses were late hits classified as misses. 3.9% of responses on the HH sections and 1.9% of responses on LL sections were late hits. There was no difference between system groups in frequency of late hits. Also no difference between system groups was present for false alarm frequencies. 3.8% of all responses on HH sections and 2.4% of all responses on LL sections were false alarms.

Table 2.
Mean hit rates [%] with standard deviations and mean response times [ms] with standard deviations by display (large/small) on route sections with high demands (HH collapsing HH1 and HH2) and low demands (LL collapsing LL1 and LL2) in the German study

	Large Display		Small Display	
	M	SD	M	SD
Hit rate [%]				
HH	85.6	7.7	88.7	6.8
LL	93.4	3.3	93.6	4.2
RT [ms]				
HH	694	125	663	113
LL	565	125	561	89

Table 3
Mean hit rates [%] with standard deviations and mean response times [ms] with standard deviations by display (large/small) on route sections with high demands (HH collapsing HH1 and HH2) and low demands (LL collapsing LL1 and LL2) in the Swedish study

	Large Display		Small Display	
	M	SD	M	SD
Hit rate [%]				
HH	68.8	22.2	67.0	17.6
LL	82.9	16.7	85.0	10.6
RT [ms]				
HH	846	155	865	124
LL	743	185	714	87

Swedish study: Corresponding results on the PDT from the Swedish study are presented in Table 3. As can be seen from the table results from both studies correspond fairly nice with each other. The highest hit rates and shortest response were observed while driving during the HH route sections. Differences between the different types of HMI (large vs. small display) were negligible.

NASA-TLX

German study: Subjective workload ratings on the subscales of the NASA-TLX were collected after section HH2 and after section LL2. Participants' ratings on the analogue scales were transformed to values between 0 and 100. The mean ratings on the subscales after HH2 and after LL2 are presented in Figure 1 separately for participants who used the large display and those who used the small display.

Means of ratings on all subscales given after LL2 are around 20 for both types of display. After HH2 the ratings on the subscales Physical Demands, Own Performance, Effort, and Frustration are also around 20 for both display groups, only the ratings of Mental Demands and Temporal Demands are in both groups above those after LL2. This difference of approximately 8 units in the ratings of mental demands and temporal demands between HH2 and LL2 is statistically significant as confirmed by ANOVAs on single subscales.

The within subjects factor "route section" (HH2/LL2) and the between subjects factor "display" (large/small) have been included in ANOVAs of ratings on single subscales. The ANOVA of mental demands ratings confirmed the main effect of route section, $F(1, 46) = 14.0$, $p < .001$; $f = .55$, and the same way the ANOVA of temporal demands ratings yielded a statistically significant main effect of route section, $F(1, 46) = 12.1$, $p < .01$; $f = .51$. No other effect reached statistical significance in these and the ANOVAs of ratings on the remaining 4 subscales.

Swedish study: Items on all scales of the NASA-TLX showed a tendency towards higher workload in the HH sections. Differences between the levels were rather small and none of them reached the level of

