Proceedings of the 17th World Congress
The International Federation of Automatic Control
Seoul, Korea, July 6-11, 2008

Toward Cooperative and Human Error-Tolerant Systems

P. Millot and F. Vanderhaegen

Laboratoire d'Automatique, de Mécanique et d'Informatique Industrielles et Humaines University of Valenciennes – LAMIH - UMR 8530 CNRS
Le Mont Houy - 59313 Valenciennes Cedex 9 - FRANCE
Email: patrick.millot@univ-valenciennes.fr , frederic.vanderhaegen@univ-valenciennes.fr

Abstract: This paper focuses on parameters related to the human machine cooperation and erroneous human behaviour affecting the system performance and safety. The concept of cooperation is presented through three prerequisites: the Know-How related to competences, Know-How-to-Cooperate related to coordination between activities and the Need-to-Cooperate to justify the activities of cooperation. It is extended to take into account normal and erroneous human behaviour. Such implementation is based on both human engineering and cognitive control principles. Examples in air traffic control illustrate these concepts for prevention support. Work perspective focuses on the integration of the correction and the containment processes of human errors.

1. INTRODUCTION

The domains of application for this research include large industrial plants or transportation networks in which human activities mainly involve decision-making: monitoring and fault detection, fault anticipation, diagnosis and prognosis, as well as fault prevention and recovery. The objectives combine the human-machine system performances (production quantity and quality) as well as the global system safety. In this context, human operators may have a double role: (1) a negative role as they may perform unsafe or erroneous actions on the process, (2) a positive role as they can detect, prevent or recover an unsafe process behavior due to another operator or to automated decision makers.

Two approaches to these questions are combined in a pluridisciplinary research way : (1) human engineering which aims at designing dedicated assistance tools for human operators and at integrating them into human activities through a human machine cooperation, (2) cognitive psychology and ergonomics analysing the human activities, the need for such tools and their use.

This paper focuses on a human-centered approach for the design of cooperative and human error-tolerant system.

2. COOPERATIVE AND HUMAN ERROR-TOLERANT SYSTEM PRINCIPLE

The human error-tolerance principle consists in integrating three levels of risk management: the prevention level to avoid the occurrence of undesirable event, the correction to recover erroneous actions and the containment to protect the human-machine system from the consequences of a danger such as an accident. Facing a potential hazardous event occurrence, prevention supports are required to control the occurrence of such events. If prevention supports fail, when erroneous actions are detectable, the correction supports aim at recovering them. When correction supports fail or when erroneous actions are undetectable, containment supports are required to control the consequences of such erroneous actions. System supports are technical or human abilities that aim at controlling a given process in order to avoid the occurrence of undesirable events, to recover or contain them. A metaphoric explanation of the genesis of human errors is adapted from Rasmussen (1997) and Polet, Vanderhaegen and Amalberti (2002). The "human operation state" is well adapted to the task when the human state remains in the centre of the area created by three operational factors, Fig. 1:

- A limit of acceptability of the global system performance demands which can be required by the human operator hierarchy.
- A limit of resources the human may involve in the task in order to satisfy these demands.
- A limit of safety of the global system.

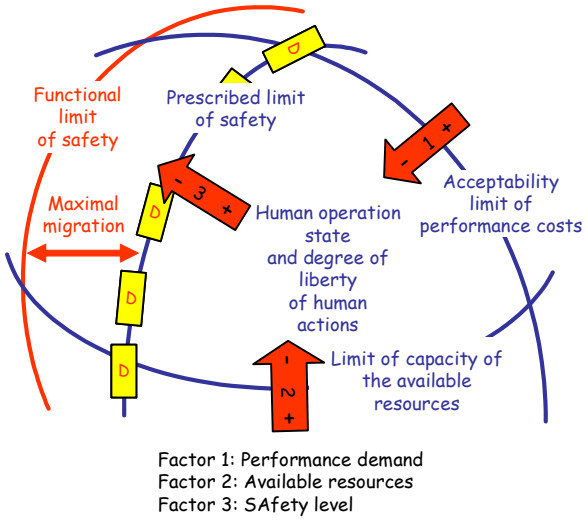

Fig. 1. Explanation of human error genesis, adapted from (Rasmussen, 1997; Polet et al., 2002)

