Methods for assessment of cognitive workload in driving tasks

Kristina Stojmenova and Jaka Sodnik
University of Ljubljana, Faculty of Electrical Engineering, Ljubljana, Slovenia
kristina.stojmenova@fe.uni-lj.si
jaka.sodnik@fe.uni-lj.si

Abstract— **In this paper, we explain the concept of mental or cognitive workload and review the most common methods and procedures for its assessment. We focus primarily on driving tasks and interaction with various in-vehicle devices and systems. Safety is one of the most important human needs and consequentially also the primary concern of the automotive industry when introducing new in-vehicle information systems (IVIS), global positioning systems (GPS), interactive displays, etc. Since the human brain and its resources are limited, the primary task of driving can be seriously challenged when secondary tasks are performed simultaneously. Several different methods have been proposed for direct and indirect measurement of the driver's cognitive workload and for the detection of its potential overload. We report briefly also on two user studies performed in our driving simulator, which illustrate the importance of correctly assessing cognitive workload in the process of evaluating new in-vehicle user interfaces.**

I. INTRODUCTION

In all developed countries the majority of adult population owns a driving licence and participates in traffic on daily basis. With the increased number of drivers and vehicles, attention to driving and everything that influences it, has increased significantly. Although the safety of cars increases every year, they are also equipped with variety of electronic devices aiming to assist the driver in the driving process, enabling navigation, communication and entertainment services. These devices can distract the driver, negatively affect his or her responsiveness and cause significant amount of unwanted workload.

Workload is generally defined as the amount of work an individual has to do [1]. The term "workload" covers a broad spectrum of human activities and corresponding tasks while the term "mental workload" focuses only on demands imposed on the human's limited mental resources [2][3]. In both cases there is always a difference between the individual's subjective perception of workload and the actual amount of work. The so called overload of workload arises when the amount of human activity and requests for various physical and mental resources surpasses the amount of the available processing resources (of human brain or more precisely of the correspondent part of the brain).

Attention on the other hand, is defined as concentration of awareness to a specific source of information, whereas distraction is the diversion of attention from one source of information to one or several other sources [4]. Humans have multiple but limited amounts of attention and processing resources available at any given time [5]. Different tasks can use different attention resources or share them. If several simultaneously performed tasks rely on the same source they usually interfere with each other and seriously compete for that resource (e.g. two separate visual tasks) [6]. In that case the brain cannot process all the information presented and the performance is significantly affected. In a lot of real life situations this can be hazardous especially if the overlooked information regards safety, health or security issues.

In the past, a mental workload and measurement of mental over-workload were investigated and considered primarily in relation to design and operation of aircrafts and onboard systems in aircrafts. The mental workload of pilots can sometimes be critically increased due to variety of functionalities and features of these onboard systems and a variety of tasks each pilot has to preform simultaneously. Similarly, the available functionalities of In-Vehicle Information Systems (IVIS) increase rapidly, expecting a driver to perform high number of tasks simultaneously. However, a human brain can process only a limited amount of information. When the amount of mental workload surpasses the amount of mental capability, the surpassed amount of information is missed. Unfortunately the brain, more or less, randomly selects the information that is going to be ignored without letting a human consciousness to choose what is more and what is less important. While in many situations this may not be so important, for drivers of vehicles it can be dangerous and critical.

Mental workload can be measured directly by measuring the cerebral activity of human brain and indirectly with a number of different experimental methods. In this paper we review several methods for measuring mental workload of drivers while operating a vehicle and performing different secondary tasks (e.g. using on-board computer or infotainment system in a vehicle). The reviewed methods are divided into four major categories:

- direct psychophysiological measurements,
- measurement of ocular activity,
- methods based on measurement of response time and
- subjective measurements based on questionnaires for self-evaluation.

We also briefly report on two examples of cognitive workload measurements in a driving simulator revealing its importance in evaluation of novel in-vehicle systems and display technologies.

