Three Quantities for Error Evaluation in Safety Critical Human Computer Interface [1]

Fabio A. Schreiber
Dipartimento di Elettronica e Informazione - Politecnico di Milano
Piazza Leonardo da Vinci 32, 20133 Milano, Italy
e-mail: schreibe@elet.polimi.it

Abstract: *Three quantities are introduced in order to evaluate the impact of the operator's errors on the total dependability figures in safety critical systems: usability - expressed as a function of the workload on the operator -, carelessness, and the difficulty of learning, modeled as a function of the time needed to use the interface without making mistakes. The proposal points out the need for experimental work to determine the values of the numerical coefficients.*

Keywords: *carelessness, human computer interface, learning difficulty, MTBF, safety critical systems, usability*

1. Introduction

The embedding of computers in control systems for safety critical applications requires an in depth evaluation of their dependability features. Figures established by standardization committees are orders of magnitude higher than what can be proved today for software, which is an important part of the system. However both certification authorities and insurance companies need quantitative tools for assessing the dependability of safety critical systems.

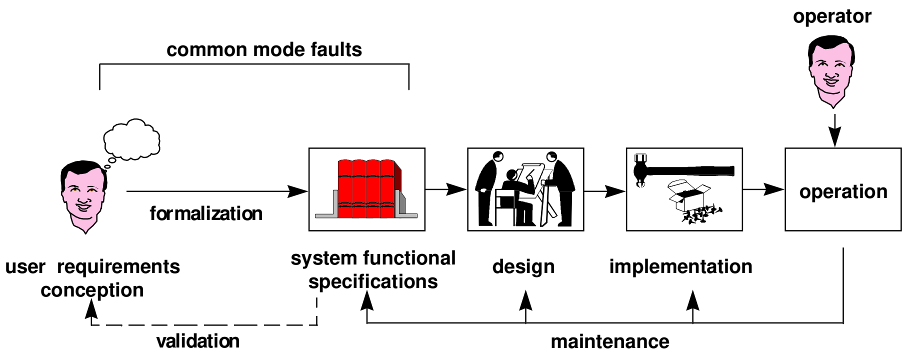

Fig. 1 - The design process

Aim of the research, this work is part of, is to understand if and how there is any chance of bridging this gap.

---

[1] Work partially supported by CESTIA-CNR and by MURST

Human beings enter in the picture since they are involved in at least two very important functions: at the very beginning (the user requirements conception) and at the end (the operator-system interface, often called Human-Computer Interface *HCI*) of the design process (Fig. 1); even if the entire design and implementation process of the system could be automatically verified and tested by formal methods, so obtaining the certainty of its correctness, the aforementioned functions would still be a source of human errors [MoSc97]. Errors in the requirements analysis phase, which are independent of the physical nature (hardware or software) of the components, are the most dangerous since they affect - as common mode faults - all the design process. However they are dealt with by means of inspections and reviews of the functional specifications, possibly produced with formal methods and languages, and international norms and standards try to quantify their impact on the overall risk [FeNe98, CENE1, CENE2]. On the other hand, the operator's error directly affects the reliability of the system just like a software bug or a hardware failure.

While some work has been made on the psychological side of the conception and of the design process [Cock94, Curt88, Holl93, Holl98, Rasm86], most researchers are dealing with the ergonomy of HCI in order to ease the job of the operator and to minimize the probability of errors in his/her work. However ergonomic studies often end with a list of guidelines and recommendations, but very few of them give a quantitative evaluation of the operator's reliability [Dhil87, GeGr96, HuBl96]. The goal of other studies, based on very detailed models, is the discovery of causal links leading to the operator's errors, in order to avoid possible future accidents [CaHo93, LeTu93].

Models entailing the use of quantitative data have been criticized on the ground that it is very difficult to obtain meaningful data [Holl98]; however we strongly feel that engineering evaluations must be rooted in quantitative analysis. In this paper three quantities constituting the core of an abstract quantitative model of the HCI are proposed. Their use is twofold: first they guide experimental researchers in collecting data about the performance of the couple {operator, HCI}, data which are used to fit the model to the experiments; after that, they can be used to evaluate the HCI. Suggestions for experimental procedures can be found at [HFE99]. Therefore, the goal of this paper is to identify and define measurable quantities *without entering into the details of the system's control features*, but meaningful about the interactions of the human operator with the system and its environment.