When the performance demands increase and/or the human resources decrease, the "human operation state" migrates over the safety limit and leads to hazardous behaviours of the human operators. This overlap of the safety limit does not always result in an erroneous action and in an accident. Nevertheless, that increases the pressure on the safety limit. Then, this limit migrates toward an ultimate limit really corresponding to the maximum of the human abilities to satisfy the performance demands.

Therefore, a cooperative and human error-tolerant system could result in two main ways, Fig. 2:

- By introducing tools fore assisting the human operators and then increasing the global human-assistance tools resources.
- By controlling the overlappings of the safety limit through a human error-tolerant system.

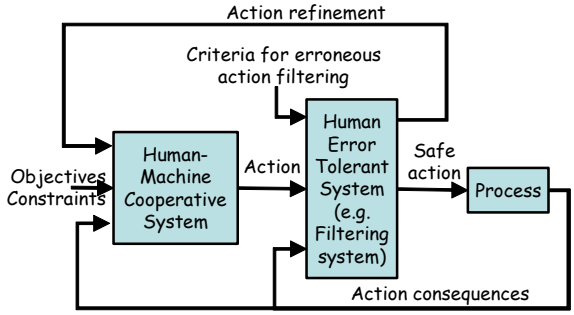

Fig. 2. Cooperative and human error-tolerant system principle

The following part describes first the human-machine cooperation principles. An implementation example in the Air Traffic Control (ATC) domain has shown their feasibility and the resulting increasing of performance, but human errors remain. Then, the following part is dedicated to human erroneous action characteristics. The last part proposes a joint approach combining both principles.

3. COOPERATIVE SYSTEM TO INCREASE RESOURCES

A cooperative system requires competences to solve problems, abilities to cooperate and abilities to identify the need to cooperate. Such abilities for a cooperative system can be implemented into several human-machine structures. The aim of this section is to draw the generic characteristics of a Decision Support System (DSS) without focusing on a dedicated application field.

3.1. Know-How-to-Cooperate and structures

A DSS is designed to assist Human Operators in order to facilitate their tasks and avoid failed performances. Both the DSS and the Human Operators are called agents. Agents can be modelled related to 3 classes of capabilities:

- the *Know How* (KH) for solving problems and performing tasks autonomously, including problem solving capabilities (knowledge, processing abilities) and communication capabilities for sharing information with the environment and other agents through sensors and control devices.
- the *Know-How-to Cooperate* (KHC) needed for Managing Interference (MI) between goals and for Facilitating other agents' Goals (FG) according to the definition of cooperation definition given in the next subsection (Millot and Hoc, 1997).
- a *Need-to-Cooperate* (NC) including :
  a) *the adequacy* of the agent's personal KH (in terms of knowledge and processing abilities) for the constraints required by the task
  b) *the abilities* to perform the task (the human agent's workload produced by the task, perceptual abilities, control capacities)
  c) *the Motivation-to-Cooperate* including the motivation to achieve the task, self-confidence, trust (Moray et al., 1995) and the confidence in the cooperation (Rajaonah et al., 2006).

In the field of cognitive psychology, Hoc (1996) and Millot and Hoc (1997) propose the following definition : *"two agents are cooperating if 1) each one strives towards goals and can interfere with the other,and 2) each agent tries to detect and process such interference to make the other's activities easier".*

From this definition, two classes of cooperation activities which constitute know-how-to-cooperate (KHC) can be derived (Millot and Lemoine, 1998):

- The first activity requires the ability to detect and Manage Interference between goals (*MI):* this interference can be positive (common goal, sub-goal ...) or negative (conflicts between goals, sub-goals ... or common resources to be shared).
- The second activity requires the ability to Facilitate the Goals of the other agents (*FG).*

Therefore the MI involves *coordination* ability, while the FG involves generous agent behavior. Such cooperative systems can be implemented following several possible structures to allocate them the required tasks to be achieved and to facilitate the interference management.

