Dispatcher reliability analysis: SPICA-RAIL experiments

F. Belmonte, J.-L. Boulanger & W. Schön
Laboratoire HEUDIASYC*, Université de Technologie de Compiègne, 60200 Compiègne, France*

Abstract

Although considered as not safety critical (safety function being performed by other sub-systems), railway supervision systems can contribute to hazardous scenarios. It is of primary importance to identify this type of scenario and evaluate the behaviour of human operators. A state of the art in human reliability is presented. This article presents an experimental protocol based on an automatic train supervision system coupled to a traffic simulator. It allows one to gather data for human reliability evaluation and man-machine performance studies.
Keywords: automatic train supervision, human reliability, simulated environment, experiments.

1 Introduction

Train control systems have evolved over time to include tactical and strategic control over the traffic. These systems are based on technological barriers that have reached a high safety integrity level. The safety analysis of railway traffic systems implies the evaluation of technological barriers reliability; this could be achieved thanks to basic dependability evaluations such as Failures Modes Effects and Critically Analysis or Fault Tree Analysis (FTA). However, in all cases, railway operation could never be seen as fully "man less" operation because the traffic must be supervised. This is the field of Automatic Train Supervision (ATS) systems. ATS monitors and controls the states of all railway network sub-systems and all train operations. ATS is located in a centralised control room called the Operation Control Centre (OCC).

ATS functions can contribute to safety in some scenarios where inappropriate or mistaken decisions can seriously affect safety (see [1]). Even protected by

WIT Transactions on the Built Environment, Vol 108,
www.witpress.com, ISSN 1743-3509 (on-line)
doi:10.2495/SAFE090611

technological barriers, railway safety must take into account dispatcher reliability. Today, neither railway authorities nor regulations impose Human Reliability Analysis (HRA) in safety cases. Currently, safety cases only demonstrate that no single failure or likely combination of failures of technical equipment can lead to serious consequences. When human beings are implied in demonstrations (for example by applying procedures) safety cases consider that humans are fully reliable. However, experience shows that many accidents involve human failures or bad dispatcher's decisions. In fact, situations always exist that are covered only by procedures and thus by dispatchers, and railway safety analysis always considers that these procedures are correctly applied.

Data remains one of the major difficulties in Human Reliability Analysis. A real traffic supervision centre cannot be observed for a long period, because the disturbance generated by observers and the supplementary stress induced by the presence of external people on supervisors in critical situations.

The French state and the "Picardie" region have sustained a research project developed in collaboration with several multi disciplinal partners.

A simulation platform was installed at the university of technology of Compiègne called SPICA-RAIL. This platform is similar to a real one (ALSTOM Transport's ATS product) includes of course a traffic simulator making it possible to carry out simulations "as if" the experimental platform is really connected to a railway network. The main interest will be the possibility of re-creating real accidental scenarios in laboratory, and to be able to calibrate the quantification phase of the HRA.

2 Human reliability analysis

HRA was developed in the sixties to estimate qualitatively and/or quantitatively human errors in human interacting environments. The basic background relies on the reliability theory for conventional equipment. HRA represents a specific scientific discipline, which combines the knowledge and experience of psychology, human factors and engineering. The definition of human reliability is given in [2] by the probability that "a job or task will be completed successfully by personnel at any required stage in the system operation within a required minimum of time if the time requirement exists". HRA can be defined as a method where human reliability is estimated. To estimate this probability named Human Error Probability (HEP), an HRA model should be developed first. This model relies on an accident model. Today accidents are considered as a combination of events (the systemic approach that explains accidents in terms of interactions and coincidences) more than a succession of failures (historical model, linear approach). Finally, HRA could take into account only the individual activities, the organisation system or the socio-economic system. Given these properties, different methodologies were developed (historically and by the domain of activity) [3]. For example, the first methodologies was based on the linear approach and focused on individual aspects only, such as the *Technique for Human Error Rate Prediction* (THERP) [4]. The more recent techniques use the systemic approach and take into account the socio-economic system such as MERMOS developed by *Electricité De France*, AMSMA method developed in air traffic control EUROCONTROL or *Cognitive Reliability and Error Analysis* (CREAM) [5]. CREAM allows one to highlight the dependence of human performance on the context and provides a useful cognitive model for both retrospective and prospective accident analysis. The specificity of CREAM is that human errors are shaped more by the context than by a stochastic process.

