EVALUATING DRIVER MENTAL WORKLOAD USING THE DRIVING ACTIVITY LOAD INDEX (DALI)

Annie Pauzié (INRETS, France)

annie.pauzie@inrets.fr

ABSTRACT: Methodologies are required to support the user-centred design and evaluation of in-vehicle Information and Communication Technology (ICT). This paper reports the outcomes of three on-road experiments conducted to assess the usability of mobile phone and guidance/navigation systems. These utilized a method for evaluating drivers' mental workload: the Driving Activity Load Index (DALI). For the guidance situation, functions were tested for two systems corresponding to two levels of technology maturity. Outputs of the DALI were used to identify which aspects of the old generation system had to be improved, and how the new generation could be designed to be more usable by drivers.

1 Introduction

Many research studies have been conducted to investigate the road safety consequences of deploying Information and Communication Technology (ICT) in vehicles [1],[2]. The objective is to create a balance between the potential interference induced by these systems versus the potential benefits that can be derived from them in supporting the driving task.

To evaluate these systems, it is necessary to have an efficient methodology that is applied according to the type of function, the type of system and the context in which the system is used [3]. A quite exhaustive overview of available methodologies, tools and techniques has been conducted within the framework of the network of Excellence HUMANIST [4]. Classically, the parameters that have been taken into consideration for safety evaluation have been related to the vehicle; for example, deviations from vehicle trajectories deriving from system use [5]. Other parameters include drivers' visual strategies, visual demand deriving from use of on-board screens [6], general driver behaviors [7],[8] and overall driver workload according to the situation [9].

The assessment of workload is coupled with the difficulty of the task experienced by the individual [10], because several reactions to task demands are possible. The individual can adapt his behaviour to increased demand, leading to a higher investment of effort but no perceptible effect on performance; or, on the contrary, he can decide to change his strategy with a lower level of performance. Then, moderate increases in task difficulty may produce few observable changes in error rate, as the driver attempts to keep performance constant by allocating more resources to the task [11]. Furthermore, inter-individual strategies are variable; some individuals develop more effective strategies which require less effort to reach a level of performance than do others. So, for all these reasons, objective performance measures, which directly measure performance, are not sufficient by themselves to evaluate the overall demands of a given situation.

Mental workload is a variable difficult to assess, in comparison with other variables. Several methods have been developed to measure mental workload [12]: measurements of physiological parameters, such as heart rate [13]; the dual-task method [14]; and methods that elicit drivers' subjective judgments about the workload they have experienced. The latter include S.W.A.T (Subjective Workload Assessment Technique; 15] and the NASA TLX - Task Load Index [16]. Subjective measures are often used in practice because they have many practical advantages over objective measures [17],[18],[19].

The next section of this paper discusses subjective methods for the evaluation of drivers' mental workload.

2 Subjective evaluation of drivers' mental workload

2.1 Methods for subjective evaluation of mental workload

The subjective method allows for the evaluation rather than the measurement of mental workload by comparing the perception of workload between situations. It can therefore be regarded as a global, and even a "crude", criterion. Subjective evaluation is often conducted in association with other workload measurement techniques [20].

The SWAT is a sophisticated workload assessment tool, composed of a two-step process: in a scale development phase, data necessary to develop a workload scale are obtained from individuals; during an event scoring phase, people rate the workload associated with a particular task [21]. The primary assumption of SWAT is that workload is function of three dimensions: *time load, mental effort load* and *psychological stress*, each dimension having three possible levels. All possible combinations of the three levels of each dimension yield a 27-cell, three dimensional, matrix to represent workload.

The NASA TLX method assumes that workload is influenced by *mental demand, physical demand, temporal demand, performance, frustration level* and *effort*. After assessing the magnitude of each of these six factors on a scale, the individual performs pair wise comparisons between these six factors, in order to determine the higher source of workload factor for each pair. A composite note quantifying the level of workload is set up by using both factor rating and relative weights computed from the comparison phase.

The NASA-TLX has been tested and used by the army; being considered as superior in terms of sensitivity than other methods and well accepted by the operator [22].

The DALI (Driving Activity Load Index) is a revised version of the NASA-TLX, adapted to the driving task [23]. As previously mentioned, mental workload is multidimensional and, among other things, depends upon the type of loading task. The NASA TLX was originally designed to assess pilot workload in the aviation domain.

The basic principle of DALI is the same as that for the TLX. There is a scale rating procedure for six pre-defined factors, followed by a weighting procedure in order to combine the six individual scales into a global score. The main difference lies in the choice of the main factors composing the workload score.