Defining the organization is one way to prevent or solve decisional conflicts between the agents, especially in human engineering where agents can be either humans or artificial DSS. Two generic structures of purely structural organization exist, called respectively vertical (i.e. hierarchical) and horizontal (i.e. heterarchical) (Millot, Taborin et al 89):

- In the vertical structure, an agent called AG1 is at the upper level of the hierarchy and is responsible of all the decisions. If necessary it can call upon the other agent called AG2 which will give advice. In Fig. 3-a, AG1 is a human operator and AG2 is a DSS.
- In the horizontal structure, both agents are at the same hierarchical level and can behave independently if their respective tasks are independent. Otherwise, they must manage the interference between their goals using their

MI and FG capabilities. In Fig 3-b, AG1 is a human operator and AG2 is a DSS. The MI activities are performed by a coordinator at the upper level called the task allocator control level that involves the KHC of each agent.

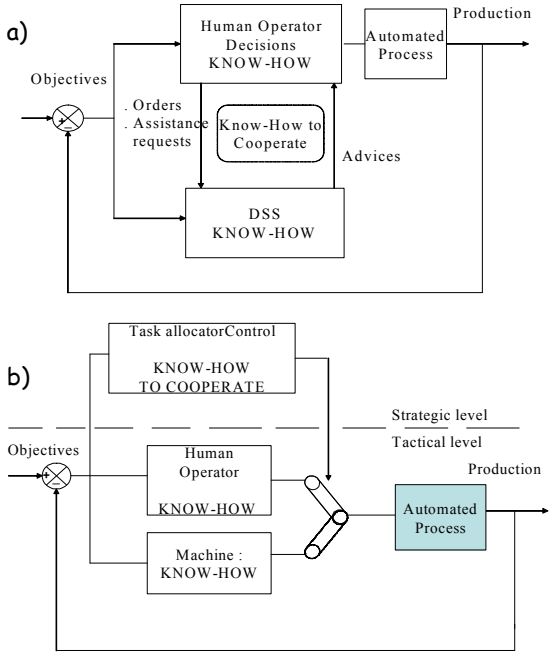

Fig. 3. Vertical (a) and horizontal (b) structures for human machine cooperation

Several combinations of these two structures are possible by breaking the task down into different abstraction levels (Rasmussen, 91) and assigning a dedicated structure to each of these levels. These cooperative structures are usually applied for studying normal system functioning without taking into account the possible erroneous action propagation.

3.2. Exemples of prevention related DSS

The application concerns the SPECTRA platform (French acronym for Experimental Task Allocation System for Air Traffic Control) developed at the University of Valenciennes to study the dynamic allocation of the air traffic control tasks between a human controller and DSS. SPECTRA experiments involved 15 human controllers (i.e. 9 experts and 6 inexperienced) in 54 conflicts between planes in the course of 3 experiments: one without using the DSS, and two experiments with the automated tool (the so-called explicit mode for a pre-emptive allocation managed by the human operator, the so-called implicit mode for a definitive allocation managed by the DSS), (Vanderhaegen *et al.*, 1994). The shared task is the supervision of conflicts between planes, including the conflict detection verifying if planes may transgress minimum separation norms, the conflict solving by sending an adequate order to planes in order to avoid conflicts, and the problem solving in order to verify if a conflict is over and to orient deviated planes to their initial way. The DSS is able to detect all conflicts but to solve and control only those between two planes. This is the DSS's KH.

The NC relates to the human workload assessment. In the explicit mode, the human controllers define this NC and manage the KHC controlling the task allocator. In the implicit mode, the DSS defines the NC and controls the task allocator regarding a task demand estimator based on the assessment of the task complexity.