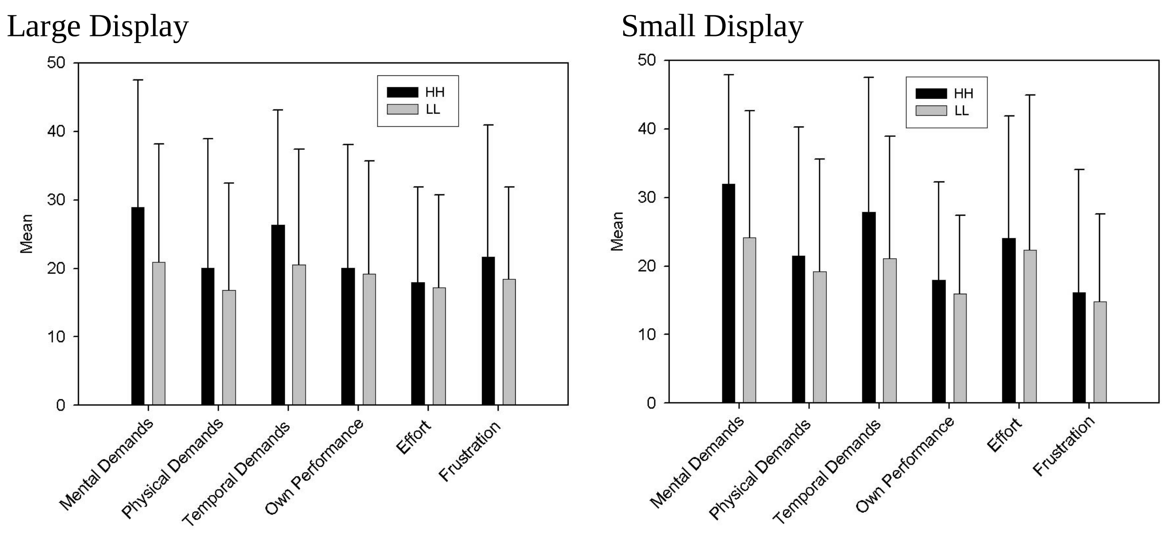

Figure 1. Mean ratings on the NASA-TLX subscales Mental Demands, Physical Demands, Temporal Demands, Own Performance, Effort, and Frustration with standard deviations (error bars) by display (large/small) after route section HH2 with high demands and route section LL2 with low demands. The left diagram shows mean ratings by the large system group, the right diagram shows mean ratings by the small system group.

statistical significance. A closer examination of individual ratings revealed large variations between participants. This indicates that the workload imposed by the different road sections was perceived very individually, and that other factors like e.g. traffic volume, personal characteristics etc., besides the demands of the road sections might have influenced the ratings. Effects of the HMI (display size) were weak as well. But at least a tendency was observed that the small display caused larger workload. More detailed information on the NASA-TLX results of the Swedish study can be found in Kircher, Östlund, Patten and Nilsson (in press).

Discussion

PDT hit rates and response times indicated a higher workload on more demanding route sections. Within the present design, effects of route demands and effects of route guidance cannot be separated, but the impairment is comparable to the documented effects of similar route demands on detection tasks (Verwey, 2000). Thus, the route effect probably reflects mainly demands of the driving task.

Nonetheless in short intervals around route guidance messages there might have been effects on PDT performance. The expected sensitivity of the PDT to short peaks in workload will be true only if short intervals are analysed or if those peaks are frequent relative to the length of the analysed interval. It is planned to have a closer look at those intervals in following analyses.

The participants in the present study were taxi drivers highly familiar with the local area. Because the route guidance systems visually indicated the street name into which the next turn would lead, the taxi drivers could make use of their knowledge and their workload therefore may have been lower than that drivers would experience who are less familiar with the city. But novelty of the systems presumably has raised workload of participants relative to usual users.

Route demands in the German and the Swedish study showed a stronger effect than in the study by Olsson and Burns (2000) who compared driving on a motorway and on country roads and did not find a difference in PDT performance. The highly demanding route sections in the present study included turns and traffic situations that may impair PDT performance through workload and through eye and head movements. In the low-demand sections turns and demanding situations were nearly absent. Obviously sections like these should be selected as baseline if effects of in-vehicle systems are to be studied. This points to a weakness of the PDT method as already mentioned: Some systems, e.g. route guidance systems, usually are in use on route sections that are highly demanding like in this study and, thus, should be assessed on such routes. Because of variance due to traffic situations PDT performance then is less sensitive to system effects.

NASA-TLX ratings indicated a low level of overall workload. And there were no significant effects of display size on NASA-TLX ratings, too. There may have been no differences in workload experienced by participants in the two groups (large vs. small display). Less probably, but also possible with the high variance between participants' ratings existing differences might have remained blurred. If participants had used both systems and would have been able to directly compare the systems, reports of differences would have been more likely. The high variance of subjective ratings is not unusual and yields them useful mainly with large samples.

The Japanese study: Assessment of driver workload using an auditory arithmetic secondary task

Although research revealed the relative demands of various driving situations (e.g. Harms, 1991), these demands were rarely described in quantitative terms or relatively during driving or operating IVIS. JARI's part of the joint IHRA-ITS project proposes a way to estimate spare capacity (bits/sec.) based on performances for auditory-presented arithmetic task, and try to assess the task load imposed by the demands of traffic situations by a visual detection task set in a driving simulator.

Procedures

It is generally accepted that, in simple situations, the reaction time in choice reaction task is delayed as a logarithmic function of the number of alternatives presented (e.g. Wickens & Hollands, 2002). Research has confirmed that choice reaction times of a subject for one to ten alternatives are distributed as a logarithmic function. The logarithms of the alternative numbers, where the base is two, agree with the theoretical units of information amount (bits) if the occurrences of each alternative are equal. Accordingly, the reaction times regress linearly with the information units. The slope of the linear regression line can be considered as the time required to process a 1-bit unit of information; i.e., the greater the incline, the longer the processing time per information unit. For the purpose of this study, the reciprocal number of the reaction time incline for information units was defined as a measure of the mental capacity, which is the amount of information processed per second (bits/sec.). Here, the bits/sec. value obtained in the choice reaction time task alone is considered to be the total capacity of the respective subject. The value obtained in the simultaneous performances of another task (e.g. arithmetic task) and choice reaction time task is expected to change according to the performance of another task.