2. The model

A detailed analysis of the human behavior with respect to the probability that an operator makes an error is presented in [Holl93, Holl98, Rasm86]; we can summarize these studies in two broad categories:

- subjective (or psychological) - errors depend on the operator's perception and skills;
- objective (or engineering) - errors depend on the engineering aspects of the design.

Therefore a model for evaluating a HCI should account for both of them [GeGr96]. In figure 2 a framework is shown in which both aspects are present:

- the definition of the operator's functions, their design and implementation represent the engineering side of the problem and the *usability* and *carelessness* are the evaluation parameters;
- the learning and training procedures depend both on the design of each function and on the skill of the single operator [Kont96] and the *difficulty of learning* is the evaluation parameter.

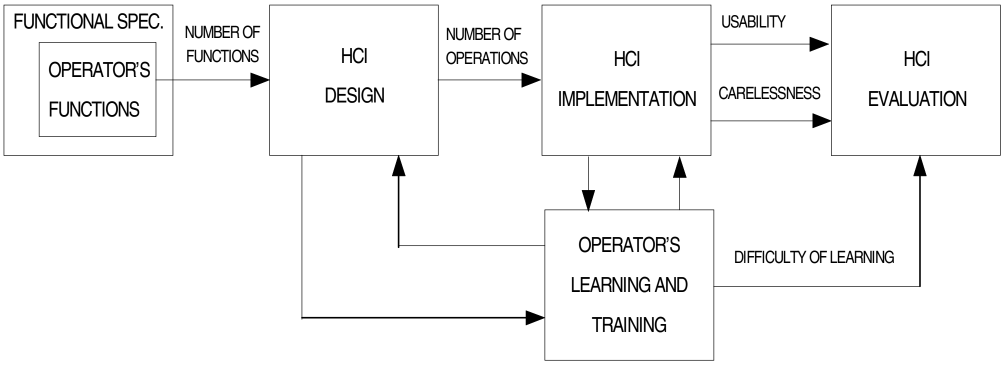

Fig. 2 - Evaluation framework of a HCI

A first point in designing a HCI is the partition of tasks between the operator and the machine; this entails a philosophical (but with very effective practical consequences) debate on how much of real *control* functions and how much of simple *monitoring* functions should a human being be assigned, the trade-off being fallibility vs. flexibility [Wien87]. In any case, in the functional specification phase, a *number of functions* (*F*) are identified as requirements for the HCI. In the design phase (*Task analysis*) each function is exploded into a sequence (or into a set) of *operations* (or *events*) to be performed in order to accomplish it; the dynamics of the operator-system interaction is also to be considered in the modeling [CaCV93, CaHo93]. Due to the output of the design phase, errors can be made by the operator *in the sequence* of the operations, *in their timing*, in taking the *wrong decisions*. In the implementation phase the *technological* factors are considered; the design and implementation phases are interactively based on engineering as well as on ergonomic and learning requirements.