Two performance indicators have been evaluated on-line:

- The first indicator is the plane kerosene consumption which relates the quality of the conflict solving. This indicator is better when the human controllers are aided by the DSS.
- The second indicator relates to the safety system evaluated by the number of conflicts solved by the team human controller-DSS in this situation of heavy traffic (i.e. 54 conflicts). The results show an increase of the number of conflicts solved when the human controllers are aided by the DSS regarding the situation when they control alone the traffic. The increase is higher in the implicit mode. A complementary level analyses human errors through two classes of events to be evaluated. The former class of events noted *e1* relates to the occurrence of the erroneous human behaviour and the latter class of events noted *e2* relates to their consequences, Table 1:
- The ratio *P(e1)* is the number of detected conflicts upon the total number of conflicts allocated to the human controllers.
- The ratio *P(e2/e1)* concerns the air traffic control safety for SPECTRA, i.e. the number of conflict correctly treated upon the detected ones.

Table 1. Results from SPECTRA experiments

	Without DSS	With DSS	
		Explicit mode	Implicit mode
P(e1)	0.89	0.97	0.97
P(e2/e1)	0.83	0.88	0.95
HR	0.73	0.85	0.92

The proposed DSS decreases the number of human erroneous actions in terms of human detection and system safety. The degree of liberty allocated in the explicit mode that is a pre-emptive mode (i.e. the human operators can modify their initial allocations) does not modify the detection abilities regarding the implicit mode that is a definitive mode (i.e., an allocation cannot be corrected). The implicit mode presents the best results in terms of human error prevention when detecting and solving conflicts between airplanes.

Another example concerns the AMANDA project (Automation and MAN-machine Delegation of Action). A DSS was developed to facilitate the interference management through a so-called Common Work Space (CWS). The KH of

the DSS takes into account anticipative behaviours. By taking the human air traffic controller strategies into account, the DSS is able to assess precise solutions and transmits the corresponding command to the plane's pilot (Pacaux-Lemoine and Debernard, 2002). These actions are jointly realised by the DSS and the human controller in order to optimise both the system performance and the system safety. In an initial experiment, the air traffic controllers frame of references were identified by coding the cognitive activities of the controllers (Guiost et al., 2003). The CWS plays a role similar to a black-board displaying the problems to be solved cooperatively: it is the support for the KHC and the NC. Results have shown that such a cooperative structure facilitates the interference management related to the anticipation of the conflicts between planes and increases the human activity quality.

Experiments of these organizations in air traffic control (Vanderhaegen et al., 1994; Millot and Lemoine, 1998) have then shown an increase of the human-machine performance and safety when the human is supported by a cooperative DSS. A second result is the need of a common frame of references between both agents (Pacaux-Lemoine and Debernard, 2002). A third result is an important reduction of the number of erroneous decisions and actions.

Nevertheless, despite the cooperative system, the human errors are reduced but still remain.

4. HUMAN ERRONEOUS ACTION CHARACTERISTICS

Unsafe or unoptimal human actions relate to intentional or unintentional behaviour, Fig. 4, (Reason, 1990):

- Slips are non-intentional and relate to skills and attention based failure
- Lapses are non-intentional and relate to skills and memory based failures
- Faults are intentional and relate to rules or knowledges based failures.
- Violations are intentional and relate to barrier removals.

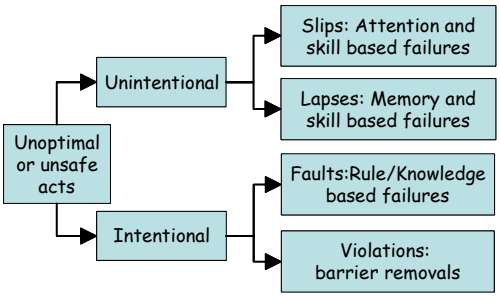

Fig. 4. Human error taxonomy and unsafe acts from (Reason, 1990).