II. PSYCHOPHYSIOLOGICAL MEASUREMENTS

Due to the multifaceted nature of the complex mental demands in in-vehicle interaction multiple measures are required. One way of measuring mental workload directly is collecting the time-varying spatial potential distribution over the scalp produced by the cortical brain activity. These measurements can be performed with the electroencephalogram (EEG) or its magnetic counterpart magnetoencephalography (MEG) [7]. The brain activity is recognized whenever a potential difference appears between the electrode with an active neural signal and the electrode that is placed in an inactive surface to serve as a reference point. Based on the type of the activity that provokes the signal (spontaneous or event-related), the obtained results are divided in two categories. First group of signals correspond to spontaneous activities which can usually be detected in the frequency range between 1 Hz and 100 Hz. The second group of signals corresponds to various sensory, cognitive or motor events, which can be detected as event-related potential (ERP) in the brain. The ERP measurements always contain some noise which comes from other bio-signals in the brain and various electromagnetic interferences in the environment. Based on the assumption that the noise can be approximated by a zero-mean Gaussian random process, the resulting signal-to-noise ratio (SNR) can be significantly improved by averaging several ERP measurements.

Another measurement of the activity of the central nervous system is based on the analysis of variation of the potential distribution across the eye. The latter reveals the information about the eye blinks and the eyes movements (electrooculogram, EOG). There are also several other methods for measuring cognitive workload based on human’s ocular activity. They are discussed in a separate chapter of this paper.

Increase or decrease of mental workload can be obtained also by monitoring the variation of the heart rate with the electrocardiogram (EKG), or by monitoring the variation of the galvanic skin responses (GSR) as well. Variation of the heart rate is recorded by skin electrodes that measure electrical activity caused by depolarization and polarization of the heart muscle, where increased mental workload leads to an increased cardiovascular activity and vice versa.

III. MEASURMENTS OF OCULAR ACTIVITY

Many ocular activity variables have been tested and validated to be correlated with mental workload, such as pupil diameter, blink rate, blink duration, blink frequency, fixation duration, fixation frequency and saccadic extent. All of these have been used to measure mental workload in vehicles. Experiments have shown that the physiological and performance measures and the remote eye tracking might provide reliable driver cognitive load estimation, especially in simulators [8].

A. *Pupil Dilation*

Pupil diameter is a physiological measure of cognitive workload. The magnitude of pupil dilation can be described as a function of processing mental workload required to perform a given task [9]. When facing an increased visual or mental workload people's pupils tend to increase in diameter. This phenomenon is called the Task Evoked Pupillary Response. Even though it can be measured with most of video based eye tracking systems, setting the system and providing accurate measurements is not effortless. This is mainly due to the visual interference that can appear and is not directly connected to the task. The pupil is very light sensitive and can react to every light change in the environment, such as for example, a vehicle driving in a dynamic road with random lighting. The size of the pupil, as seen by the eye tracking system, depends on the person’s gaze angle. This issue is an important source of error in the measurement of pupil dilation. However, these technical issues can always be omitted and fixed. The procedure of measuring mental workload by observing changes in pupil diameter can therefore, be considered as a reliable indicator of changes in the mental workload.

B. *Blink Duration and Blink Frequency*

Eye blink parameters such as blink duration and blink frequency have been associated with level of drowsiness and information processing. Blink frequency has been reported to increase with greater workload in a way that the number of blinks increases as a function of time in ongoing tasks [10]. Blink duration, however, decreases when the driver is at the beginning of a new task and increases when the driver experiences fatigue or drowsiness. These conditions decrease driver’s mental capabilities and increase overall cognitive workload.

Although all studies confirm that there is a clear correlation between these parameters and the mental workload, the actual results suggesting which parameter is more suitable differ drastically. Fakuda et al. found a correlation between pupil diameters and blink frequency which does not depend on task completion time or amount of information which needs to be processed [10]. The latter suggest that blink frequency can be considered as accurate indicator as the pupil dilation but much easier to detect and measure. Benedetto et al. on the other hand, suggest that blink duration is even more sensitive and reliable indicator of driver’s visual workload as the before mentioned blink frequency [11].