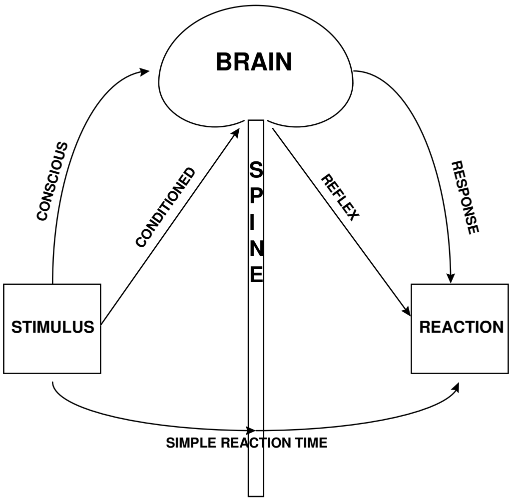

Fig. 3 - The stimulus-reaction chain

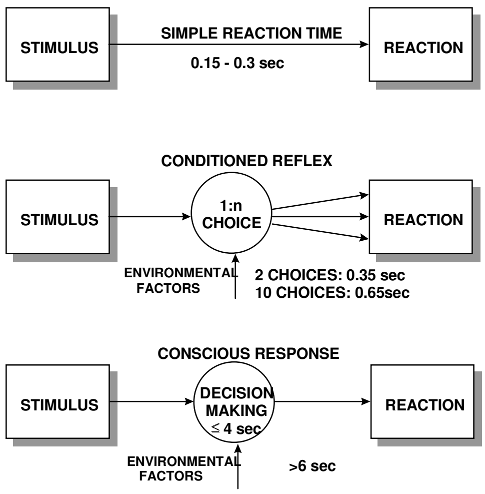

Fig. 4 - Typical reaction times

A very important issue to be considered is the type of stimulus-reaction chain required by a certain operation since it determines the workload on the operator. Figure 3 shows the three basic mechanisms: *simple reaction*, when the instinctive reaction is immediately triggered by the stimulus through the spine; *conditioned reflex*, when a choice has to be made among a (small) number of predefined responses; *conscious response*, when the brain has to explicitly process the input to obtain a decision; in [Holl93] an extensive analysis of cognition and of cognitive processes has been made; in [Rasm86] a "ladder abstraction" is described for the sequence of mental activities between initiation of response and the manual action .

In figure 4 some typical mean response times are shown for the three mechanisms; these times can greatly change on the basis of different environmental factors such as: age and sex of the operator, type of sensorial stimulus (hearing, touch, sight, pain, etc.), signal intensity, and many others [Wood92]; moreover, reaction times can be reduced by training and practice [Rasm86].

In modeling the human-computer interaction for control and command systems and the error probability associated with it one must consider two different environmental situations:

- normal routine, when the operator's attention is focused on planned and mostly independent (even if sometimes overlapping) events and actions; in this case simple reaction or conditioned reflex responses are mostly used.
- accident management, when the operator, beside being stressed by emotional pressure, has often to deal with unforeseen situations and sequences of possibly overlapping dependent stimuli; in this case conditioned reflex and conscious reaction are the most common behaviors.

3. The measurable quantities

While being conscious of the great complexity of modeling the full process of human (possibly erroneous) response to external stimuli, we try to formulate a model based on measurable elements which could be actually quantified. In the following we introduce three major parameters as the basis of the model: *usability*, *carelessness*, and the *learning difficulty* of how to operate with the HCI under evaluation.

Usability (*U*): a HCI is "usable" when one can *easily learn* exactly *when* and *what* to do in order to control the system and this is *within the range of his/her skills*. However many authors [AmDe92, Kont96] relate and restrict the usability of a system to the workload W the operator(s) must support to monitor/control it. Usability is also affected by the overhead introduced by the HCI with respect to a virtual direct chain of the type "I think it - I do it", analogous to the simple reaction chain, this aspect being heavily dependent also on the technological choices.

In the following definitions we separate the learning (subjective) from the workload (objective) aspect and we do not take into account the technological aspects emerging from the implementation phase, but only focus on the manual/decisional issues; further work should include the technological aspects as well. Let us define:

$$U = \rho_s - \rho_o \tag{1}$$

where $\rho_s \leq 1$ represents the operator's saturation workload, i.e. the maximum stress he/she can afford within safety limits, which should be experimentally measured under different environmental conditions, and

$$\rho_o = \frac{t_{busy}}{T}$$

represents the *operator's utilization factor*, i.e. the fraction of a time interval T in which the operator's attention/action is required. Therefore a zero or negative value for U would show an unbearable workload for the operator.

In this way, *usability results in the distance of the operator's utilization factor from his/her saturation workload.*

To evaluate ρ_o, let $E=\{e_1, ..., e_m\}$ be a set of events each requesting the operator's attention for a time τ_i $(1 \leq i \leq m)$ and let $n_1, ..., n_m$ be the number of times each event happens in T (n_i/T being the *frequency* of event e_i). Then

$$\rho_o = \frac{\sum_{i=1,m} n_i \cdot \tau_i}{T}$$

However this simple formula holds only if events e_i are all independent and sequential. In case of overlapping independent events $\{e_i, e_j\}$ we can suppose that a *time overhead* τ_{ij} (which is zero if the events are sequential) is required of the operator to manage the couple, which adds to the times required by the single events. Therefore we should add a term of the form

$$\frac{n_i}{T} \cdot \frac{n_j}{T} \cdot \tau_{ij}$$

so obtaining

$$\rho_o = \frac{1}{T} \sum_{i=1,m} n_i \cdot \tau_i + \frac{1}{T^2} \sum_{\substack{i=1,m \\ j=1,m \\ i \neq j}} n_i \cdot n_j \cdot \tau_{ij} \qquad (2)$$

This assumption amounts to considering the human operator as a time sharing system where the τ_{ij} represent the context switching times which add to the task processing times.

