Evaluation of human error probabilities based on classical HRA models: an application to railway systems

Siqi Qiu, Mohamed Sallak, Walter Schön, Zohra Cherfi-Boulanger

Evaluation of human error probabilities based on classical HRA models: an application to railway systems

Siqi Qiu, Mohamed Sallak, Walter Schön
Heudiasyc CNRS UMR 7253,
Université de Technologie de Compiègne, France
Email: {siqi.qiu, mohamed.sallak, walter.schon}@utc.fr

Zohra Cherfi-Boulanger
Roberval CNRS UMR 7337,
Université de Technologie de Compiègne, France
Email: zohra.cherfi-boulanger@utc.fr

***Abstract*—This paper presents an experimental protocol which aims to study human reliability in railway systems. The experiment is conducted on a railway traffic management system that places operators (experimental subjects) in simulated situations involving failures. Some classical HRA (Human Reliability Analysis) models are used to interpret the experimental results and to evaluate the probability of human error.**

***Keywords*—Human Reliability Analysis, Human error probability, Railway.**

I. Introduction

According to the statistics of the Federal Railroad Administration Office of Safety Analysis [1], human factors are the most significant cause of train accidents. In 2011, 36.35% of train accidents in the US were caused by human factors, 33.58% by track defects, 11.60% by equipment defects, 1.71% by signal defects, and 16.77% are ascribed to miscellaneous causes. These figures underline the need for human factors to be studied, with the aim of preventing or reducing the number of train accidents.

Human reliability refers to the reliability of humans in many fields, including the transport systems. Human reliability can be affected by many factors, especially the human errors. According to Spurgin [2], human errors are sometimes thought of as spontaneous errors made by individuals and crews; however, most errors are induced by the situation under which persons operator. Swain and Guttman [3] defined human error as any member of a set of human actions that exceeds some limit of acceptability.

In railway systems, human error has been defined as a behavior of the human operator which leads to accidents in railroad systems [4]. The normal operation of the railroad system depends on the activities of human beings and machines. Advances in science and technology have meant that mechanical reliability has been significantly improved. Human error is an increasingly significant factor in train accidents. Humans can deal with accidents and unusual situations, but they also make mistakes. Therefore, in order to evaluate the performance of a railway system, it is necessary to model human operators involved in the railway system.

In the literature, there exist a variety of HRA models. HRA models are used to evaluate the Human error probability (HEP) throughout the completion of a task. Spurgin [2] summarized three categories of existed HRA models according to their characteristics: task-related models, time-related models, and context-related models. A model is not developed to handle all the issues addressed in human reliability. Each model is developed only for certain issues. Thus, an appropriate HRA model should be chosen according to the characteristics of the research subject.

In this paper, an experimental protocol is developed to conduct an experiment on a railway platform, Route Control Centre System (RCCS), provided by Ansaldo STS. The main objective of the experiment is to assess the HEP of human operators. Several experimental subjects are invited to conduct the experiment under different conditions. The obtained experimental result is later analyzed by some classical Human Reliability Analysis (HRA) methods which estimate the HEP of each subject. Finally, we propose a discussion on the analytic results.

The reminder of the paper is organized as follows. Section II presents three used classical HRA models. Section III details the experimental protocol and evaluate the HEPs using presented HRA models. Section IV gives some conclusions and perspectives.

II. Classical HRA models

A. THERP

Technique for Human Error Rate Prediction (THERP) is a task-related HRA model. THERP (Swain and Guttman [3]) decomposes a task down into a number of subtasks. Swain and Guttman make this subtask array into an assembly of discrete HRA subtasks, forming an HRA event tree. The appropriate HEPs are selected to represent the failure probabilities of the subtasks in the HRA event tree. The failure probabilities can be found in relative THERP tables according to the patterns of the subtasks. The HEPs in the tree are summed to give an overall HEP. Fig. 1 depicts an example of the HRA event tree. The task is broken into three subtasks: A, B, and C. Upper case letters represent errors, while lower case letters represent successes.

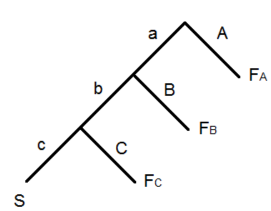

Figure 1: An HRA event tree.

B. HCR

The Human Cognitive Reliability (HCR) model developed by Hannaman [5] is based on Rasmussen's human behavior model [6] and the simulation studies of Oak Ridge and General Physics [7]. There are two hypotheses for this model: all the behavior types of human actions can be classified into skill-based, rule-based, and knowledge-based according to the Rasmussen's human behavior model; the probability of every behavior error is only related to the proportion of permitted time to execution time $t/T_{1/2}$ and is approximated with Weibull distribution [8], [9]

$$p = e^{-\{\frac{t/T_{1/2}-\gamma}{\alpha}\}^{\beta}} \quad (1)$$

$$T_{1/2} = T_{1/2}, n \times (1 + k1) \times (1 + k2) \times (1 + k3) \quad (2)$$

where t is the time window. It represents the allowable time in which the operator must take action to correctly resolve the situation. $T_{1/2}$ is the median response time. $T_{1/2}, n$ is the nominal response time. $k1, k2, k3$ are Performance Shaping Factor (PSF) coefficients. A PSF is an aspect of the human's individual characteristics, environment, organization, or task that specifically decrements or improves human performance, thus respectively increasing or decreasing the likelihood of human error [10]. $k1$ represents the operator experience, $k2$ represents the stress level, $k3$ represents the quality of operator/plant interface. α, β, γ are coefficients associated with the type of predominant cognitive process. Values of parameters $k1, k2, k3$ and α, β, γ are given in Table I and Table II respectively.

k1: Operator experience	
Advanced	-0.22
Good	0
Insufficient	0.44
k2: Stress level	
Serious emergency	0.44
Heavy workload/potential emergency	0.28
Excellent/normal conditions	0
Vigilance problem (very low stress)	0.28
k3: Quality of operator/plant interface	
Excellent	-0.22
Good	0
Sufficient	0.44
Poor	0.78
Extremely poor	0.92

Table I: PSF coefficients and their values.

