Toward an on-line and non-obtrusive workload assessment method

Marianne Pichon, Patrick Millot, Simon Enjalbert

Toward an on-line and non-obtrusive workload assessment method

Marianne PICHON [1,2,3]**, Patrick MILLOT** [1,2,3]**, Simon ENJALBERT** [1,2,3]

1 Univ Lille Nord de France, F-59000 Lille, France

2 UVHC, LAMIH, F-59313 Valenciennes, France

3 CNRS, FRE 3304, F-59313 Valenciennes, France

marianne.pichon@gmail.com,, patrick.millot@univ-valenciennes.fr, simon.enjalbert@univ-valenciennes.fr

Abstract: Workload is an index introduced in the 70th for ergonomic purposes, for evaluating the adequacy of tasks to the human operator abilities. Methods based on a self evaluation such as SWAT and TLX gave the best results, but mainly for assessing a total workload after the task has been performed. But in highly dynamical systems as transportation, the driver (or pilot) abilities can be enhanced by driving assistance tools on-line. Therefore new challenges appear which needs new methods for assessing Workload on-line, in real time and without disturbing the human operator. At LAMIH, an online assessment method has been developed and partially validated by Millot in the past. This method is based on one hand, on two workload generators: temporal demands (time pressure) and functional demands (task difficulties). On the other hand, to cope with a possibility of assessment on-line, the workload is defined through the analogy with the physical notions of "power" and "energy". The instantaneous workload Wl(t) is seen as the "power" the human operator invests on-line in the task in order to cope with the task demands. After a time available denoted TA, the human operator has spent a quantity of "energy" WL defined as the sum of the successive instantaneous Wl(t) along TA. These ideas have been validated in multitask situations for discrete as well as continuous tasks like driving tasks, but especially with temporal demands. This paper first compares several methods with the LAMIH's method. It then proposes an extension of this LAMIH's method in order to cope with the new dynamical constraints. Finally it proposes experimental protocols for validating the new LAMIH's method by comparing it with SWAT and TLX methods.

Keywords: Workload assessment methods, multi-task context, sampling period, temporal load, functional load

1. INTRODUCTION

The present work is a part of the European FP7 project ITERATE (Information Technology for Error Remediation And Trapping Emergency). A review of workload assessment methods was needed in order to build driver behavioural model for cars, ships, trains face to critical situations.

The base of this study is the workload definition given by Sperandio was chosen (1978): workload is the rate of activity, *i.e.* physical mental and sensori-motor, supplied by the operator in order to perform a task.

From this definition, several characteristics of a workload assessment method of workload can be highlighted. The workload measure must take into account the different kinds of activities performed by the operator. Particularly, Casali and Wierwille introduced four generic classes of activities, *i.e.* perception, mental, communication and psychomotor (Casali and Wierwille, 1983; Wierwille et al., 1985). Each activity generates a dedicated *i.e.* perceptual load, mental load, communication load and physical load, each contributing to the global workload.

Different workload assessment methods were then tested and compared according to several criteria, *i.e.* sensitivity, diagnosticity, selectivity, obtrusiveness, bandwidth and reliability (O'Donnel and Eggemeier, 1986). Particularly, obtrusiveness can be observed when the same human processing mechanisms are used for the evaluation as well as the task execution. One of the major results were that the self-rate assessment workload methods gave the best results according to the criteria given previously. Two self-rate methods will be analysed in this paper, SWAT (Subjective Workload Assessment Technique) (Reid et al., 1987) and TLX (NASA Task Load Index, Hart & Staveland, 1988).

In our present dynamical application contexts, two complementary constraints for selecting a relevant workload assessment method must be added to the criteria above. The assessment method must be used on-line during the task performing and does not disturb the operator (no obtrusiveness). A track to cope with the obtrusiveness issue can be inspired from the Sheridan's method developed by Sheridan and based on time measures (Sheridan, 1979). Another way of solution is the LAMIH's method introduced by Millot (1988). The advantages of this method are the assessment on-line and the low obtrusiveness.