For the NASA TLX, one of the factors to be rated is called the Physical Demand component and is usually defined in the following terms : " How much physical activity was required ? -pushing, pulling, turning, controlling, activating,...-" It appears that this question would not be very relevant when considering the driving activity where the control of the vehicle is quite automatic for an experienced driver, and where maneuvers are not supposed to be physically demanding in modern cars.

Another example is the Mental Demand component defined in the TLX as follows " How much *Mental* and perceptual activity was required? - thinking, deciding, calculating, remembering, looking, searching,...". This statement covers both perceptual and cognitive aspects of workload, and we think it would be interesting in the context of the driving task to be able to identify these various modalities.

Finally, the evaluation of the Performance factor can be made using objective data. The subjective rating of a good performance by the driver can show discrepancies with the measured one, but this difference might be due to many factors other than the mental workload itself - low or high self-esteem, motivations to fit to the standard performance,...-

DALI was derived principally by asking various experts involved in driving task studies to define which were, in their opinion, the main factors inducing mental workload for people driving a vehicle equipped with an on-board system (car phone, driving aid system, radio,...). This investigation led to the development of six workload dimensions for DALI: Effort of attention, Visual demand, Auditory demand, Temporal demand, Interference and Situational stress.

Key results of previous and recent studies using the DALI tool are summarized in the following paragraph. An overview is provided of the advantages and the limits of this task load index according to the purposes of the investigation and the various contexts in which it is used.

2.2 Evaluation of driver’s workload using Hand-Free mobile Phone

Using a mobile phone while driving raises the issue of road safety. Unlike other Information and Communication technology developed to support the driving task, the activity of phoning is disconnected from the driving task itself. There is then no benefit of this function in terms of enhancement of the driving task process for the driver. Use of the mobile phone while driving would induce a high probability of interference in terms of attentional demand for the driver [24],[25].

Nevertheless, some diversified findings are encountered in the literature, linked to the modalities of experimental conditions. In order to test the load of conversation, one author [26] used secondary verbal task techniques to assess mental load while driving. According to this author, drivers showed no significant changes in their driving performance while verbalizing, which made him assume that the driving task priority was maintained as intended. Brown, Tickner & Simmonds [27] used a verbal reasoning task based on grammatical transformation in order to assess the effect of phoning while driving. Results indicated increased errors in judgment of gaps, decreased skill in steering through narrow gaps and decreased speed. Drory [28] measured driver behaviour when using a mobile phone in a driving simulator. No serious performance decrement was found concerning the driving activity, except when the subjects actually dialed the number. Mikkonen & Backman [29] studied the influence of phone conversation on driving performance in a familiar urban environment. In this case, drivers paid more attention to their task, increasing their alertness and their anticipation behavior. Tokunaga & col. [30] showed the negative impact of the complexity of the phone conversation on reaction time and on NASA-TLX values, both for young and older drivers.

In order to evaluate the efficiency of the DALI as a tool for the assessment of mobile phone use, an experiment was carried out in a real road context (see Pauzié & Pachiaudi [31] for detail of the experimental protocol and results). The objective was to investigate evaluation of perceptual and cognitive load for the driver in this phoning condition, knowing that specific factors of the DALI were dedicated to these aspects. Results indicated that the global value of mental load increased significantly when phoning and driving in comparison with the reference situation corresponding to no system use. The load index was significantly high for “auditory” and “interference” factors, in addition to “stress”. The effort of “attention”, although higher than during a simple driving task, does not increase significantly. So, in terms of subjective evaluation of the workload, drivers identified the disturbance induced by phoning, through the perceptive channel of audition, on managing the driving task, which induced stress. Through this example, it is possible to illustrate how a tool based upon subjective evaluation and driver awareness can allow for the understanding of mental workload and cost of the task.