Type of cognitive process	α	β	γ
Skill	0.407	1.2	0.7
Rule	0.601	0.9	0.6
Knowledge	0.791	0.8	0.5

Table II: Behavior type parameters α, β, γ.

C. SPAR-H

Context-related HRA models are different from task-related and time-related models. In task-related and time-related models, task and time are important in evaluating HEP value. However, for context-related models, the context under which the action takes place is important. For example, when an accident occurs, the crews should response to the accident and take some actions. The response of crews and their actions depend on some context elements such as the training of crews, the communication among crews, and the quality of man-machine interface. Thus, the HEP in context-related models is determined by influential context elements.

The Standardized Plant Analysis Risk-Human reliability (SPAR-H) can be considered as a task-related or context-related model because of the strong contextual influence of PSFs involved in deriving the crew HEP. The SPAR-H model will be presented here as a context-related model.

In SPAR-H [11], there is a diagnosis and action model for crew and personnel responses to accident conditions. The model consists of probabilities associated with diagnosis and action. The HEP values are usually set to be 0.01 and 0.001 for diagnosis and actions. The effective HEP is made up of these elements along with modifiers from the context.

For the case when the number of PSFs is less than 3, the base HEP equals the diagnosis HEP or action HEP multiplied by weighting factors defined in eight categories: available time, stress/stressors, complexity, experience/training, procedures, fitness for duty, and work processes. Table III shows these PSFs, levels and multipliers for each PSF. Each category has several levels. For example, in the case of experience/training, there are 3 levels: low, nominal, and high. A weighting value is allocated to each level. The final HEP is calculated by multiplying the nominal HEP by the weighting factors. The diagnosis HEP and action HEP are calculated in this manner. The overall HEP is the sum of diagnosis HEP and action HEP.

$$HEP = NHEP \times PSF_{composite} \quad (3)$$

For the case when the number of PSFs, for which the weighting factor is greater than 1, is not less than 3, the base HEP is given by the following formula

$$HEP = \frac{NHEP \times PSF_{composite}}{NHEP \times (PSF_{composite} - 1) + 1} \quad (4)$$

where HEP is the effective error for either diagnostic or action error; $NHEP$ is the nominal HEP (0.01 for diagnosis and 0.001 for action); $PSF_{composite}$ is the product of all PSFs.

PSF	PSF level	Multiplier
Available time	Expansive time	0.01
	Extra time	0.1
	Nominal time	1
	Barely adequate time	10
	Inadequate time	HEP=1.0
Stressors	Nominal	1
	High	2
	Extreme	5
Complexity	Nominal	1
	Moderately complex	2
	Highly complex	5
Experience/training	High	0.5
	Nominal	1
	Low	3
Procedures	Nominal	1
	Available, but poor	5
	Incomplete	20
	Not available	50
Ergonomics/HMI	Good	0.5
	Nominal	1
	Poor	10
	Missing/misleading	50
Fitness for duty	Nominal	1
	Degraded fitness	5
	Unfit	HEP=1.0
Work processes	Good	0.8
	Nominal	1
	Poor	2

Table III: SPAR-H PSFs, levels and multipliers for each PSF.

To take dependence into account, SPAR-H uses a defined decision tree with the following headings: crew (same or different), time (close or not close), location (same or different), and cues (additional or no additional). The results of all pathways are complete, high, moderate, low or zero dependency. A dependency condition table equivalent to the decision tree has been constructed. The user can follow the choices on the four headings through the dependency condition table to arrive at a level of dependency (zero to complete).

III. Experiment

In this section, the railway platform, the experimental protocol, and the analysis of the experimental results are detailed successively.

A. RCCS platform

RCCS is a Centralised Traffic Management System used to manage the traffic. It is currently used on important lines such as the Cambrian lines in United Kingdom and the high speed train connection between Figueras (Spain) and Perpignan (France). It provides complete railways traffic solutions including expert functions like automatic conflict resolution, automatic possession setting. It is used on complex networks including mixed traffic, big stations and terminals.

Figure 2: RCCS Platform.

Fig. 2 shows the RCCS platform in Heudiasyc Laboratory. The RCCS platform is composed of one server and four workstations, including two signaller workstations, a supervisor workstation and a maintenance workstation. Fig. 3 shows the complete server/client architecture on PC via Ethernet. There are five servers in this architecture. The railway platform in Fig. 2 is a simplification of the architecture in Fig. 3.

The RCCS provides the central control function of the Channel Tunnel Rail Link (CTRL) rail traffic. It enables to manage: the CTRL rail traffic, the route settings based on a timetable, the delays and other incidents to be identified, the reports linked to the rail traffic management to be edited.

Signaller tasks include train running monitoring and control, route setting (automatic and manual) management, possessions/isolations monitoring and control, emergency control of the CTRL, alarm monitoring, incident and daily reporting. Supervisor tasks include CTRL monitoring, signaller support and assistance. Maintainer tasks include system monitoring, parameter maintenance, system maintenance and configuration.