An analysis of these four workload assessment methods and a comparison is then presented, in order to highlight the method which reaches the maximum of the criteria.

2. ANALYSIS OF FOUR WORKLOAD ASSESSMENT METHODS

2.1 Two self-rate assessment methods: SWAT and TLX

Several studies have shown that SWAT and TLX are sensitive (Hill et al., 1992), non-intrusive (Rubio et al., 2004) and usable workload assessment methods (Mitchell et al., 2005).

TLX is a multi-dimensional rating procedure that provides an overall workload score based on a weighted average of rating on six dimensions: mental demands (MD), physical demands (PD), temporal demands (TD), self satisfaction regarding the operator own performance (OP), effort (EF), and frustration (or stress, FR).

Each of the six dimensions is quoted R_i, $i \in [0, 6]$ and is subjectively evaluated by the human operator her/himself by pointing a mark on a continuous scale between 0 to 1; $R_i \in [0, 1]$.

Then quantifying workload on a unique scale, needs the aggregation of each R_i in weighting each one regarding the others trough a weighting factor α_i, as shown in the expression (1).

$$WL = \sum_{i=1}^{6} \alpha_i . R_i, \text{ with } \sum_{i=1}^{6} \alpha_i = 1 \qquad (1)$$

The human operator is asked to evaluate each α_i. For that purpose the analyst builds combination of two dimensions (pairs) and asks the operator to choose the most important one. Therefore with the 6 dimensions, we build 15 different pair-wise comparisons of dimensions. This operation is performed after the end of the task. Despite this method cannot be used on-line, it satisfies the criteria given by O'Donnel and Eggemeier. Then it gives a total workload value, *i.e.* the whole resources supplied by the operator during the task duration.

SWAT assumes that workload is made of 3 dimensions; time load (temporal pressure), mental load (functional demands) and psychological stress load. Each dimension is quantified with 3 levels; Low (L), Medium (M) and High (H). Therefore the workload measure is on a range of 27 values defined by the triplet (time load level, mental load level, psychological stress load level) and asks to the operator to classify these cards from the lowest to the highest workload value. This operation is performed before the experiment. This classification depends on the appreciation of the operator who evaluates her/his own workload in a general situation. But an important issue is that it remains constant along time (Valot et al., 1997).

Then after this first step of calibration, the evaluation can be done online in specific situations, by asking the operator to evaluate the level of time load, mental load and psychological stress load. The place of the corresponding card among the 27 cards allows estimation of operator's workload. Experiments on aircraft pilots allow validating the method with questions, asked on-line with a sampling period of 5 minutes.

SWAT takes into account the different generators of load and satisfies to the criteria given by O'Donnel and Eggemeier. Contrary to TLX, this method can be used on-line but with a quite long sampling period. Moreover the operator has to stop her/his task to evaluate her/his workload and this can be obtrusive for some being of tasks which need the same processing mechanisms as the estimation.

2.2 Sheridan's Workload assessment

For mainly temporal tasks Sheridan defined workload as the operators' occupation rate (2) (Sheridan, 1979).

$$WL = \frac{TR}{TA} \qquad (2)$$

Where TA is the time available to perform the task and TR, the time required by the human operator to perform the task.

Therefore TA only depends on the task demands and TR takes into account the difficulties that the operator has to face with. More the operator is face to a difficult task, more her/his time required TR is long. The implementation can be simple as the analyst is able to detect the instants when the task begins and ends and the instants when the human starts and stops. It is also non obtrusive. Furthermore it gives a total workload.

2.3 LAMIH's workload assessment

LAMIH's method distinguishes the task demands on one hand which corresponds to the intrinsic task difficulties and workload on the other hand which results of the human activities for performing the tasks. These activities can be more or less difficult to perform according to the operators own capabilities for instance, skill, expertise, health, etc.

The method follows two steps: the first step deals with the formulation of the task demands, the second one consists in observing how the human takes these demands into account (Millot, 1988).