IV. DETECTION RESPONSE TASK MEASUREMENTS

Detection response task (DRT) or Peripheral detection task (PDT) is a method for measuring the amount of driver’s mental workload with the use of a secondary task distraction. DRT has been used in simulator and driving studies in recent years to assess changes in workload during driving, and to assess workload and distraction caused by in-vehicle information systems [12]. The most widespread version of this method consists of a red dotted light that is placed in front of the driver and a physical button placed on the steering wheel as seen in Fig. 1. The red light emitter turns on randomly every 2 to 5 seconds and emits for 1-2 seconds unless it is turned off earlier by the driver with the remote button. The task of the driver is to press the button each time he or she spots the red light. If a new interval starts without the driver pressing the button for the previous one, it counts as miss. The system keeps track of the time needed for the driver to respond to the visual stimuli and all the missed targets. It also counts as a miss, if the driver presses the button more than once for one light interval. All this data is then collected and time-aligned with primary and secondary tasks. The longer response time and the higher number of missed targets signify greater mental workload of the driver.

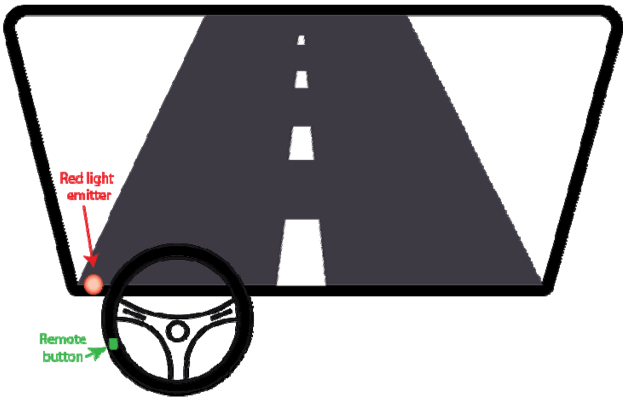

Figure 1. DRT method

Krause M. and Conti A. proposed the DRT measurement device as a simple mobile application which they named MDT [13]. It is available for Android devices and it is free of charge. The MDT works in the same way as DRT, the only difference is that the remote button and the light emitter are now both presented by the telephone's screen (see Fig. 2), which turns red every 2-5 seconds and can be turned off by pressing directly on the screen.

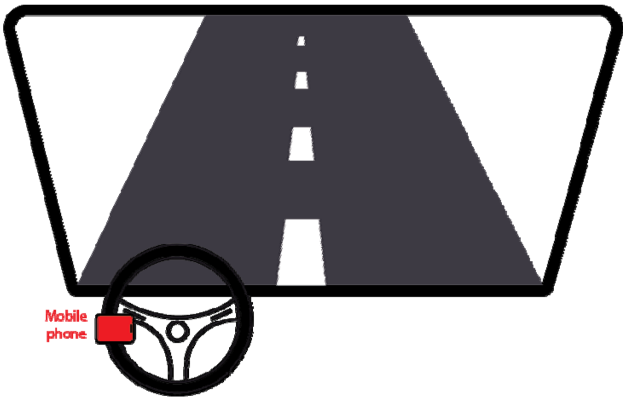

Figure 2. MDT method

Bruyas and Dumont split DRT methods into three categories, based on the stimulus used and its physical placement relatively to the driver [14]. The first method is called Head-mounted DRT (HDRT) since the small diode emitting red light is mounted directly on a driver's head, right in front of the eye (e.g. at the distance of approx. 10 cm). In the other method called Remote DRT (RDRT) the light source is placed in the lower right corner of the driver's windshield. While these two versions provide visual distractions, the third method called Tactile DRT (TDRT) uses tactile stimuli. A small vibrating sensor sends vibration impulses directly to the driver's skin.

The Detection Response Task measurement is currently being discussed in the ISO working group TC22/SC13/WG8 as the basis of a standard to assess the effect of cognitive workload on driver's attention [14].