B. Experimental protocol

The protocol was defined in order to evaluate the performance of human operators, especially movement inspectors and pointsmen, in railway systems.

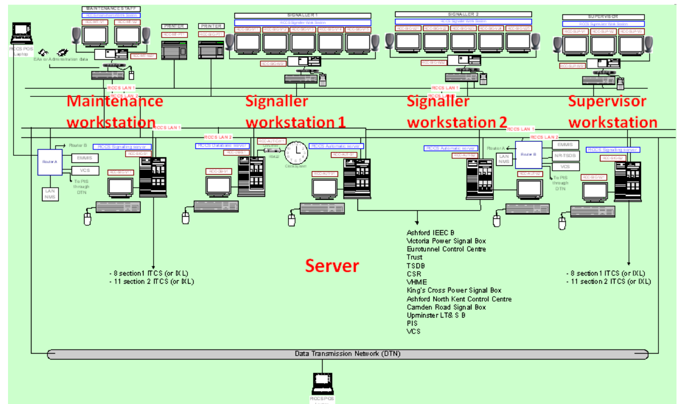

Figure 3: Architecture of RCCS Platform.

Five experimental subjects participated in the experiments. Because of the considerable amount of time spent by each subject, it is not practical to perform experiments by more subjects. Significant parameters were evaluated by the experiments. Before the experiments, each subject was trained to detect six representative types of failures:

- Points end detection is out of correspondence. Out of correspondence means that a piece of equipment was required to do a task, but the indication coming back shows that it did not perform the task. When this term is used to refer to a point, it means that the point was required to be controlled normal but was detected to be reverse, or was required to be controlled reverse but was detected to be normal. Fig. 4 illustrates the scenario where the point 2055 is detected to be out of correspondence. The yellow circle indicates the position of point 2055. When a point is detected to be out of correspondence, it twinkles. This kind of failure

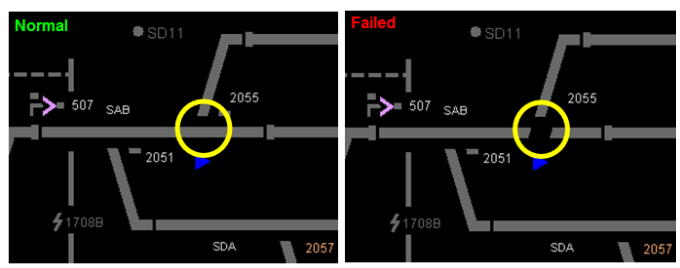

Figure 4: Points end detection is out of correspondence.

can be detected by the maintenance workstation. The corresponding message is shown as follows:

07/04/14	17:05:54	STRA	2055	Points end detection is out of correspondence
07/04/14	17:05:54	STRA	2055	Swing nose is out of correspondence

It means the failure is located at STRA (Stratford). The failed equipment is the point 2055. The type of failure is that points end detection is out of correspondence. For those points which have swing noses, the second message will also appear on the maintenance workstation. For those points which have fixed noses, only the first message appears on the maintenance workstation.

- Interval track circuit fails. Track circuits can fail due to many reasons, for example, a short circuit caused by water. Fig. 5 illustrates the scenario where the interval track circuit PAM fails. This kind of failure

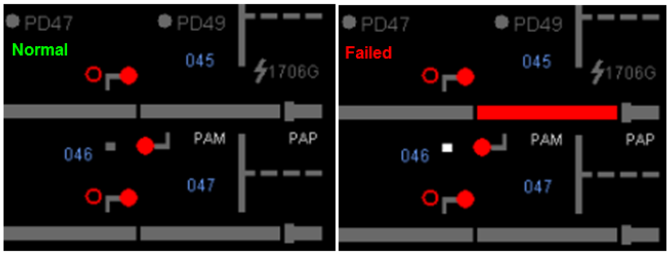

Figure 5: Interval track circuit failure.

can be detected by the maintenance workstation. The corresponding message is shown as follows:

07/04/14	17:07:32	ST PA	PAM	Track circuit equipment defect 437_TCWR_PAM

It means the maintainer detects a track circuit equipment defect. However, this message may appear not only due to the track circuit failure but also due to the departure of a train. Thus, the experimental subject is always demanded to locate the involved track circuit on the detailed view of the signaller workstation and confirm the reason of the appearance of the message. The shown message means the failure is located at ST PA (St_Pancras). The failed equipment is the interval track circuit PAM. The type of failure is track circuit equipment defect.

- Diamond Crossing Track circuit fails. Fig. 6 illustrates the scenario where the diamond crossing track circuit 2022_2023 fails. This kind of failure can be detected

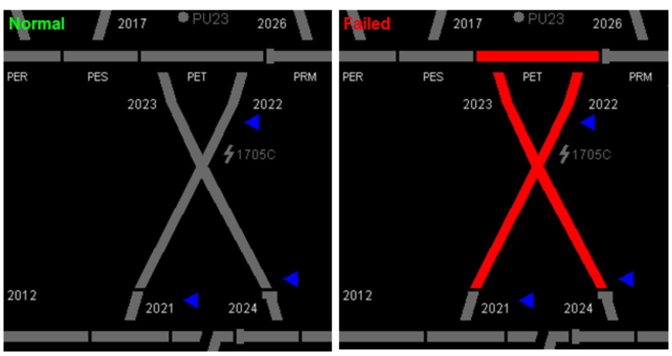

Figure 6: Diamond Crossing Track circuit failure.

by the maintenance workstation. The corresponding message is shown as follows:

07/04/14	18:14:54	ST PA	2023	Track circuit equipment defect 440_P_2023
07/04/14	18:14:54	ST PA	2022	Track circuit equipment defect 440_P_2022

It means the failure is located at ST PA (St_Pancras). The failed equipment is the diamond crossing track circuit 2022_2023 (the position is expressed by the two points included in the diamond crossing track). The type of failure is track circuit equipment defect.