For the task demands formulation, the time available is a function of time TA(t). TA(t) has a linear expression (3) Fig. 1. , for discrete tasks, *i.e.* tasks which has a beginning and a final instants.
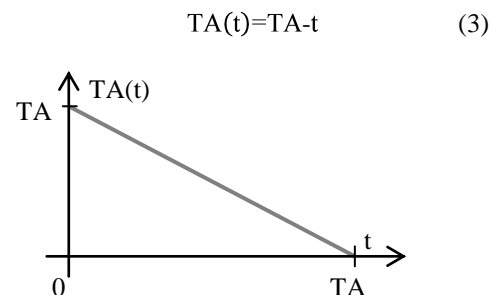
$$TA(t) = TA - t \qquad (3)$$

Fig. 1. Time Available variable TA.

For a continuous task, TA(t) has another definition, for instance in the context of a car driving, the Time Available is given (4) and Fig. 2. (Riera et al., 1990).
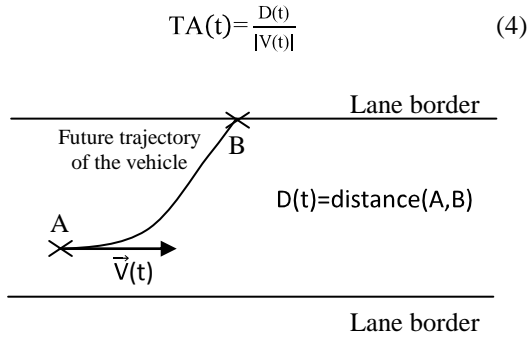
$$TA(t) = \frac{D(t)}{|V(t)|} \qquad (4)$$

Fig. 2. The car driving context.

These expressions are related to pure temporal demands. Millot introduced the functional part of the demands trough an artefact. In fact, the effect of an increase of a task difficulty can be seen as a decrease of the available time, *i.e.* the increase of the temporal demands. For that he introduced the gravity of the situation G(t) representing the functional demand of the task, felt by the operator. The resultant Time Available is called subjective time available TA_S calculated by the expression (5).

$$\frac{1}{TA_S(t)} = \frac{1}{TA(t)} \cdot G(t) \qquad (5)$$

For the second step aiming at formulating instantaneous workload Wl(t), Millot used an analogy with the Physic notions of power and energy. When performing a task the operator is supposed to invest a "power" sufficient to satisfy the task demands. Therefore instantaneous workload can be written by expression (6).

$$WL(t) = \frac{1}{TA_S(t)} \qquad (6)$$

With the same analogy, he defines total workload WL the expression (7), as the "energy" spent along the task duration TR. WL can be seen as a value of the fatigue due to the task *i.e.* the resources which have been consumed.

$$WL = \int_0^{TR} \frac{dt}{TA_S(t)} \qquad (7)$$

Generally, the gravity G(t) is not a continuous function as we can see Fig. 3. ; Therefore WL cannot be easily calculated with (8). Millot introduced Wl_j called sampling workload and showed that the Wl_j is a good estimator of Wl(t) if the sampling period ΔT is very short (Millot, 1988).

$$Wl_j = \int_{(j-1)\Delta T}^{j\Delta T} \frac{dt}{TA_S(t)} \qquad (8)$$

Where Wl_j is the sampling workload value during the jth sample and ΔT is the sampling period

These formulations were tested by experiments in the context of supervision of discrete variables and compared with values of workload assessed by the MCH scale (Modified Cooper-Harper Scale) (Millot and Kamoun, 1988; Debernard et al., 1990). Different levels of difficulty

of this task were used during experiment, low, medium and high. Workload was assessed with the expression (8) for an empirical evolution of G, drawn Fig. 3. The obtained values were compared with a dedicated MCH (Millot, 1988). A good correlation between the method and the MCH, and the sensitivity of the LAMIH's method with the task difficulty has been shown.
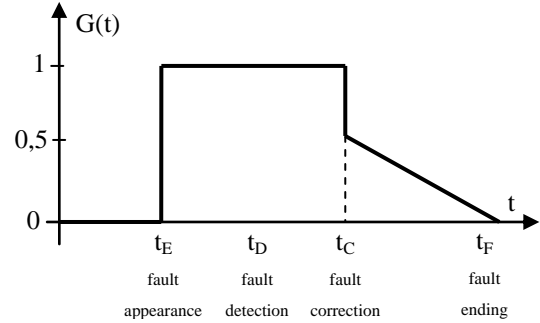

Fig. 3. Example of a gravity function G(t).