V. SUBJECTIVE MEASUREMENTS

By subjective measurements we refer to questionnaire based self-evaluation procedures where test subjects are asked to answer different questions regarding their subjective perception of cognitive workload. These questionnaires are always completed after participation in various tasks and experiment conditions.

A. NASA-TLX

The Nasa Task Load (NASA-TLX) is a multi-dimensional rating procedure that provides an overall workload score based on a weighted average of ratings on six subscales [15]:

- Mental Demands,
- Physical Demands,
- Temporal Demands,
- Own Performance,
- Effort and
- Frustration.

Initially there were nine categories including also fatigue, stress and frustration. These three categories were later abandoned as they were found irrelevant for the final assessment of perceived cognitive workload.

The NASA-TLX procedure consists of two parts or two sets of questions (in a paper or an electronic version):

- rating of workload in different categories and
- estimation of weights for individual subscales.

The first part of the procedure requires the test subject to rate each of the subscales (e.g. magnitude of load) on how much is it present and important for a given task. Individual subscales are rated on a scale from "low" to "high" divided into 20 equal intervals (e.g. the final score ranges from 0 to 20).

Estimations of weights is performed by comparing different categories (subscales) pair-wise, forming 15 different comparisons. Whenever an individual subscale (e.g. the source of workload) is selected to be a greater contributor to the anticipated workload for a specific task its counter is incremented by one. The final number of counts represents the weight for individual subscale and ranges from 0 to 5. The overall workload for a selected task is than calculated by summing products of ratings and corresponding weights. Additionally the sum is divided by 15 (the sum of the 15 paired weights) to normalize the final score.

NASA-TLX has been used in various experiments for estimating cognitive workload of drivers offering a reliable comparison of different experiment conditions and tasks ([6], [8], [11], [12], [19], [20], etc.).

B. DALI

The Driving Activity Load Index (DALI) was developed with as a subset of NASA-TLX. While the latter was initially developed to measure pilot's cognitive load, DALI on the other hand is a reversed version which has been adapted for accessing driver's cognitive load [16]. It also consists of 6 subscales:

- Effort of attention,
- Visual demand,
- Auditory demand,
- Temporal demand,
- Interface and
- Situational stress.

These changes were applied since some of the subscales in TLX cannot be entirely correlated with the task of driving (e.g. "physical demand" or "performance"). Pauzé discusses that "physical demand" for example cannot be considered relevant for driving a car since today's cars demand only negligible physical effort in order to be operated efficiently[17]. She also talks about how the "performance" estimation may vary from person to person due to its self-esteem, motivation to fit the expected standards and other similar factors that are not directly related to the mental workload.

The original NASA-TLX is still the simplest and the most often used method for the estimation of cognitive workload also for driving tasks and simulator studies. Consequentially, it is also the most commonly cited method enabling different researchers to directly compare their results. In the following chapter we report briefly on two of our experiments in driving simulator where NASA-TLX was used for estimation of cognitive workload and showed some significant differences between different display technologies in vehicles.

VI. EXAMPLES OF USING NASA TLX FOR THE EVALUATION OF AUDITORY AND VISUAL INTERFACES IN VEHICLES

The main research goal of the first study was to establish if a pure auditory interface in a vehicle can perform better than a classical visual interface displayed on a Head-down Display (HDD) [18]. We intended to find out if it is less distracting for a driver by measuring cognitive workload of drivers and evaluating their driving performance (e.g. counting driving errors and potential dangerous situations on the road, etc.). The user study took place in a driving simulator, consisting of a large projection screen, steering wheel, pedals and the in-build IVIS supporting a variety of tasks related to communication, navigation and multimedia. The IVIS was operated through a custom-made interaction device attached to the steering wheel, which enabled the drivers to enter commands and select functions while holding the steering wheel. The feedback or the output of the system was based on three different display technologies:

- visual HDD mounted on the dashboard,
- mono auditory interface played through a speaker and
- spatial auditory interface played through a 7.1 speaker surround system.