- Overhead elementary section (OES) is powered off. Electrical zones transmit electrical energy to trains by overhead lines. Fig. 7 illustrates the scenario where the overhead elementary section 1704B is powered off. This kind of failure can be detected by the maintenance workstation. The corresponding message is shown as follows:

213	07/04/14	18:37:07	ST PA	1704B	OES status is off

It means the failure is located at ST PA (St_Pancras). The failed equipment is the overhead elementary section 1704B. The type of failure is that overhead elementary section status is off. However, the information in red is updated so quickly that it's hard to be detected by eyes.

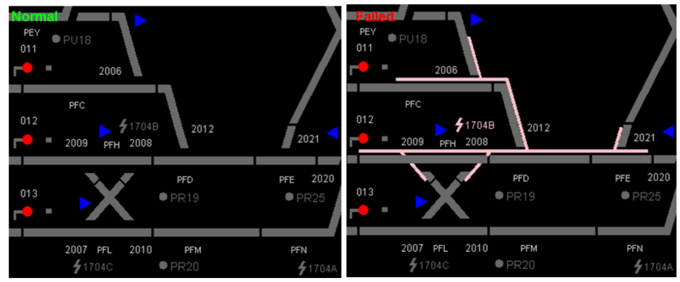

Figure 7: Overhead elementary section is powered off.

- Absolute Stop Marker Route Failure Control forces a traffic light to be red when a traffic light fails or the following track is occupied or broken. Fig. 8 illustrates the scenario where the traffic light AF121 is turned red due to the Absolute Stop Marker Route Failure. This

Figure 8: Absolute Stop Marker Route Failure.

kind of failure can be detected by the maintenance workstation. The corresponding message is shown as follows:

213	07/04/14	18:33:44	EBBS	AF121	Marker status is closed

It means the failure is located at EBBS (Ebbsfleet). The traffic light AF121 is forced to be red. The message is that marker status is closed. However, the message in red is updated so quickly that it's hard to be detected by eyes.

- Emergency Replacement Switch is on. When an emergency replacement is required, a switch positioned in the zone of emergency replacement forces a row of traffic lights to be red, so that no train can pass. Fig. 9 illustrates the scenario where the emergency replacement switch DSR is on. This switch controls the traffic lights AF721, AF723, AF725, AF727, and AF729. This kind of failure can be detected by the maintenance workstation. The corresponding message is shown as follows:

213	07/04/14	18:35:57	DAGE	AF729	Marker status is closed
213	07/04/14	18:35:56	STRA	AF727	Marker status is closed
213	07/04/14	18:35:56	STRA	AF725	Marker status is closed
213	07/04/14	18:35:56	STRA	AF723	Marker status is closed
213	07/04/14	18:35:56	STRA	AF721	Marker status is closed
213	07/04/14	18:35:56	STRA	DSR sw	Switch status is on