Another issue introduced by the LAMIH's method is the context of the task operating, *i.e* mono-task or multi-task context. The method supposed that in multitask situations, the operator defines a new available time for each task which is reduced proportionally to the number of tasks (9).

$$TA_i^r(t) = \frac{TA_i(t) \times \max_{k \in [1,n]} TA_k(t)}{\sum_{l=1}^{n} TA(t)} \qquad (9)$$

Where i is the number of the task, n, the quantity of tasks and TA_i^r, the time available recalculated by the operator.

This expression was tested with the supervision task of about 100 discrete variables and compared with values of workload assessed by the MCH Scale. The gravity of the situation used is drawn Fig. 3. Finally a new experimental campaign mixed continuous and discrete tasks in a car drivers context (Berger et al., 1989; Riera et al., 1990). The result of these experiments showed a good sensitivity of the LAMIH's method.

This method can be used on-line and is not obtrusive. The workload evolution is sampled with a shorter sampling period than SWAT and is assimilated to instantaneous value of workload due to its short sampling period. Another practical issue is the possibility to take into account, the mono-task or multi-task context which is excluded from the other methods previously presented. However the time load is the only one load generator taken into account by the method. In theory the gravity function G(t) represents the functional demand of the task. But in practice this gravity function is difficult to determine *a priori*.

2.4 Comparison of the workload assessment method

Table 1 compares the four workload assessment methods, according to the different criteria given above.

The LAMIH's method is the most relevant method according to the criteria presented in Table 1. and particularly, the unique method which allows a detailed task analysis in a multi task context. Unfortunately, this method takes into account temporal load only.

Table 1. Comparison of four workload assessment methods.

	TLX	SWAT	Sheridan	Millot
Load generators	Temporal demand, Mental demand, Physical, demand Own performance, Effort Frustration (or stress)	Time load (temporal pressure), Mental load (functional demands) Psychological stress load	Temporal demand	Temporal demand
S	✓	✓		✓
O			✓	✓
U	✓	✓	✓	✓
On-line measure	Measure at the end of the task	Sampling period around 5 minute	Measure at the end of the task	Sampling period around 1 second
Multitask context				✓

Legend: S, sensitivity; O, obtrusiveness; U, usability.

The next section presents a way to improve the LAMIH's method, in order to develop a workload assessment method which reaches criteria presented in this paper.

3. PROPOSAL OF A WORKLOAD ASSESSMENT METHOD DEVELOPMENT

Our objective is to develop of a general workload assessment method which can be used on-line and in the multi-task context; and is sensitive, non obtrusive and easily usable (mathematical expressions).

The idea is to take different load generators into account, allowing assessment on-line with a sampling period around 1 second and without disturbing for the operator.

A selection of the generators of load which influence workload must be made. For example in TLX, the own performance of the operator is used as a generator of load. But this term, performance, may be ambiguous. According to the operator, the same performance can be observed but the resources supplied by operators can be different regarding their various expertise levels for instance. Therefore one first proposal is to reuse the same generator of charge as SWAT, *i.e. temporal* load, *functional* load related to the task and *stress*, as they match the workload definition given by Sperandio.

Temporal load can be easily assessed as it was described by the LAMIH's method. Stress is partly due to time load. More the time available decreases, more the stress of the operator increases. If the functional demand of the task is high due to a risk of danger for example, the operator is stressed also. A second proposal is to consider the time load and the functional load only and find an expression estimating the correlation of the two parameters with stress.

Furthermore a study must be performed in order to find an estimation of the functional load by a mathematical expression, as made in the LAMIH's method.