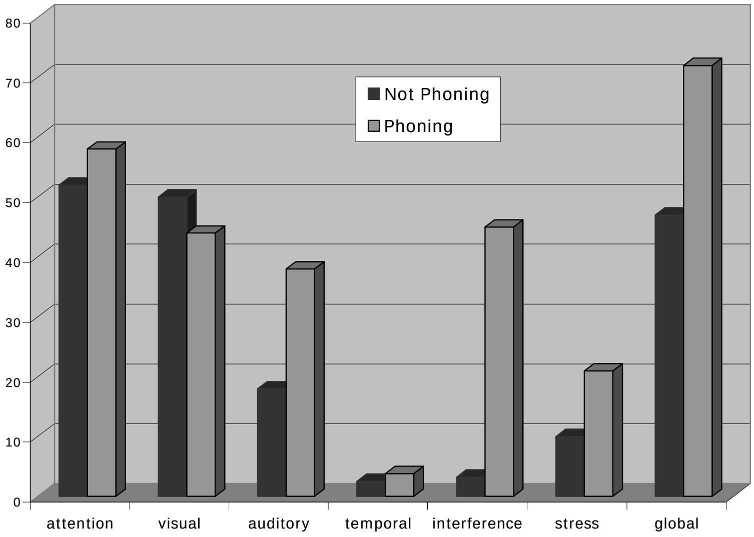

Fig.1. Factors and Global Value of the DALI for hand-free mobile phone use.

<u>Advantages</u>: Allow to better understand how the implementation of a system in a vehicle can be experienced by driver (in this example, the most significant factor was “interference” due to phoning while driving).

<u>Limits</u>: Subjective evaluation gives information only on driver's awareness about workload he/her experienced in a defined context

2.3 Evaluation of driver's workload using Navigation/Guidance function

Navigation and guidance functions have been developed to support drivers at the strategic level of the driving task, by supporting the navigation process. They also support driving at the operational level, by supporting drivers to anticipate upcoming maneuvers. Theoretically, driver reliance on auditory and visual instructions to support their way finding decisions should decrease mental workload and reduce driving errors. However, this objective can only be reached if the system is correctly designed, that is to say avoiding misconception, with correct timing for displayed messages, not too soon and not too late, and clear, legible and visible visual information. Evaluation of drivers' mental workload, in addition to driving errors, is necessary to enhance the effectiveness, usability and acceptability of systems developed to support these functions [32],[33].

Two experiments conducted with a 9-year time lapse [34] are especially revealing about the importance of making a distinction between the "benefit of the function", such as "instructions to guide the driver", and "design of the system for this function" such as "modalities of displayed instructions, timing of auditory messages, legibility and understandability of messages". The first experiment described was testing the first generation GPS-based system; the second was conducted using a new-generation version of the system, both implemented by the same car manufacturer.

2.3.1 Drivers' workload for old-generation guidance system

It was proposed that this system provide the option of auditory guidance instructions or an electronic map display. The experiment was conducted in a real road context and the purpose of it was to compare drivers' workload for these two options with a reference situation (that is, no system but a paper map; Pauzié & Pachiaudi, 1997).

According to the DALI "global" score, use of the system corresponded to a significantly higher workload for the driver in comparison with the "reference" situation (guidance or electronic map). The DALI factors, "auditory" and "temporal" demands were both critical factors, rather than "visual" load, with high values for "stress" and "attention". Based upon this data, it was possible to determine that the messages displayed by this specific system, whatever the option of guidance (display of arrows) or navigation (display of electronic map), has a poor timing when delivering auditory instructions to the driver. These results showed also that the navigation option induced a high rate of interference, in comparison with the two other contexts of guidance and reference situations.

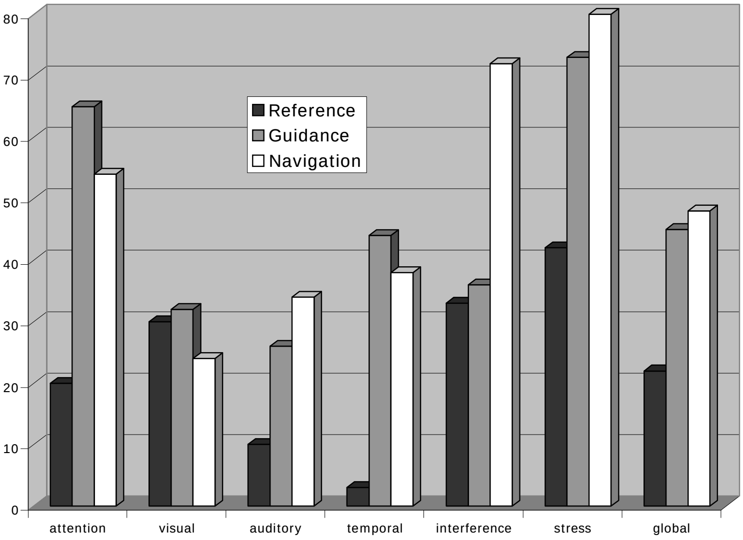

Fig.2. Factors and Global Value of the DALI for Guidance and Navigation

Old generation system.