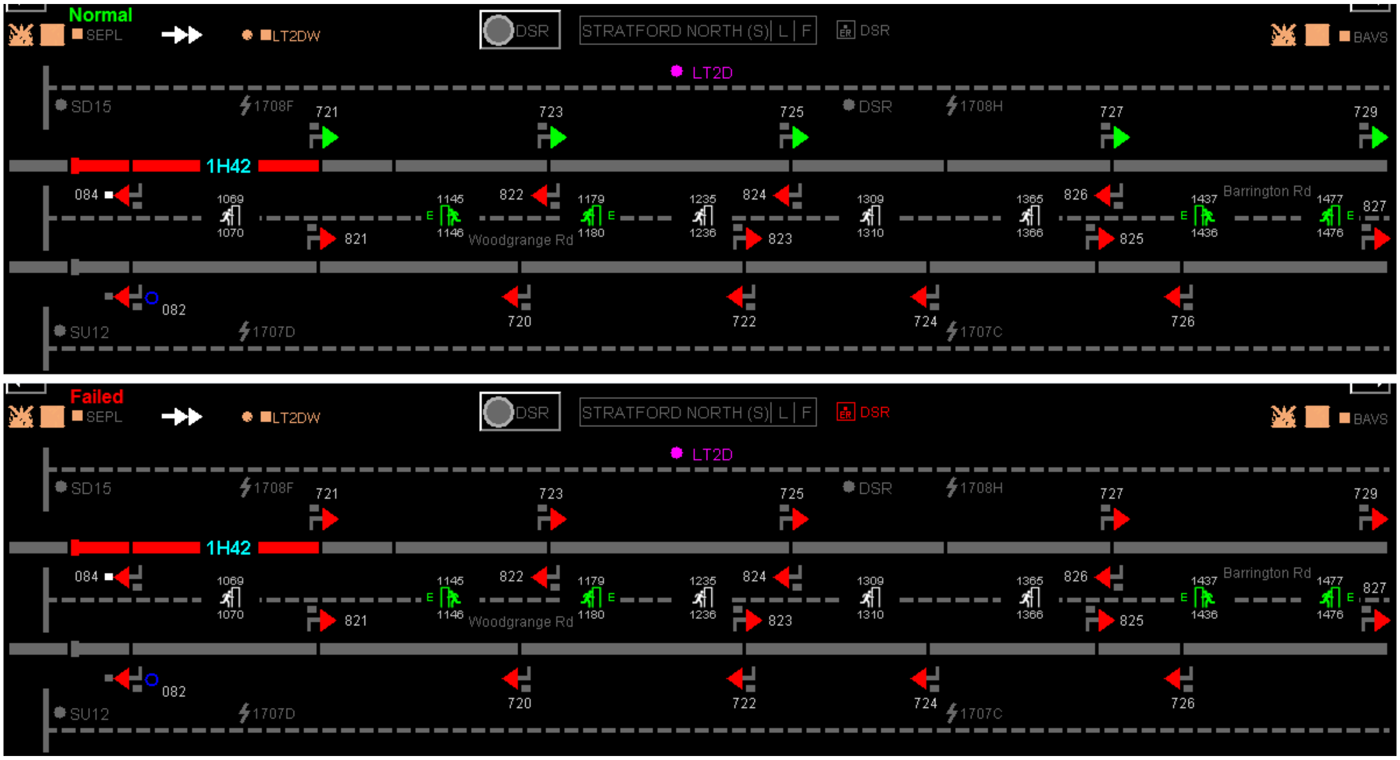

Figure 9: Emergency Replacement Switch is on.

It means the failure is located at STRA (Stratford). The emergency replacement switch DSR is turned on. DSR switch forces traffic lights AF721, AF723, AF725, AF727, and AF729 to be red. The message is that switch status is on. However, the message in red is updated so quickly that it's hard to be detected by eyes.

Four variables were included in the protocol in order to evaluate the performance of the experimental subjects in different traffic supervision configurations.

- The first variable is the knowledge level of a subject. The frame of $Knowledge$ is {0,1,2}. Three levels are set to express the knowledge level of experimental subjects: inadequate (0), medium (1), adequate (2). The knowledge level is decided by the training before the experiments and the knowledge acquired from other ways. In our research, the knowledge level is set for the subjects involved in our experiment, not for real experts in railway systems.
- The second variable is the fatigue. The frame of $Fatigue$ is {0,1}. A subject may be in tired (1) or not tired (0) state.
- The third variable is the workload. When the number of trains increases, some kinds of failures are more difficult to be detected. In this case, the workload increases. The frame of $Workload$ is {0,1}. '0' means there is less workload, '1' means there is more workload.
- The fourth variable is the experience. Each time when a subject finishes a test, his/her experience about the platform increases. This may improve his/her performance. The frame of $Experience$ is {0,1,2,3}. 4 levels are set to express that each subject performances four tests in total.

The performance of a subject is evaluated by four parameters: the detection time, the rate of correct detection, the rate of false detection, and the rate of non-detection. The detection time of an operator is defined as the difference of the time when a failure occurs and the time when the operator detects and interprets that failure. For each failure, the detect result may be correct, false or non-detected. Correct detection means that the operator detects a failure and interprets it correctly. False detection means that the operator detects a failure but interprets it incorrectly, or the operator detects a nonexistent failure. Non detection means that the operator misses a failure. When we calculate the rates for each subject, because a subject may detect nonexistent failures, the total number of recorded failures may exceed the predefined number of failures.

4 scenarios were developed to implement all the six kinds of failures. The differences among scenarios lie in the number of each kind of failures and the time when failures occur. Each scenario lasts 30 minutes. Each subject has to perform the experiments in all the 4 scenarios.

Training which presents all types of failures will be given to subjects before the experiments. Experimental subjects do not know how many failures there are and the time when they occur. Each failure will be repaired automatically after 2 minutes. Experimental subjects should detect failures, locate failures in the detailed view of signaller workstation and distinguish which kind of failures they are. The corresponding time when subjects detect failures will be recorded.

C. Experimental results

As said before, each subject has to perform the experiments in all the 4 scenarios. Thus, each subject has four performance results. Because each failure will be repaired after 2 minutes, for those non-detected or falsely detected failures, the detection

Scenario		Detection time (sec)	Rate of correct detection	Rate of false detection	Rate of non-detection
Subject1_1	Knowledge = {1}, Fatigue={0}, Workload={0}, Experience={0}	76.7	0.6	0.2	0.2
Subject1_2	Knowledge = {1}, Fatigue={0}, Workload={0}, Experience={1}	62.4	1	0	0
Subject1_3	Knowledge = {1}, Fatigue={0}, Workload={1}, Experience={2}	64.4	0.8	0	0.2
Subject1_4	Knowledge = {1}, Fatigue={0}, Workload={1}, Experience={3}	63.3	0.8	0.1	0.1
Subject2_1	Knowledge = {0}, Fatigue={0}, Workload={0}, Experience={0}	101.1	0.455	0.09	0.455
Subject2_2	Knowledge = {0}, Fatigue={0}, Workload={0}, Experience={1}	102.3	0.364	0.091	0.545
Subject2_3	Knowledge = {0}, Fatigue={0}, Workload={1}, Experience={2}	66.7	0.7	0	0.3
Subject2_4	Knowledge = {0}, Fatigue={0}, Workload={1}, Experience={3}	71.2	0.8	0	0.2
Subject3_1	Knowledge = {2}, Fatigue={0}, Workload={0}, Experience={0}	58.1	0.727	0.091	0.182
Subject3_2	Knowledge = {2}, Fatigue={0}, Workload={0}, Experience={1}	63.9	0.727	0.091	0.182
Subject3_3	Knowledge = {2}, Fatigue={0}, Workload={1}, Experience={2}	62.6	0.9	0	0.1
Subject3_4	Knowledge = {2}, Fatigue={1}, Workload={1}, Experience={3}	79.4	0.636	0.091	0.273
Subject4_1	Knowledge = {1}, Fatigue={1}, Workload={0}, Experience={0}	76.5	0.636	0.091	0.273
Subject4_2	Knowledge = {1}, Fatigue={1}, Workload={0}, Experience={1}	73.9	0.7	0	0.3
Subject4_3	Knowledge = {1}, Fatigue={0}, Workload={1}, Experience={2}	55.5	0.8	0	0.2
Subject4_4	Knowledge = {1}, Fatigue={0}, Workload={1}, Experience={3}	66.8	0.9	0	0.1
Subject5_1	Knowledge = {0}, Fatigue={0}, Workload={0}, Experience={0}	83.8	0.6	0	0.4
Subject5_2	Knowledge = {0}, Fatigue={0}, Workload={0}, Experience={1}	85.2	0.6	0	0.4
Subject5_3	Knowledge = {0}, Fatigue={0}, Workload={1}, Experience={2}	77.4	0.7	0	0.3
Subject5_4	Knowledge = {0}, Fatigue={0}, Workload={1}, Experience={3}	59.3	0.8	0.1	0.2