As proposed by Wierville and Casali, a task is defined by different natures of activities, perception, mental, communication and psycho-motor. These activities imply different resource supplies from the operator. For example, pull a glass of water on the grass is less resources demanding than pull the same glass of water on the dehydrated plant. The first task demands only psychomotor activity. The second one demands psychomotor activity, but also a mental activity in order to detect the plant which really needs water. Therefore it is possible to give a degree of the functional load, low, medium or high (by comparison with SWAT), by a detailed task analysis. This way of functional load estimation is moreover linked to the task, which allows a better adaptation of the method to different operators, such as car driver, train driver, for example.

The first step of future study is to find different combinations of activities (perception, mental, communication or psychomotor) for an operator, performing a specific task; in order to find the way to create low, medium or high functional load contexts.

Secondly experiment will be performed, by using the previous study. in the task we study the temporal load remains constant and a variation of the functional load level will be simulated. Workload of the subjects will be assessed by three methods, SWAT, TLX and LAMIH's method taking into account initially only temporal load.

A comparison between the value obtained by SWAT and TLX with the proposed method, will allow obtaining discrepancies. This will be performed according to Fig. 4.

Some details have to be given concerning the comparison of the values obtained by SWAT and by TLX, due to the different sampling period of these methods. Indeed the nature of the sampling period characterizes the nature of the estimated workload. Therefore the results given by LAMIH's method at a sampling period of 1 second cannot directly be compared neither with SWAT nor with TLX. We must calculate the sampling Wl_j with the same period ΔT (for instance 5 minutes) as used in SWAT. Finally according to expression (8), to be compared with TLX, the LAMIH's results must be integrated all over the task duration.

When these particularities will be taken into account, we expect the respective discrepancies between SWAT and TLX and the LAMIH's method will represent the functional load of the task due to the specific activities performed by the operator during the experiment. From the discrepancies collected during experiment a data analysis should allow finding a formulation representing the

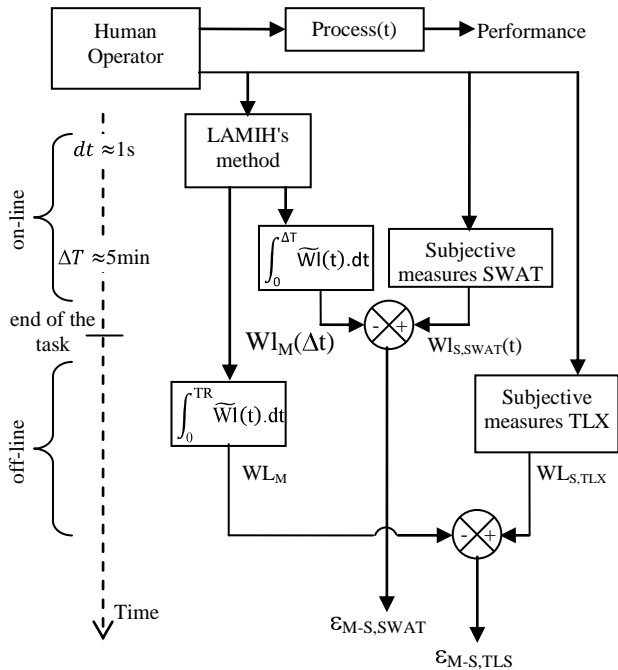functional load of a task due to the activities which it implies.

Fig. 4. Experimental set-up for functional load assessment.

Where dt and ΔT are the respective sampling period.

4. CONCLUSION

Four workload assessment methods were analysed according to different criteria: the generators of charge which are taken into account, a formulation of the sensitivity, the obtrusiveness, the usability and the sampling period of measurement.

Among the four workload methods, two allow a short sampling period, SWAT and the LAMIH's method.

By analysing the LAMIH's method, the need to take into account the multi-task context is another criterion of workload assessment method which has been highlighted. The LAMIH's method is the only one which proposes a solution for evaluating in multi-task contexts. This method will be the base of the development of our new workload assessment method.

However the LAMIH's method, contrary to the SWAT and the NASA-TLX methods, evaluates the temporal load only. An improvement of the LAMIH's method is then suggested by including functional load and stress.

Stress is a double manifestation of temporal load and functional load. The interest is taken in the temporal load and functional load only.