Hollnagel's taxonomy (Hollnagel, 1998) describes such erroneous actions by different parameters called phenotypes that are caused by genotypes, e.g., individual, systemic or environmental causes. These paramaters concern the erroneous actions characteristics, Fig 5:

- Erroneous goal, e.g. the achievement of an action relates to a wrong action or a wrong objective.
- Erroneous sequence, e.g. the achievement of an action relates to an omission, an interruption, an inversion, a repetition, an intrusion.
- Erroneous duration, e.g. the processing time of an action is too large or too small.
- Erroneous time, e.g. the action is omitted or achieved too early or too late.
- Erroneous distance, e.g. the achievement of an action is too far away from the target or too close to the source.
- Erroneous speed, e.g. the action is realized too quickly or too slowly.
- Erroneous space, e.g. the direction, the movement or the orientation associated to an action is wrong.
- Erroneous intensity, e.g. the effort done to achieve an action is too high or too low.

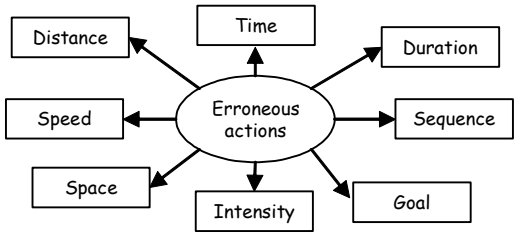

Fig. 5. Erroneous action taxonomy from (Hollnagel, 1998).

A given erroneous action can also be assessed in terms of consequences. The so-called Benefit-Cost-Deficit (BCD) approach developped in (Polet et al., 2002; Vanderhaegen, 2004) analyzed intentional or unintentional human erroneous actions using three distinct consequences on several evaluation criteria such as safety, service quality, workload or production quantity, Fig. 6:

- *B*: the possible benefits due to the occurrence of the erroneous action. The knowledge of the controlled system behaviour in terms of safety and performances may increase with the management process of the human errors.

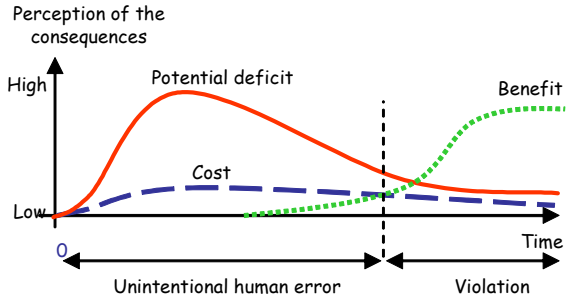

Fig. 6. Example of the BCD parameter evolution

- *C*: the acceptable costs due to the occurrence of the erroneous action. The human error occurrence may require additional physical or cognitive behaviours to recover or control the situation.

- *D*: the unacceptable possible deficit related to the occurrence of the erroneous action. A human error may lead to unacceptable and unrecoverable situations in case of failed action.

The BCD model is able to assess the benefits and the acceptable cost of a given action in case of successful human error control and the potential unacceptable deficits or dangers in case of failed human error control. Objectively, these BCD parameters may be combined with others ones in order to assess the utility level *U(s)* of an erroneous action *s* :

$$U(s) = p(s)[\alpha.B + \beta.C] + (1 - p(s))[\gamma.D] + \varepsilon \qquad (1)$$

B, *C* and *D* are the benefits, the costs and the potential deficits or dangers ponderated by α, β and γ respectively, occurring after the task achievement *s* for which a probability of success of the error control process is given by *p(s)*. Error assessment ε on all the parameters BCD can occur.

The BCD model was originally used to describe or predict barrier removal made by human operators on the field of industrial rotary press (Polet et al., 2003), of simulated train traffic control (Vanderhaegen, 2004), and of simulated car traffic control (Chaali-Djelassi and Vanderhaegen, 2006).