<u>Advantages</u>: Allow to identify the weakness of the design characteristics that induced workload for the driver (in this example, poor timing of the auditory messages)

<u>Limits</u>: Necessity to analyze objective variables such as driving errors to complete the investigation (in this example, the value of driving errors could modulate the conclusions regarding reference situation versus guidance situation; it can be costly to use the system for the driver but it can induce less driving errors).

2.3.2 Driver's workload and new-generation of guidance system

Diversified contexts of orientation processes have been set up, varying according to their level of workload for the driver, in order to test the validity of the method DALI. Overall, 4 situations were identified, from HIGH to LOW demand: "complex system" requiring cognitive and perceptive attentional demand, "paper map" with no system, "guidance system" correctly designed and "human co-pilot" giving instructions to the driver. The four situations have been processed in real road contexts in a turnover order between drivers to avoid effect of practice and learning [34].

According to the DALI global score, there is a significant difference between the 4 experimental sessions (Wilcoxon, Z= 3,007, p=0,003; Z= 2,224, p=0,026, Z= 2,539, p=0,011; Z= 3,923, p<0,001). More precisely, contrary to the previous experiment, "use of guidance instructions" induced generally a lower workload than "use of a paper map", identified as "reference" in the previous paragraph. Looking at the detail of the DALI factors, it appears that support of the system for the driver is in terms of "stress", "interference between driving and finding his route", "temporal", "visual" and "attentional" demand, with significant differences. Of course, "auditory" demand was not rated by the driver in the context of the paper map use.

Hence, these results demonstrate that a guidance system correctly design in terms of visual and auditory messages (timing, loudness, content) is an added value for the driver by making the orientation task lighter in terms of cognitive and perceptive processes. Furthermore, the DALI results showed that there is a higher level of workload while using the system in comparison with relying on the human co-pilot. Hypothesis can be made that this system could require a phase of training longer than the timing of this experiment, in order for the driver to be fully comfortable with the system. Additional testing with a longer training phase could indicate if the system can be equivalent to a human co-pilot or not. At least, the DALI results indicated that this system is superior to a paper map.

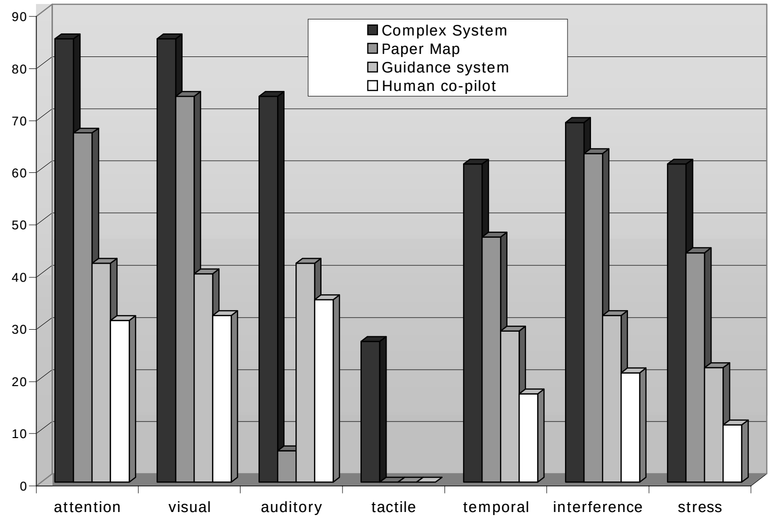

Fig.3. Factors and Global Value of the DALI for Guidance and Navigation.

New generation system

<u>Advantages</u>: Allow identifying correctly design functions of in-vehicle system, able to support driving task, in comparison with situation with no system (in this example, driving with a well- designed guidance system in terms of visual and auditory messages (timing, loudness, content) induced less workload for the following factors “attention”, “visual”,” temporal”, “interference”, “stress” in comparison with the situation of paper map use).

<u>Limits</u>: Necessity to set up several contexts to be able to use this tool (reference situation versus tested situation or several tested systems), as it allows relative and not absolute results.

2.3.3 Driver’s workload and maturity of technology

The two previous experiments illustrated evolution of the design characteristics of a system for a given function in relation to the improvement of the technology. They highlighted the need to keep in mind, while testing in-vehicle ICT, that both functions and system design can have an impact on usability and safety, but at a different level. At this stage, and based upon literature investigations, it can be considered that the guidance function by itself has a positive impact on the driving task (less driving errors and hazardous behavior, less workload in comparison with paper map use) as long as the guidance instructions are correctly designed. Indeed, in this experiment, a specific care on the HMI design of the guidance instructions was applied: enhanced contrast, simplified visual display, well-paced auditory message. The objective was to evaluate the added value brought by the guidance functions in an optimized context of HMI design. The chosen HMI design was not the only solution, but avoid poor interface such as monochrome screen (example: mobile phone on the left of the figure 4) or complex map display.