The temporal demand is assessed on-line with a sampling period around 1 second and without disturbing for the operator, as shown by the Millot's work. The issue is the determination of the functional load expression which has the same characteristics.

Firstly, a list of different functional load level (low, medium and high) has to be classified as function of the activities performed in a specific task. Task analysis is the background of this study.

The operator will be placed in different experimental context with the same temporal load but with increasing functional loads. A comparison will be made between SWAT and TLX and the LAMIH's workload estimation; calculated at the relevant sampling periods. A data analysis following the experiments expects to find a formulation of the functional load.

6. ACKNOWLEDGEMENT

7. REFERENCES

Berger, T., Riera, B., Angue J.C and Millot, P. (1989). Real time assessment of human operator workload placed in a simulated multitask context of a monitoring and supervising process. *ESA Workshop "Human factors engineering: a task-oriented approach*, pp. 13

Casali, J.G. and Wierwille, W.W. (1983): A comparison of rating scale, secondary-task, physiological, and primary-task workload estimation techniques in a simulated flight task emphasizing communications load. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 25, 6, pp. 623-641

Debernard, S. and Vanderhaegen, F. and Debbache, N. and Millot, P. (1990). Dynamic task allocation between controller and AI systems in air-traffic control. *9th European Annual Conference on Human Decision Making and Manual Control*.

Hart, S., & Staveland, L. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Human mental workload,* 1, pp. 139-183.

Hill, S.G., Iavecchia, H.P., Byers, J.C., Bittner, A.C., Zaklad, A.L. & Christ, R.E. (1992). Comparison of four subjective workload rating scales. *Human Factors, 34*, pp. 429-439

Millot, P. (1988): *Supervision des procédés automatisés et ergonomie*. Hermes, Paris

Millot, P. and Kamoun, A. (1988). An implicit method for dynamic task allocation between man and computer in supervision posts of automated processes, In *3. FAC/IFIP/IEA/IFORS Conference on Analysis*, pp. 77-82

Mitchell,L., Pickup,L., Norris, S. and Watts, D. (2005). Assessment of New Workload Tools for the Rail Industry. In John Wilson and Beverley Norris, University of Nottingham, UK, Theresa Clarke, Network Rail, UK and Ann Mills, Rail Safety and Standards Board, UK, *Rail human factors: supporting the integrated railway*, pp.330-340. Ashgate Publishing, Aldershot

O'Donnell, R. and Eggemeier, F. (1986): Workload assessment methodology. *Handbook of perception and human performance*, 2, pp. 42-1

Reid, G., Potter, S., & Bressler, J. (1987). *Subjective Workload Assessment Technique (SWAT): A user's guide.* Wright-Patterson AFB, OH: Harry G. Armstrong Aerospace Medical Research Laboratory.

Riera, B., Simon, P., Vandwale P., Millot, P. and Angue, J. C. (1990) : Activity modelling and workload assessment of a car driver. *Ninth European Annual Conference on Human Decision Making and Manual Control*, pp.15

Rubio,S., Díaz, E., Martín, J. and Puente, J. M. (2004): Evaluation of Subjective Mental Workload: A Comparison of SWAT, NASA-TLX, and Workload Profile Methods. *Applied Psychology: An international review*, 53, 1, pp. 61–86

Sheridan, T. B., & Stassen, H. G. (1979). Definitions, Models and Measures of human Workload. In, *Mental Workload: Its Theory and measurement*, pp. 219-233. NATO Scientific Affairs, New York and London.

Sperandio, J. C. (1978). The regulation of working methods as a function of workload among air traffic controllers. *Ergonomics*., 21, pp. 195-202

Valot, C., Grau, J., Romans, M., Ferret, S., Gervais, T. (1997). A method of designing ergonomics for activity dynamics: an aeronautical example. *Engineering Psychology and Cognitive Ergonomics: Transportation systems,* 13, p. 121.

Wierwille, W.W. and Rahimi, M. and Casali, J.G. (1985): Evaluation of 16 measures of mental workload using a simulated flight task emphasizing mediational activity. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 27, 5, pp. 489-502