We could generalize expression (2) to consider the overhead associated with k independent overlapping events by adding $\binom{m}{k}$ terms, some of them possibly null, obtained in a similar way as in (2), but we leave it as a mathematical exercise to the reader. The maximum value of k has been argued to be "seven plus or minus two" due to the short term buffering limitations of the human memory [Mill56].

The definition in (2) allows a very fine tuning of the usability parameters; however more approximated models can group the events/actions into broader equivalence classes with the criterion of having similar reaction times.

It should be noted that, even if *the structure* of this abstract formulation is irrespective of the application environment, this is taken into account through *the values* of the various coefficients.

A further step is to consider that often events depend on each other; in this case the frequencies products in (2) must be substituted with the conditional probabilities $p\{e_i|e_j\}= p\{e_ie_j\}/p\{e_j\}$.

Carelessness (*C*): from the analysis of a number of accidents in flight as well as in industrial control systems, it is well known that lowering too much the workload on the operator can be a potential cause of errors owing to the possible inadequacy of his/her response time and of the quality of the decision in case of emergency [Kont96]; experimental data show a difference of up to 20% in reaction times between expected and unexpected events [Wood92]. The inadequacy may arise from inadvertence, frustration, lack of skill or of training; therefore it should be important to quantify the ability of the HCI of keeping the operator alert and of avoiding safety measures elusion (avoidance of default values, etc.). The technological choices can greatly influence this parameter as, for instance, in the case of multi mode displays and command consoles; in an extensive research on the perceived human factors problems of flight deck automation [Funk96] 3% of the citations claimed that *mode awareness may be lacking* and this may lead to reduced situation awareness and errors, while 0.7% pointed out that pilots can inadvertently select the wrong automation mode.

It is very difficult to give an analytic expression for *C*; it should account both for the real workload and for the workload possibly added to keep the operator's attention high. At the moment we can only conjecture *it correlates to the inverse of the workload* ρ_o.

$$C = g\left(\frac{1}{\rho_o}\right)$$

In any case, calling ρ_C the minimum workload needed for keeping the operator's attention awake, we can say that *the operator's utilization factor should satisfy the constraints* (3)

$$\rho_C \le \rho_o \le \rho_s \tag{3}$$

Also ρ_C should be experimentally evaluated under different environmental conditions.

Learning difficulty (*L*): let us suppose the harder is the job the longer it takes to learn how to do it correctly. We define $t(f_i)$ to be the *learning time* for function f_i (i.e. the time needed to learn how to perform f_i); then we can define *L* as in (4), that is *the learning difficulty of a HCI is represented by the average learning time of its functions.*

$$L = \frac{1}{F}\sum_{i=1}^{F} t(f_i) \tag{4}$$

However we must define what the learning time is. To this purpose, let τ_{ij} be the time from the beginning of the training session to the most recent error incurred by a trainee operator o_j on function f_i. We can say that o_j *learned* f_i *if he/she did not make* *further errors over a time period* $t_j(f_i)=\gamma*\tau_{ij}$. This definition closely looks like that of a transient phenomenon in a physical system, where usually γ has a value of 4 or 5; however in physical systems the transient is represented by a continuous function and τ represents its time constant (fig. 5a), while the typical learning curve is discontinuous. Therefore our definition of τ_{ij} is conservative and it does not pose any constraint on the form of the learning function, as it is shown in figure 5b.