5. TOWARD COOPERATIVE AND HUMAN ERROR-TOLERANT SYSTEM

The future architecture integrating both the cooperative system and the human error-tolerant system principles to manage the prevention, the correction and the containment processes of the system danger generated by human errors is proposed on Figure 7. The human error-tolerant process is controlled by making both HO and DSS cooperate and by respecting the system constraints and demands. The identification of the human error has to focus on the taxonomies presented section 4 and the control of the human errors is done in a cooperative ways such as those presented section 3. The BCD assessment process related to erroneous actions will be used to refine the human abilities to perceive, assess and make decisions in a preventive way. It will be then used to describe the consequences of an erroneous action, but also to recover or refine it, to prohibit its transmission and to contain its dangerous consequences.

Considering normal and erroneous human behaviors, cooperation related concepts have to be extended in order to consider possible erroneous KH, KHC or NC at each step of the prevention, correction and containment control processes. Moreover, such KH, KHC and NC may evolve dynamically for a given process. For example, the KH may increase and the erroneous KH may be corrected. Learning effects are then important and the cooperation process may take into account how an agent may learn from the normal or erroneous behaviors of the other agent.

The present attempts for answer use a context dependant approach consisting in defining a list of prohibited commands. Variants place barriers around the process, either to avoid erroneous actions as above or to avoid unexpected process behaviors (Polet et al., 2002). General more context free approaches could be based on more achieved models of the human errors. This needs progresses in cognitive sciences and artificial intelligence as well.

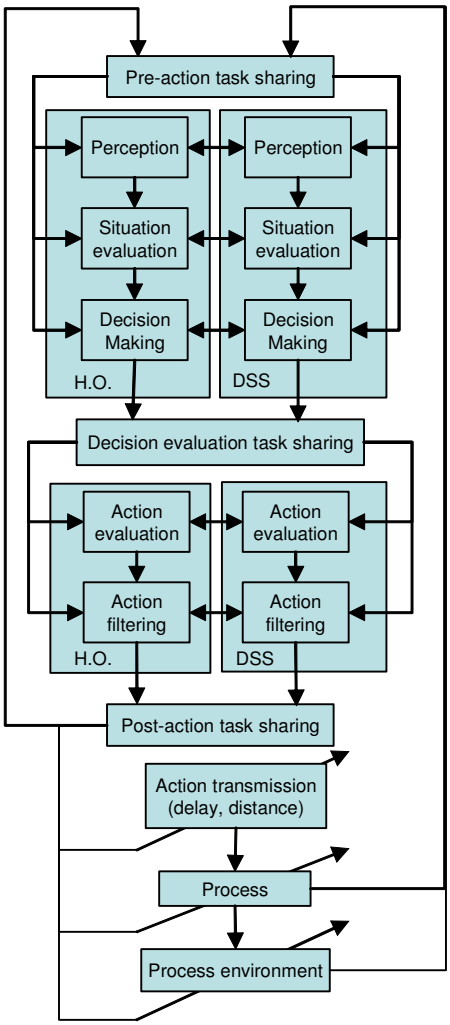

Fig. 7. Toward cooperative and human error-tolerant system structure

Numerous other problems emerge now, for instance linked to the need for understanding fixations (also called diabolic errors) (Van der Vlugt and Wieringa, 2003) or trying to understand why some operators cross or remove barriers (Polet et al., 2002).

Behind these conceptual frameworks, concrete aspects regarding implementation possibilities of such ideas must also be studied.

5. CONCLUSION

This paper has developed human error-tolerance and human-machine cooperation concepts in order to define cooperative and human error-tolerant system architecture and ability. The KH, KHC and NC abilities for a given human operator has to be implemented into a DSS in order to increase the capacitiy of the human-machine system to cooperate. The integration of human error models or taxonomies may facilitate the prevention, the recovery or the containment of erroneous human actions. A BCD model framework was detailed in

order to describe human intentional or non-intentional erroneous actions in terms of benefits, costs and potential deficits or dangers. It will be used to facilitate human activities and to control normal and abnormal situations.