Table IV: Experimental results of 5 subjects.

time is set to be 2 minutes. The average of detection times of a subject in a scenario is regarded as the detection time of the subject in that scenario. Table IV shows the simulation results of the five subjects.

Compared to the real environment, the detection times are relatively unrealistic in our experiments because of the following reasons:

- Subjects know that failures will occur in each scenario. In real world, the occurrence of failures is unknown.
- Each scenario lasts only 30 minutes. In real world, the operators work more than 30 minutes.
- The number of kinds of failures is limited to 6. In real world, there are much more kinds of failures.
- Subjects are only required to detect failures. In real world, some corresponding actions need to be taken simultaneously.

Some conclusions can be yielded from the experimental result:

- From the results of the five subjects, we find that the training influences really the performance of subjects.
- As the experience of each subject increases, the performance increases at a certain extent.
- According to the performance of subject 3, fatigue may strongly influence the performance of subjects.
- The workload related to the number of trains does not have a significant impact on the performance of the majority of experimental subjects. The reason may be that the difference between the two levels of workload is not large enough to influence the performance of subjects.

D. Analysis of the result by the THERP + HCR model

Facing a failure, the reaction procedure of a subject contains three steps: detection, diagnosis, and action. Therefore, the whole task of this experiment is broken into three subtasks: detection, diagnosis, and action. There are many HRA models in the literature. In this work, we choose the THERP model which is used to analyze the task-related experiment. Fig. 10 shows the HRA event tree of the experiment. The overall HEP is the sum of the HEPs of the three subtasks in the event tree:

$$HEP = p_1 + p_2 + p_3 \tag{5}$$

where p_1 is the HEP of detection, p_2 is the HEP of diagnosis, and p_3 is the HEP of action.

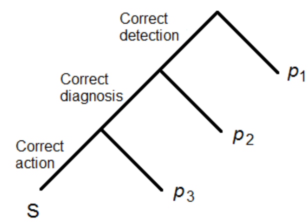

Figure 10: HRA event tree of the experiment.

Because the detection procedure is time-related, the HEP of detection is assessed by the HCR model. In the HCR model, we find the following formula to evaluate the HEP of detection:

$$p_1 = e^{-\{\frac{t/T_{1/2}-\gamma}{\alpha}\}^{\beta}} \tag{6}$$

The HEPs of diagnosis and action are

$$\begin{aligned} p_2 &= \frac{p_F}{p_C + p_F} \\ p_3 &= p_N \end{aligned} \tag{7}$$

where p_C is the rate of correct detection, p_F is the rate of false detection, and p_N is the rate of non-detection.

In the experiment, $t = 120s$ (each failure lasts 2 minutes), $T_{1/2}, n = 30s$ (we suppose that the average time that an expert detects failures is 30 seconds). Because our experiment is based particularly on the skill and the knowledge, we use the average values from the second and fourth rows in Table II as the values of α, β, γ in our model: $\alpha = 0.599$, $\beta = 1$, and $\gamma = 0.6$.

Table V lists the parameters $k1, k2, k3$ and the response time $T_{1/2}$ calculated by Equation 2 for each subject. Table VI gives the HEP of each subject obtained by the THERP + HCR model. Because all the subjects are not experts in railway systems, their HEPs are relatively high in this experiment. As the HEPs are too high for railway systems, a real expert will be invited to do the experiment.

Subject 1	Subject 2	Subject 3	Subject 4	Subject 5
$k1 = 0$	$k1 = 0.44$	$k1 = -0.22$	$k1 = 0$	$k1 = 0.44$
$k2 = 0.28$	$k2 = 0.28$	$k2 = 0.28$	$k2 = 0.28$	$k2 = 0.28$
$k3 = 0$	$k3 = 0$	$k3 = 0$	$k3 = 0$	$k3 = 0$
$T_{1/2} = 38.4s$	$T_{1/2} = 55.3s$	$T_{1/2} = 30.0s$	$T_{1/2} = 38.4s$	$T_{1/2} = 55.3s$

Table V: Parameters and execution time of each subject.