Based on pilot research it was suggested that bits/sec. value can be estimated by the reaction time of the arithmetic task (adding digits from 0 to 15). That is, if the arithmetic task is used as a secondary task and its reaction times are measured while driving, information processing demands of traffic situations can be estimated using the reaction time of arithmetic task. Thus, subjects were required to perform the arithmetic task during simulated driving. The arithmetic task has some advantages to be used as a secondary task: It is relatively easy to carry out, and is considered to be hard to interfere on the driving task performance.

Traffic situations were provided using a driving simulator with motion-base, and information processing demands were estimated based on the measured reaction times of arithmetic task which was performed during driving.

A straight shape, two-lane (3.5m width for each lane) rural road of 2 km length was set on the driving simulator. Three driving situations were provided; i.e., straight, obstacle A, and obstacle B. In the straight condition, there were no obstacles on the road, and drivers were merely required to drive straight with keeping at the speed 60 km/h. In obstacle conditions, number of 10 obstacles were placed at 167 m interval for "obstacle A", and 20 obstacles were placed at 84 m interval for "obstacle B". The drivers were required to keep the speed of 60 km/h, so they had to negotiate the obstacles each 20 sec. (in obstacle A condition) or 10 sec. (in obstacle B condition).

In order to reproduce at least partially the aspects of using in-vehicle information devices, a peripheral visual detection task was also added to the straight condition. Red LEDs were installed in the simulator cabin at intervals of 10 degrees over an area of 85 degrees on left side to 65 degrees on right side of the driver's seat, and one of 14 LEDs was randomly lit up at interval of 5 to 11seconds. The visual detection task required driver to press a horn switch if a LED lit up.

Subjects drove under four conditions (straight, straight + visual detection, obstacle A, and obstacle B) while simultaneously carrying out the arithmetic task. Before driving simulator experiment, baseline of mental capacity for each subject was measured in not-driving (static) situation. Subjects were 13 male drivers who have been participated in the above mentioned experiment that estimated the mental capacities.

RESULTS

Figure 2 shows the average values of processing demands estimated by subtracting of bits/sec. values in each driving condition from baseline value. Estimated processing demands varied corresponding to the driving conditions. Largest demand was observed in obstacle B (5.4 bits/sec.) condition, and smallest one was straight condition (4.0 bits/sec.). That is, difference among driving conditions in this experimental setting was estimated as 1.4 bits/sec. This difference was relatively small than the demand in straight condition (4.0bits/sec.) that required drivers' simple driving task. This result reveals that driving itself creates large processing demands even if it is done under relatively simple driving conditions.

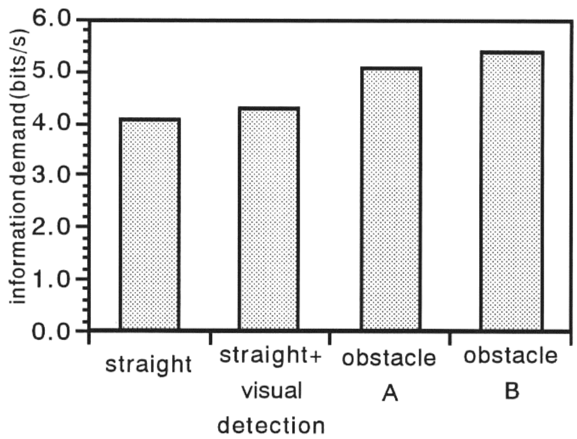

Figure 2. Processing demands by driving conditions in the Japanese simulator study

On the other hand, there were few increases in processing demand caused from the visual detection task. Simple detection of visual stimuli that can be presented on in-vehicle information devices is considered to induce relatively small information processing demands in comparison with that induced from driving tasks.

GENERAL DISCUSSION

The sensitivity of the PDT as suggested by van Winsum et al. (1999) to demands of the driving task has been demonstrated in the German-Swedish field studies. The workload effects of route guidance systems turned out to be weaker than the effects of the demands of traffic situations. However, further analyses will have to consider the possibility that this result can be explained by effects of confounding the temporal length of measurement intervals.

The disadvantage of blurred PDT sensitivity to effects of in-vehicle systems in the presence of demanding traffic situations may be difficult to avoid in field evaluations of certain IVIS, especially route guidance systems. In the German-Swedish study the NASA-TLX has been proven sensitive to driving demands, but may be useful mainly with large samples and within-subjects designs because of high variance between participants' ratings. Further analyses will include PDT performance on interesting shorter intervals and glance behaviour to uncover possible effects of system display size.