Therefore the learning time L_i for function f_i can be taken as the average performance over a number O of operators (5)

$$L_i = t(f_i) = \frac{1}{O}\sum_{j=1}^{O} t_j(f_i) \tag{5}$$

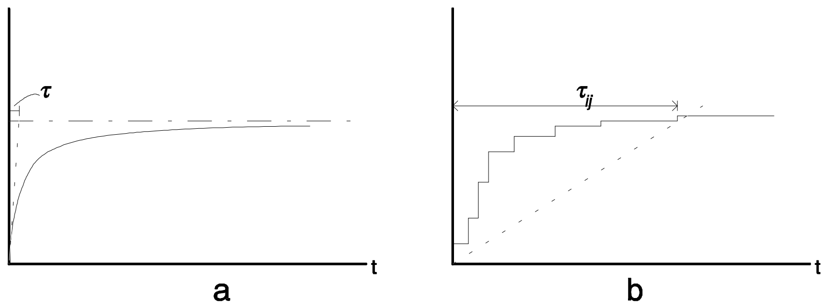

Fig. 5 - The learning function

The learning difficulty then becomes

$$L = \frac{1}{F}\sum_{i=1}^{F} L_i = \frac{1}{F*O}\sum_{i=1}^{F}\sum_{j=1}^{O} t_j(f_i) \tag{6}$$

We can also define a derived measure which, even if not directly included in the HCI model, can be useful in evaluating the operators performance. Let us call L_j the learning time of operator o_j averaged over all of the functions of the HCI, as in (7);

$$L_j = \frac{1}{F}\sum_{i=1}^{F} t_j(f_i) \tag{7}$$

then we can define the *skill* S_j of operator o_j the ratio of his/her learning time to the learning difficulty of the HCI.

$$S_j = \frac{L_j}{L} \tag{8}$$

Obviously this is a very rough averaged index, which could only be used in mass screening; a much finer analysis of operators skills can be found in [DiWa96].

4. Future work

The formulas introduced in the previous section give some hints on the way to define and measure some performance/reliability parameters of a HCI; however, up to now, they have little experimental support and further work is needed between this abstract model and its practical effects. A substantial amount of experiments must be made before they can be used in actual evaluation procedures; in particular:

- the saturation workload ρ_s, and the minimum attention workload ρ_C, which may change from different application domains and environmental situations, must be the result of statistical analysis on historical data and of field measures;
- the response times τ_i and the overheads τ_{ij} in the usability definition (2) could be inferred from task analysis and from general data on reaction times [Wood92]. However, accounting for the dynamics of the operator-system interaction, still requires experimental activity;
- the coefficient γ and the learning time constants τ_{ij} embedded in (6) must be experimentally measured; while γ should be mostly invariant with respect to the different applications, the τ_{ij} must be measured for each function of the HCI.

Probably some preliminary experimental work can be based on simulation, however field test should be the main source of significant data. The reward for the expensive empirical activity, that in any case takes a large part of this kind of studies [AmDe92, HuBl96], should be the availability of meaningful numerical values, homogeneous with those coming from other classes of physical components of the system, obtained without "throwing the dices".

5. Conclusions

The final goal of this work is the evaluation of a Mean Time Between Errors (*MTBE*) due to the HCI, homogeneous with the *MTBF* (Mean Time Between Failures) arising from the analysis of the other components (hardware and software) of the system, which could be used in reliability analysis methodologies such as FMEA (Failure Modes and Effects Analysis) - a "bottom-up" approach to the analysis of system reliability which requires the listing of all the system's subsystems and parts, their failure modes, and the failure rate of each mode - or FT (Fault Tree analysis) - a "top-down" approach which, leaving from an "undesirable event" taken as the root, builds a tree of all the possible events which could be the cause of the root event, connecting them by means of logic gates (AND, OR, …), and evaluating their occurrence probability [Dhil87].

At the moment we can only conjecture that *MTBE* (9) is directly correlated to Usability and inversely correlated to the number of Functions to be performed, to the Carelessness, and to the Learning time; however an analytical expression is not for now and a strong experimental effort is needed here too to fit an actual function to field data.

$$MTBE = h\left(U, \frac{1}{F}, \frac{1}{C}, \frac{1}{L}\right) \tag{9}$$

In the previous sections we discussed the three major components of *MTBE* with the goal of estimating the probability that a human operator makes an error owing to stress, inadvertence or poor training. It results that Usability, expressed as a function of the workload on the operator, should be the main factor in determining the operator's reliability since, besides being the measurable stress cause, it also depends on the total number of functions performed by the HCI.

Furthermore, the Usability and the Carelessness values can guide the designer in the choice of the amount of "decision power" to be given to the automated system with respect to that of the human operator. Finally, being able to evaluate the learning difficulty of each command and control function, the designer is helped in determining the bottlenecks in the operator's training activity and in optimizing the HCI under this aspect.