Subject	Nb Scenario					Mean HEP
S1	1	p_1=0.0148	p_2=0.25	p_3=0.2	HEP=0.4648	
	2	p_1=0.0148	p_2=0	p_3=0	HEP=0.0148	
	3	p_1=0.0148	p_2=0	p_3=0.2	HEP=0.2148	
	4	p_1=0.0148	p_2=0.1111	p_3=0.1	HEP=0.2259	HEP_{S1}=0.2301
S2	1	p_1=0.0727	p_2=0.165	p_3=0.455	HEP=0.6927	
	2	p_1=0.0727	p_2=0.2	p_3=0.545	HEP=0.8177	
	3	p_1=0.0727	p_2=0	p_3=0.3	HEP=0.3727	
	4	p_1=0.0727	p_2=0	p_3=0.2	HEP=0.2727	HEP_{S2}=0.5390
S3	1	p_1=0.0034	p_2=0.1112	p_3=0.182	HEP=0.2966	
	2	p_1=0.0034	p_2=0.1112	p_3=0.182	HEP=0.2966	
	3	p_1=0.0034	p_2=0	p_3=0.1	HEP=0.1034	
	4	p_1=0.0034	p_2=0.1252	p_3=0.273	HEP=0.4016	HEP_{S3}=0.2746
S4	1	p_1=0.0148	p_2=0.1252	p_3=0.273	HEP=0.413	
	2	p_1=0.0148	p_2=0	p_3=0.3	HEP=0.3148	
	3	p_1=0.0148	p_2=0	p_3=0.2	HEP=0.2148	
	4	p_1=0.0148	p_2=0	p_3=0.1	HEP=0.1148	HEP_{S4}=0.2644
S5	1	p_1=0.0727	p_2=0	p_3=0.4	HEP=0.4727	
	2	p_1=0.0727	p_2=0	p_3=0.4	HEP=0.4727	
	3	p_1=0.0727	p_2=0	p_3=0.3	HEP=0.3727	
	4	p_1=0.0727	p_2=0.1111	p_3=0.2	HEP=0.3838	HEP_{S5}=0.4255

Table VI: HEP of each subject obtained by the THERP + HCR model.

E. Analysis of the result by the THERP + SPAR-H model

In this subsection, we use another classical HRA model to analyze the experimental result. The whole task is also broken into three subtasks: detection, diagnosis, and action. The same as the THERP + HCR model, the overall HEP in THERP + SPAR-H model is

$$HEP = p_1 + p_2 + p_3 \tag{8}$$

where p_1 is the HEP of detection, p_2 is the HEP of diagnosis, and p_3 is the HEP of action.

In this model, the HEP of detection reflects the probability that an operator does not detect a failure. Thus, it is computed by

$$p_1 = p_N \tag{9}$$

where p_N is the rate of non-detection.

The SPAR-H model is used to evaluate the HEPs of diagnosis and action influenced by contextual elements. According to Eq. 3, we have

$$p_2 = NHEP_{diag} \times PSF_{composite} \tag{10}$$

$$p_3 = NHEP_{action} \times PSF_{composite} \tag{11}$$

where $NHEP_{diag} = 0.01$, $NHEP_{action} = 0.001$. Table VII shows all the PSFs and their levels of each subject in 4 scenarios. For certain PSFs, a subject may have different levels in the 4 scenarios. The corresponding number of scenario is given in the parentheses after the level. Table VIII gives the HEP of each subject obtained by the THERP + SPAR-H model. Because all the subjects are not experts in railway systems, their HEPs are relatively high in this experiment.

F. Comparison

Table IX lists the HEPs of each experimental subject calculated by the above two methods. Though the results of the two methods are different, we find some similarities. The HEPs of subject 2 and subject 5 are much larger than the HEPs of other subjects. The HEPs of the other three subjects do not have large difference. This is because that the subject 2 and the subject 5 are those who have inadequate knowledge, while the other three subjects are those who have medium or adequate knowledge. In fact, the subject 3 is the one who has adequate knowledge. The other two are those who have medium knowledge. However, from the results of the two methods, the HEPs of these three subjects have no big difference. This proves that when the subjects have medium or adequate knowledge, knowledge is no longer the factor who has the biggest influence on HEP, and other factors become also important.

	THERP + HCR	THERP + SPAR-H
HEP_{S1}	0.2301	0.1415
HEP_{S2}	0.5390	0.4245
HEP_{S3}	0.2746	0.1862
HEP_{S4}	0.2644	0.2568
HEP_{S5}	0.4255	0.3745

Table IX: HEP of each subject obtained by the two methods.

IV. CONCLUSION

As shown by the statistics of the Federal Railroad Administration Office of Safety Analysis [1], a large amount of accidents in railway transport are caused by human errors. Although human factors are not taken into account in railway standards defining RAMS requirements such as EN50126 [12], the necessity of taking human errors into account in railway accidents analysis is an idea widely accepted by all the experts.

The railway standards provide nevertheless a non-exhaustive list of human factors that may have an impact on RAMS parameters of systems. In this paper, an experiment was designed to evaluate the HEP of human operators, especially movement inspectors and pointsmen, in railway systems. Currently, the three subtasks of human operation are

	Subject 1	Subject 2	Subject 3	Subject 4	Subject 5
Available time	Nominal	Nominal	Extra	Nominal	Nominal
Stressors	Nominal	Nominal	Nominal	Nominal	Nominal
Complexity	Nominal(1,2) Moderately(3,4)	Nominal(1,2) Moderately(3,4)	Nominal(1,2) Moderately(3,4)	Nominal(1,2) Moderately(3,4)	Nominal(1,2) Moderately(3,4)
Experience/training	Nominal	Low	High	Nominal	Low
Procedures	Nominal	Nominal	Nominal	Nominal	Nominal
Ergonomics/HMI	Nominal	Nominal	Nominal	Nominal	Nominal
Fitness for duty	Nominal	Nominal	Nominal(1,2,3) Degraded(4)	Nominal(3,4) Degraded(1,2)	Nominal
Work processes	Nominal	Nominal	Nominal	Nominal	Nominal

Table VII: PSFs of each subject.