To summarize, the use of the DALI can support the designer in identifying which aspects of the driver-vehicle interface are inducing perceptual and cognitive demand for the driver due to poor design of the system, and to improve next generation of prototypes, allowing also to take into account the quick evolution of the technology in this domain through the iterative processes of design and evaluation (figure 4: examples of two successive generations of mobile phones HMI).

Fig.4. Old and new generation of mobile phone giving guidance information to the driver

3 Conclusion

The measurement of driver workload complements other workload metrics in bringing additional information and allowing broader understanding about the complex interactions between drivers and the systems they use while driving. DALI, as a subjective workload evaluation tool, allowed for the gathering of data that was usable by the designer in improving his system prototype. It allowed enabled identification of the impact of a given system implementation by comparing results with a reference situation with no system. One of the main advantages of this tool is that it makes it possible to identify the origins of driver workload, allowing for corrective action at the identified level (e.g., high interference and visual load indicate that an in-vehicle system has a demanding visual display). The possible design improvement would be to add factors linked to specific aspects of the driving task useful to evaluate the impact of ADAS functions (e.g. level of stress to keep distance with the vehicle ahead, in the case of a system having an impact on this specificity of the driving task). It is planned to conduct further investigations to improve this method by varying these types of situations. The “DALI tool kit”, comprising the detailed method and procedures for the automatic computation of statistics and the display of graphical outputs, will soon be available on the INRETS web site (www.inrets.fr), allowing any researcher to use it in his/her scientific context.

4 References

[1] Carsten, O.M.J. and Nilsson, L. (2001) Safety Assessment of Driver Assistance Systems. European Journal of Transport and Infrastructure Research, 1 (3). pp. 225-243. ISSN 1567-7141.

[2] Vaa T., Gelau C., Penttinen M. Spyroupolou I., 2006, its and effects on road traffic accidents - State of the art, 13 World Congress on ITS, London, 9 October 2006.

[3] Bekiaris E. & Stevens A., 2005, Common risk assessment methodology for advanced driver assistance systems, Transport Reviews, Volume 25, Issue 3 May 2005 , pages 283 – 292.

[4] Gelau C., Stevens A. & Cotter S., 2004, Impact of IVIS on driver workload and distraction: Review of assessment methods and recent findings, Deliverable D.2/E.2, HUMANIST NoE, 2004.

[5] Zwhalen H.T.& Balasubramanian K.N. (1974). A theoretical and experimental investigation of automobile path deviation when driver steers with no visual input, Transportation Research Record, 520, 25-37.

[6] Pauzié A., 2002, In-vehicle communication systems: the safety aspect, Injury Prevention Journal; 8, 0–3.

[7] Vitense H.S., Jacko J.A., Emery V.K., 2003, Multimodal feedback: an assessment of performance and mental workload, Ergonomics, Volume 46, Numbers 1-3 / 15, Taylor & Francis, January 2003, pp 68 – 87.

[8] Brookhuis K., de Waard D. and Janssen W., 2001, Behavioural impacts of Advanced Driver Assistance Systems–an overview, EJTIR, 1, no. 3 (2001), pp. 245 – 253.

[9] Lansdown T.C., Brook-Carter N., Kersloot T., 2004, Distraction from multiple in-vehicle secondary tasks: vehicle performance and mental workload implications, Ergonomics, Volume 47, Number 1 / 15, Taylor & Francis, January 2004, pp 91 - 104.

[10] Gopher D. & Donchin E. (1986). Workload - an examination of the concept. In K.R. Boff, L. Kaufman & J.P. Thomas (eds.), Handbook of perception and human performance. Volume II, Cognitive processes and performance. New York : Wiley.

[11] Zeitlin L. R., 1995, Estimates of Driver Mental Workload: A Long-Term Field Trial of Two Subsidiary Tasks, Human Factors, Vol. 37, 1995

[12] De Waard D. (1996). The measurement of drivers' mental workload, PhD thesis, Traffic Research Centre VSC, University of Groningen, The Netherlands, 125 p.

[13] Casali J., and Wierwille W.A. (1984). "On the measurement of pilot perceptual workload : a comparison of assessment techniques addressing sensitivity and intrusion issues". Ergonomics, 27, 10, pp 1033-1050.