References

AmDe92 Amalberti R., Deblon F. - *Cognitive modeling of fighter aircraft process control: a step toward an intelligent on-board assistance system* - Int. J. Man-Machine Studies, Vol. 36, pp. 639-671, 1992

CaCV93 Cacciabue P.C., Carpignano A., Vivalda C. - *A dynamic reliability technique for error assessment in Man-machine systems* - Int. J. Man-Machine Studies, Vol. 38, pp. 403-428, 1993

CaHo93 Cacciabue P.C., Hollnagel E. - *Human models in reliability and safety analysis of interactive systems* - Proceedings PSA 93, Clearwater Beach (Fl.), 1993

CENE1 CENELEC - EN 50128: Railway applications - Software for railway control and protection systems - July 1998

CENE2 CENELEC - ENV 50129: Railway applications - Safety related electronic systems for signalling - June 1997

Cock94 Cockram T. et Al. - *Human error in the software generation process* - in Technology and assessment of safety-critical systems, Redmill F., Anderson T. (eds.), Springer-Verlag, 1994

CuRo91 Cushman W.H., Rosenberg D.J. - *Human factors in product design* - Elsevier, 1991

Curt88 Curtis B. et Al. - *A field study of the software generation process for large systems* - CACM, Vol. 31, n. 11, pp. 1268-1287, 1988

Dhil87 Dhillon B.S. - *Human reliability in information systems* - in: Information systems failure analysis, Wise J.A., Debons A. (eds.), ASI Vol. 32, Springer-Verlag, 1987

DiWa96 Dillon A., Watson C. - *User analysis of HCI - the historical lessons from individual differences research* - Int. J. Human-Computer Studies, Vol. 45, pp. 619-637, 1996

FeNe98 Fenton N.E., Neil M. - *A Strategy for Improving Safety Related Software Engineering Standards* - IEEE Transactions on Software Engineering, Vol. 24, n. 11, pp. 1002-1013, 1998

Funk96 Funk K., Lyall B., Riley V. - *Perceived Human Factors Problems of Flightdeck Automation* - FAA Grant n°. 93-4-039, Phase 1 Final Report, http://www.engr.orst.edu/~funkk/Auto/autoprob.html

GeGr96 Gentner D.R., Grudin J. - *Design models for computer-human interfaces* - IEEE Computer, Vol. 29, n. 6, pp. 28-35, 1996

HFE99 http://www.cs.umd.edu/~zzi/UsabilityHome.html

Holl93 Hollnagel E. - *Human Reliability Analysis Context and Control* - Academic Press, 1993

Holl98 Hollnagel E. - *Cognitive Reliability and Error Analysis Method CREAM* - Elsevier 1998

HuBl96 Hubona G.S., Blanton J.E. - *Evaluating system design features* - Int. J. Human-Computer Studies, Vol. 44, pp. 93-118, 1996

Kont96 Kontogiannis T. - *Stress and operator decision making in coping with emergencies* - Int. J. Human-Computer Studies, Vol. 45, pp. 75-104, 1996

LeTu93 Leveson N.G., Turner C.S. - An Investigation of Therac-25 accidents - IEEE Computer, Vol. 26, n. 7, pp. 18-41, 1993

Mill56 Miller G.A. - *The Magical Number Seven Plus or Minus Two: Some Limits on Our Capacity for Processing Information* - Psychological Review, Vol. 63, n. 2, pp. 81-96, 1956

MoSc97 Morzenti A., Schreiber F.A. - *Analysis of techniques in the assessment of ultra-reliable computerized systems* - ACM Software Engineering Notes, Vol. 22, n. 2, pp. 87-91,1997

Rasm86 Rasmussen J. - *Information Processing and Human Machine Interaction: an Approach to Cognitive Engineering* - North Holland, 1986

Wien87 Wiener E.L. - *Fallible humans and vulnerable systems: lessons learned from aviation* - in: Information systems failure analysis, Wise J.A., Debons A. (eds.), ASI Vol. 32, Springer-Verlag, 1987

Wood92 Woodson W.E., Tillman B., Tillman P. - *Human Factors Design Handbook* - McGraw-Hill, 1992