The project performed by JARI proposed a possible procedure to estimate information processing demand as bits/sec. value using an arithmetic task performance, and the demands were estimated for some of driving situation set in a driving simulator. The results show that driving itself has a concrete information processing demand if the driving situation is simple. It was also suggested that the simple visual detection task which was designed to simulate HMI performance on an IVIS induced only a relatively small demand compared with the demands induced by the primary driving task. This finding confirms the results and conclusions from the German-Swedish study.

REFERENCES

Byers, J.C., Bittner, A.C. & Hill, S.G. (1989). Traditional and raw task load index (TLX) correlations: Are paired comparisons necessary? In A. Mital (Ed.), *Advances in industrial ergonomics and safety* (pp. 481-485). London: Taylor & Francis.

Fastenmeier, W. (1995). Die Verkehrssituation als Analyseeinheit im Verkehrssystem. In W. Fastenmeier (Ed.), *Autofahrer und Verkehrssituation* (pp. 27-78) Köln: Verlag TÜV Rheinland.

Gail, J., Nicklisch, F., Gelau, C., Friedel, B., Bolte, F. & Sievert, W. (2002). Bestandsaufnahme der Umsetzung der EU-Empfehlung: "Europäischer Grundsatzkatalog zur Mensch-Maschine-Schnittstelle von Informations- und Kommunikationssystemen" in der Bundesrepublik Deutschland durch die Bundesanstalt für Straßenwesen. *Zeitschrift für Verkehrssicherheit,* 48, 113-116.

Haller, R. (1999). HMI - New Technologies and Safety Perspectives. *Proceedings of the 10th International Conference "Traffic Safety on Two Continents".* Malmö, September 20-22, 1999.

Harms, L. (1991). Variation in drivers' cognitive load: Effects of driving through village area and rural junction. *Ergonomics, 34*, 151-160.

Hart, S.G. & Staveland, L.E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P.A. Hancock & N. Meshkati (Eds.), *Human mental workload* (pp. 139-183). Amsterdam: Elsevier.

Kircher, A., Östlund, J., Patten, C. & Nilsson, L. (in press). PDT as workload measure in driving – the Swedish experience. Submitted at Transportation Research Part F.

Miura, T. (1986). Coping with situational demands: A study of eye movements and peripheral vision performance. In A.G. Gale, I.D. Brown, C.M. Haslegrave, P.Smith, & S. Taylor (Eds.), *Vision in Vehicles II* (pp. 205-216). Amsterdam: Elsevier.

Nunes, L. & Recarte, M.A. (2002). Cognitive demands of hands-free-phone conversation while driving. Transportation Research Part F 5, 133-144.

O'Donnell, R.D. & Eggemeier, F.T. (1986). Workload assessment methodology. In K. Boff, L. Kaufmann & J. Thomas (Eds.), *Handbook of perception and human performance*. New York: Wiley.

Sprenger, A. (1999). Ablenkung durch MMI-Lösungen: Der Blick zu Instrumenten im Fahrzeug als Unfallursache. In F. Meyer-Gramcko (Hrsg.), *Verkehrspsychologie auf neuen Wegen: Herausforderungen von Strasse, Wasser, Luft und Schiene, 37. BDP-Kongreß für Verkehrspsychologie* (S. 185-191), Bd. 1. Bonn: Deutscher Psychologen Verlag.

Tijerina, L. (2001). *Issues in the evaluation of driver distraction associated with in-vehicle information and telecommunication systems.* East Liberty, O.H.: Transportation research Center.

Verwey, W. (2000). On-line driver workload estimation: Effects of road situation and age on secondary task measures. *Ergonomics, 43,* 187-209.

Wickens, C.D. & Hollands, J.G. (2000). *Engineering psychology and human performance* (3^{rd} ed.). Upper Saddle River, N.J.: Prentice Hall.

Williams, L.J. (1985). Tunnel vision induced by foveal load manipulation. *Human Factors, 27,* 221-227.

Williams, L.J. (1986). Peripheral target recognition and visual field narrowing in aviators and nonaviators. *The International Journal of Aviation Psychology, 5,* 215-232.

Winsum, W. van, Martens, M., Herland, L., (1999). *The effects of speech versus tactile driver support messages on workload, driver behaviour and user acceptance* (Report No. TM-99-C043). Soesterberg, The Netherlands: TNO Human Factors.