Subject	Nb Scenario					Mean HEP
S1	1	p_1=0.2	p_2=0.01	p_3=0.001	HEP=0.211	
	2	p_1=0	p_2=0.01	p_3=0.001	HEP=0.011	
	3	p_1=0.2	p_2=0.02	p_3=0.002	HEP=0.222	
	4	p_1=0.1	p_2=0.02	p_3=0.002	HEP=0.122	HEP_{S1}=0.1415
S2	1	p_1=0.455	p_2=0.03	p_3=0.003	HEP=0.488	
	2	p_1=0.545	p_2=0.03	p_3=0.003	HEP=0.578	
	3	p_1=0.3	p_2=0.06	p_3=0.006	HEP=0.366	
	4	p_1=0.2	p_2=0.06	p_3=0.006	HEP=0.266	HEP_{S2}=0.4245
S3	1	p_1=0.182	p_2=0.0005	p_3=0.00005	HEP=0.18255	
	2	p_1=0.182	p_2=0.0005	p_3=0.00005	HEP=0.18255	
	3	p_1=0.1	p_2=0.001	p_3=0.0001	HEP=0.1011	
	4	p_1=0.273	p_2=0.005	p_3=0.0005	HEP=0.2785	HEP_{S3}=0.1862
S4	1	p_1=0.273	p_2=0.05	p_3=0.005	HEP=0.328	
	2	p_1=0.3	p_2=0.05	p_3=0.005	HEP=0.355	
	3	p_1=0.2	p_2=0.02	p_3=0.002	HEP=0.222	
	4	p_1=0.1	p_2=0.02	p_3=0.002	HEP=0.122	HEP_{S4}=0.2568
S5	1	p_1=0.4	p_2=0.03	p_3=0.003	HEP=0.433	
	2	p_1=0.4	p_2=0.03	p_3=0.003	HEP=0.433	
	3	p_1=0.3	p_2=0.06	p_3=0.006	HEP=0.366	
	4	p_1=0.2	p_2=0.06	p_3=0.006	HEP=0.266	HEP_{S5}=0.3745

Table VIII: HEP of each subject obtained by the THERP + SPAR-H model.

dependent. In the future, we will deal with the independence and mutual exclusion of these subtasks. A possible perspective is to propose a formal quantitative model of human factors using importance measures in order to integrate it into a global model of the accident risk analysis.

ACKNOWLEDGEMENT

This work was carried out and funded in the framework of the Labex MS2T. It was supported by the French Government, through the "Investments for the future" program, managed by the National Agency for Research (Reference ANR-11-IDEX-0004-02).

REFERENCES

[1] Federal Railroad Administration Office of Safety Analysis, "Federal Railroad Administration Office of Safety Analysis," 2013. [Online]. Available: http://safetydata.fra.dot.gov/officeofsafety/default.aspx

[2] A. J. Spurgin, *Human reliability assessment Theory and Practice*. CRC Press, 2009.

[3] A. D. Swain and H. E. Guttmann, "Handbook of human reliability analysis with emphasis on nuclear power plant applications," US Nuclear Regulatory Commission, Washington, D.C., Tech. Rep., 1983.

[4] D. S. Kim, D. H. Baek, and W. C. Yoon, "Development and evaluation of a computer-aided system for analyzing human error in railway operations," *Reliability Engineering & System Safety*, vol. 95, no. 2, pp. 87–98, 2010.

[5] G. Hannaman, A. Spurgin, and Y. Lukic, "Human Cognitive Reliability Model for PRA Analysis," Electric Power Research Institute, Palo Alto, CA, Tech. Rep., 1984.

[6] J. Rasmussen, "On the Structure of Knowledge - a Morphology of Mental Models in a Man- Machine System Context," RISO National Laboratory, Roskilde, Denmark., Tech. Rep. November, 1979.

[7] E. Kozinsky, L. Gray, A. Beare, D. Barks, and F. Gomer, "Criteria for Safety-related Operator Actions: Final Report," U.S. Nuclear Regulatory Commission, Washington D.C., Tech. Rep., 1983.

[8] L. Dai, Z. Li, and P. Li, "HRA in China: Model and data," *Safety Science*, vol. 49, no. 3, pp. 468–472, Mar. 2011.

[9] M. Madonna, G. Martella, L. Monica, E. P. Maini, and L. Tomassini, "The Human Factor in Risk Assessment: Methodological Comparison between Human Reliability Analysis Techniques," *Prevention Today*, vol. 5, no. 1, pp. 67–83, 2009.

[10] H. S. Blackman, D. I. Gertman, and R. L. Boring, "Human error quantification using performance shaping factors in the SPAR-H method," in *52nd Annual Meeting of the Human Factors and Ergonomics Society*, 2008.

[11] D. Gertman, H. Blackman, J. Marble, J. Byers, and C. Smith, "The SPAR-H Human Reliability Analysis Method," U.S. Nuclear Regulatory Commission, Washington, D.C., Tech. Rep., 2004.

[12] International Electrotechnical Commission, "EN50126 Railway applications - Specification and demonstration of reliability, availability, maintainability and safety (RAMS)," Tech. Rep., 2000.