A new framework for cooperative and human error-tolerant system was then proposed. It aims at (1) integrating DSS to provide human operators with assistance for making human tasks easier and then avoiding human error occurrence, (2) integrating a filter of erroneous actions in order to control the actions and to recover or to contain the erroneous ones.

ACKNOWLEDGEMENTS

This project is supported by european, national and regional (Nord - Pas de Calais region) fundings for the CISIT project (International Campus on Safety and Intermodality in Transportation).

REFERENCES

Chaali-Djelassi, A. and Vanderhaegen, F. (2006). Predication of violations in road transportation system. Proceedings of the 4th International Conference on Safety and Reliability, Krakow, May 30- June 02 2006, pp. 57-68.

Hoc, J.-M., (1996). *Supervision et contrôle de processus, la cognition en situation dynamique*. Presses Universitaires de Grenoble, 1996.

Hollnagel, E. (1998). Cognitive Reliability and Human Analysis Method CREAM. Oxford, Elsevier Science Ltd.

Millot P., Taborin V., Kamoun A., 89, « Two approaches for man-computer Cooperation in supervisory Tasks », 4th IFAC/IFIP/IFORS/IEA Symposium "Analysis Design and Evaluation of man-machine Systems, XiAn China september.

Millot P, Hoc JM. 97, "Human-Machine Cooperation: Metaphor or possible reality?" European Conference on Cognitive Sciences, ECCS'97, Manchester UK, april.

Millot P., Lemoine M.P (1998). "An attempt for generic concepts Toward Human-Machine Cooperation", IEEE SMC, San Diego, USA, october 12-14.

Moray N., Lee, Muir, 95. Trust and Human Intervention in automated Systems. In Hoc, Cacciabue, Hollnagel (Eds). *Expertise and Technology cognition and Human Computer Interaction*. Lawrence Erlbaum Publ., 1995.

Pacaux-Lemoine, M.-P., Debernard, D. (2002). Common work space for Human-Machine Cooperation in Air Traffic Control. *Control Engineering and Practice*, **10**, 571-576.

Polet, P., Vanderhaegen, F. and Amalberti, R. (2003). Modelling border-line tolerated conditions of use (BTCU) and associated risks. *Safety Science*, **41** (2003) 111-136.

Polet, P., Vanderhaegen, F. and Wieringa, P. (2002) Theory of safety related violation of system barriers. *Cognition Technology & Work*, **4**, 171-179.

Rajaonah, B., Tricot, N., Anceaux, F., Millot, P (2006). Role of intervening variables in driver-ACC cooperation. *International Journal of Human Computer Studies*, 2006, in press.

Rasmussen J. (1991). Modelling distributed decision making. In Rasmussen J., Brehmer B., and Leplat J. (Eds). *Distributed decision-making : cognitive models for cooperative work*, pp 111-142, John Willey and Sons, Chichester UK, 1991.

Rasmussen, J. (1997). « Risk management in a dynamic society: a modelling problem », Safety Science, vol. 27, n° 2/3, p. 183-213, 1997

Reason, J. (1990). Human Error. Cambridge University Press, Cambridge, UK.

Vanderhaegen, F. (2004). *The Benefit-Cost-Deficit (BCD) model for human analysis and control*. Proceedings of the 9th IFAC/IFORS/IEA symposium on Analysis, Design, and Evaluation of Human-Machine Systems, Atlanta, GA, USA, 7-9 September 2004.

Vanderhaegen, F., Crévits, I., Debernard, S. and P. Millot (1994). Human-Machine Cooperation: toward an Activity Regulation Assistance for Different Air Traffic Control Levels, *Int. J. on Human-Computer Interaction*, **6**, 1, 65-104.

Van der Vlugt, M., Wieringa, P.A. (2003). *Searching for ways to recover from fixation : proposal for a different view-point*. Proceedingd of the Conference on Cognitive Science Approach for Process Control CSAPC'03, Amsterdam, September 17-19, 2003.