The three technologies were used separately in three isolated experiment conditions. The simulator setup is shown on Fig. 3.

Figure 3. Driving simulator used for evaluation of spatial auditory interfaces and HDD [18]

An important research question was also if the use of spatial audio and multiple simultaneous sounds can increase the efficiency of such auditory interfaces (e.g. by comparing mono and spatial audio). Users were asked to perform different secondary tasks of varying complexity while driving a simulated vehicle (e.g. find a song in the multimedia device and play it, call somebody from the list of contacts, send a short txt message to somebody, etc.) Their performance was evaluated through the measurement of task completion times, counting of driving errors and anomalies and also NASA TLX test. The latter was included to compare individual interfaces by the amount of cognitive workload they put on a driver.

Both auditory interfaces proved to be significantly safer as the HDD resulting in much lower number of driving errors and hazardous situations on the road. They also proved to be equally fast and effective for performing secondary tasks while driving and causing less cognitive workload to the driver. No significant differences were found between the two auditory interfaces in terms of task completion time or driving errors. On the other hand, the results of NASA TLX test reviled some important differences between the two auditory interfaces (see Fig. 4).

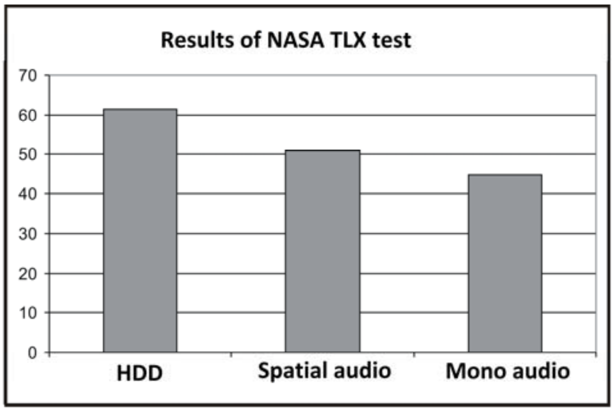

Figure 4. Overall scores of NASA TLX test comparing three experiment conditions [18]

The use of spatial sound and multiple simultaneous sounds resulted in significantly higher cognitive workload. Majority of test subjects substantiated this by clearly explaining that perception of spatial sound in such form is inappropriate and too much mentally demanding for in-vehicle environments.

In the second study we changed the HDD to a Head-up Display (HUD) where output of IVIS was projected directly to the windshield [19]. In this way, it is much easier to read the content and less “eyes off the road” situations occur. The HUD was compared to the mono auditory interface and to a multimodal display (e.g. using HUD and audio output simultaneously). The main research question was how the change of HDD to HUD will reflect in task completion time, safety of driving and cognitive workload. We were primarily interested if a visual display projected as a HUD can be comparable to pure auditory display. Beside the display technologies, the simulator software was also changed in this study resulting in an even more realistic driving experience (see Fig. 5).

Figure 5. Driving simulator used for evaluation of auditory, visual (HUD) and multimodal interfaces [19]

In comparison to the previous study, the results in this case did not show any significant differences between the HUD and the auditory interface. The visual interface did not show to be any less safe than the auditory interface, resulting in a comparable number of driving errors and anomalies. The NASA TLX scores on the other hand, indicated some significant differences between the interfaces. In this study, the estimated cognitive workload caused by the auditory interface proved to be higher than the workload caused by the HUD or by the multimodal interface (see Fig. 6). The latter proves that HUD and multimodal displays should be the next step in the field of in-vehicle display technology providing better safety and lower cognitive workload.

These are just two examples illustrating the importance of correct estimation of cognitive workload in driving tasks. The NASA TLX tests showed some important differences between the evaluated displays and technologies which were not registered through the measurement of other objective parameters (e.g. task completion times and driving errors).

Recently we have started using the DRT method in our research, also for the assessment of cognitive workload in various driving tasks. We are currently conducting an extensive research study in which we are comparing the newly proposed audio version of the DRT to the well reported and commonly used visual and tactile versions of the DRT test.