[14] Shingledecker C. (1982). Performance evaluation of the embedded secondary task technique. Aerospace Medical Association Annual Scientific Meeting, 151-152.

[15] Reid G., Shingledecker C., Nygren T. & Eggemeier T. (1981). Development of multidimensional subjective measures of workload, International Conference on Cybernetics & Society, Atlanta, 403-406.

[16] Hart, L. A.,, Staveland, L., "Development of the NASA Task Load Index (TLX): Results of empirical and theoretical research, In P. A Hancock and N. Meshkati (Eds.), Human Mental Workload, pp. 139-183, Amsterdam, North-Holland, 1988.

[17] O'Donnell R.D., Eggemeier F.T., 1986, New York: John Wiley and Sons, Boff KR, Kaufman L, Thomas JP, eds. Handbook of Perception and Human Performance. Volume II.

[18] Wierwille, W. W., & Eggemeier, F. T. [1993). Recommendations for mental workload measurement in a test and evaluation environment. Human Factors, 35, 263-281.

[19] Colle H.A., 1998, Context Effects in Subjective Mental Workload Ratings, Human Factors, Vol. 40, 1998.

[20] Sheridan T. (1990). Human factors of driver-vehicle interaction in the IVHS environment, Center of Transportation Studies, MIT, n° DTNH22-89-Z-07595, NHTSA, U.S. Department of Transportation.

[21] Reid G. and Nygren T. (1988). The subjective workload assessment technique : a scaling procedure for measuring mental workload". In Human Mental Workload, P.A. Hancock & N. Meshaki (eds.), Elsevier, North-Holland.

[22] Hill S., Lavecchia H., Byers J., Bittner A., Zaklad A. & Christ R. (1992). Comparison of four subjective workload rating scales, Human Factors, 34, 4, 429-439.

[23] Pauzié A. & Pachiaudi G., 1997, Subjective evaluation of the mental workload in the driving context, in " Traffic & Transport Psychology : Theory and Application " , T. Rothengatter & E. Carbonell Vaya (eds.), pp 173-182, Pergamon.

[24] Tokunaga R.A., Hagiwara T., Kagaya S. Onodera Y., 2000, Cellular telephone conversation while driving : Effects on driver reaction time and subjective mental workload, Human Performance : Driver Behavior, Road Design, and Intelligent Transportation Systems. Annual Meeting of the Transportation Research Board No79 , 2000, no 1724, pp. 1-6.

[25] Patten C, Kircher A., Östlund J. and Nilsson L., 2004, Using mobile telephones: cognitive workload and attention resource allocation Accident Analysis & Prevention, Volume 36, Issue 3, May 2004, pp 341-350.

[26] Wetherel A. (1981). The efficacy of some auditory-vocal subsidiary task as measures of the mental load on male and female drivers, Ergonomics, 24 (3), 227-248.

[27] Brown I.D., Tickner A.H. & Simmonds D.C.V. (1969). Interference between concurrent tasks of driving and telephone. Journal of Applied Psychology, 53, 419-424.

[28] Drory A. (1985). Effects of rest versus secondary task on simulated truck driving task performance. Human Factors, 27 (2), 201-207.

[29] Mikkonen V. & Backman M. (1988). Use of the car telephone while driving. Technical Report n° A39. Department of Psychology, University of Helsinki.

[30] Tokunaga R.A., Hagiwara T., Kagaya S. Onodera Y., 2000, Cellular telephone conversation while driving : Effects on driver reaction time and subjective mental workload, Human Performance : Driver Behavior, Road Design, and Intelligent Transportation Systems. Annual Meeting of the Transportation Research Board No79 , 2000, no 1724, pp. 1-6.

[31] Pauzié A. & Pachiaudi G., 1997, Subjective evaluation of the mental workload in the driving context, in " Traffic & Transport Psychology : Theory and Application " , T. Rothengatter & E. Carbonell Vaya (eds.), pp 173-182, Pergamon.

[32] Ashby M.C., Fairclough S.H., Parkes A.M., A comparison of route navigation and route guidance systems in an urban environment, Proceedings of ISATA Conference, 1991.

[33] Pauzié A., Manzano J., 2007, Evaluation of driver mental workload facing new in-vehicle information and communication technology, Enhanced Safety in Vehicle Conference, Lyon, France.

[34] Pauzié A., Manzano J., 2006, Subjective evaluation of the driver's workload in real road experiment, AIDE Deliverable 2.2.6.