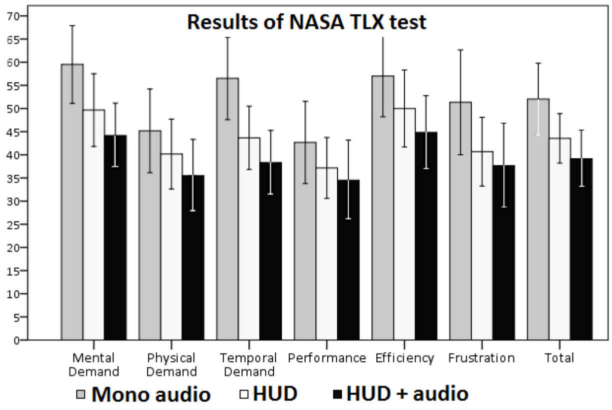

Figure 6. NASA TLX scores of individual subscales comparing three experiment conditions [19]

VII. DISSCUSSION AND CONCLUSION

Nowadays mobility represents an essential part of our lives and majority of the tasks that we used to perform in the office are being done also while walking, running, cycling and even driving a car. All these primary and secondary activities compete for a limited capacity of our cognitive and mental resources. If the cognitive workload of numerous simultaneous activities exceeds the available mental resources, our performance drops significantly and leads to dangerous and unpredicted situations. It is therefore very important to correctly estimate and measure the cognitive workload caused by individual activities and by the use of specific electronic devices and user interfaces. In this paper, we summarized the most relevant methods for assessment of cognitive workload of drivers when driving a car and performing various secondary tasks in a vehicle.

The psychophysiological measures represent a group of the most advanced methods for assessment of cognitive workload, since they require a set of very expensive and complex equipment as well as well-trained professionals capable of recording and interpreting the brain signals. They offer continuous observation with high time resolution and collection of data without disturbance and intrusion into primary tasks [20].

Similarly, the measurements of ocular activity could also represent an accurate and very objective measurement of cognitive workload in various environments and tasks. However, due to the specifics of the driving environment and the complexity as well as high price of the required equipment they are not as commonly used as the DRT methods.

DRT methods are rapidly gaining on their popularity and are being more and more used also for observing in-vehicle activities and tasks. However, the visual version of DRT test itself causes a significant amount of workload and high demand specifically for visual resources. It is therefore unsuitable for complex visually-demanding secondary tasks and should be replaced with the tactile or perhaps auditory version of the same test (e.g. the use of tactile or auditory stimulus).

The subjective questionnaire-based procedures are still most widely used and cited methods for assessing cognitive workload in driving tasks. The results obtained with NASA TLX test can, for example, be directly compared to the biggest number of results of other similar studies and researches. Another major benefit of NASA TLX is also its simplicity, since it requires no specific or sophisticated electronic equipment. On the other hand, the problem with subjective measures is that it is always performed post-hoc and does not measure time-varying qualities. The answers in the questionnaires are for example strongly influenced by events towards the end of the task (e.g. they are more related to the latest events instead of all events which occurred throughout the task) [21].

In general, it is hard to estimate which method is more suitable than the other. This is because the outputs of various methods are not comparable and do not have the same reference parameters. Some measure only the visual workload while the others measure the total workload without knowing which resource centre is actually used for successful performance of various tasks.

Finally, it is important to point out that the mental workload in driving tasks should not be to low either. Small amount of mental demand can cause the driver to turn from an active to a passive participant in the process, making the driving process monotonous and dull (e.g. when driving in semi-autonomous or autonomous vehicle). This can quickly lead to sleepiness and lack of motivation. The final result and mental state can be as dangerous as a mental overload.

In the future, it will be therefore, very important to constantly estimate the level of cognitive workload of drivers and sustain it in the predefined and safe range.

REFERENCES

[1] Workload, From Wikipedia. Available from: http://en.wikipedia.org/wiki/Workload.

[2] Wickens, C.D. and Hollands, J.G., “Engineering Psychology and Human Performance”, Prentice Hall, Saddle River, New York, 2000.

[3] N. Meshkati and P.A. Hancock (editors), “Human mental workload”, *Elsever*, p.8, September 2008.

[4] James, W., “The Principle of Psychology”, New York, NY, USA: Holt, 1890.

[5] Navon, D. and Gopher, D., “On the economy of the human processing system,” *Psychological Review*, vol. 86, no. 3, 1979, pp. 214–255.

[6] Harms, L. and Patten, C., “Measuring distraction from navigation instructions in professional drivers—a field study with the peripheral detection task”. In: *Proceedings from 9th International Conference of Vision in Vehicles*, Brisbane. Elsevier, Exeter, 2001.

[7] Borghinia G., Astolfia L., Vecchiato G., Mattiaa D., and Babiloni F., “Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness”, *Neuroscience and Biobehavioral Reviews* vol. 44 pp.58–75, 2014

[8] Palinko O., Kun A. L., Shyrokov A., and Heeman P., “Estimating cognitive load using remote eye tracking in a driving simulator”. In Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications, 2010, pp.141-144.

[9] Beatty J., “Task-evoked pupillary responses, processing load, and the structure of processing resources”. In Psychological Bulletin, vol.91 (2), 1982, pp.276–292.

[10] Fakuda K., Stern J. A., Brown T. B. and Russo M. B., “Cognition, Blinks, Eye-Movements, and Pupillary Movements During Performance of a Running Memory Task”. In: Aviat Space Environ Med, vol. 76, 2005, pp.75-85.

[11] Benedetto S., Pedrotti M., Minin L., Baccino T., Re A. and Montanari R., “Driver workload and eye blink duration”. In: Transportation Research Part F, vol. 14, 2011, pp. 199-208.

[12] Jahn G., Oehme A., Krems J. F. and Gelau C., “Peripheral detection as a workload measure in driving: effects of traffic complexity and route guidance system use in a driving study”, Transportation Research Part F:Traffic Psychology and Behaviour, vol. 8, May 2006, pp.255-275.

[13] Krause M. and Conti A., “Mobile detection taks”. Available from: http://www.lfe.mw.tum.de/en/research/open-source/mdt/.

[14] Bruyas M. P. and Dumont L., “Sensitivity of detection response task (DRT) to the driving demand and task dificulty”. In *Proceedings of the Seventh International Driving Symposium on Human Factors in Driver Assesment, Training, and Vehicle Design*, 2013.

[15] NASA TASK LOAD INDEX (TLX) Paper and Pencil Manual. Available from: http://humansystems.arc.nasa.gov/.

[16] Pauzié A. And Pachiaudi G., “Subjective evaluation of the mental workload in the driving context”. In: Traffic and Transport Psychology: Theory and Application, 1997, pp. 173-182.

[17] Pauzié A, “Evaluating driver mental workload using the driving activity load index (DALI)”. In Procedings European Conference on Human Centred Design for Intelligent Transport Systems, 2008, pp. 67-77.

[18] Sodnik, J., Dicke, C., Tomažič, S., Billinghurst, M., “A user study of auditory versus visual interfaces for use while driving”. In: International Journal of Human-Computer Studies vol. 66(5), 2008, pp. 318-332.

[19] Jakus, G., Dicke, C., Sodnik, J., “A user study of auditory, head-up and multi-modal displays in vehicles”. In: Applied ergonomics, vol. 46, 2015, pp. 184-192.

[20] Wilson, G.F. and Russell, C.A., “Performance enhancement in an uninhabited air vehicle task using psychophysiologically determined adaptive aiding”. In: Human Factors, vol. 49, 2007, pp. 1005-1018.

[21] Insko B.E., “Measuring presence: Subjective, behavioral and physiological methods”. In: Being There: Concepts, effects and measurement of user presence in synthetic environments, 2003.