All of these methods have the same underlying points given by three important steps:

- analysis of the working environment;
- quantification of the possible human errors;
- evaluation of the procedures and of the consequences of human error.

Current HRA practice implies sometimes arbitrary quantitative evaluation of Human Error Probability due to the data validation problem [6, 7]. Indeed, the quantification lies in tables of human error probabilities or on probability distributions. The quantification needs a calibration and a validation in each studied domain of activity. These two phases are the most difficult and the most critical steps of the methodology. Data is collected from feedback experience and/or simulations and is essential for probability calibration and validation. For this reason, numerous HRA quantifications are still subjective. Therefore, [8, 9] provide useful methodologies for rescaling a subjective scale containing at least two empirical anchors into an objective scale probability. Other approaches use uncertainty studies based on fuzzy logic or belief function theory; see [10] for a complete review. Finally, a recent enhancement of CREAM allows one to estimate a mean failure rate directly without invoking the notion of human error [11].

The role of ATS activities in safety systems stems from the detection of safety critical events.

The Human Cognition Reliability HCR model (see [12, 13]) was elaborated in order to calculate the probability that a control team answers correctly to a safety critical event on time. HCR calculates this probability from the time available before the accident and the type of human behaviour required to recover the situation. The basic assumption of this model provides that this probability mainly depends on the time available before the accident and the nature of the cognitive activity required by operators. In HCR the cognitive model stems from the Rasmussen SRK ladder (see [14]). Results are given by curves for each nature of cognitive behaviour, where the probability of success is a function of the time available before the accident. This model could integrate some performance shaping factors that influence the reaction time.

This paper presents an experimental protocol elaborated in order to evaluate the train traffic supervisor time detection of three types of incident. The situation of the ATS operator of Ladbroke Grove (UK) in 1999 reveals the importance of the time detection of critical incidents. At eight o'clock in the morning a train passed through an absolute stop signal and then faced a high speed train coming from the opposite direction. In spite of the fact that the ATS operator was not responsible for this accident (the origin coming from the signal passed at danger), the inquiry [15, 16] shows that the ATS operator saw the accident coming on the ATS interface, but the time to react was too short. He realised what happened 20 seconds after

www.witpress.com, ISSN 1743-3509 (on-line)

the signal passed at danger and put a signal to red just in front of the high speed train, which engaged the emergency break. This action came too late to prevent the collision since a high speed train requires a long distance in which to stop.

The objective of experimentation presented below was to evaluate the efficiency of the joint system composed of the ATS–Man Machine Interface (MMI) and the human operator to detect equipment failures. This allows one to more deeply study the cognitive behaviour of the human operator.

3 Experiments

3.1 Platform SPICA-RAIL

SPICA-RAIL is a real *Automatic Train Supervision* (ATS) product developed by ALSTOM Transport. The ATS system is connected to the interlocking system and the *Automatic Train Control* (ATC) system (which includes *Automatic Train Operation* and *Automatic Train Protection*, see [17] for descriptions of these systems). The ATS supervises all the traffic from staff and rolling stock management to signalling and route setting monitoring and control. These last two functionalities, intensively automated in this last decade, remain safety critical operations when degraded circumstances occur [1]. The interlocking and the ATC are simulated in SPICA-RAIL by the traffic simulator system developed by ALSTOM Transport in order to validate and test ATS projects. The traffic simulator allows one to simulate train traffic operations by a scripting informatics language.

The ATS delivered by ALSTOM is the clone of a recent project. The line supervised is twenty one kilometres long, has two side lines and is of the suburban traffic type. It includes a bifurcation, commercial stations and several interline switch points. This track plan is a typical suburban railway line operating homogeneous trains (same speed, same size, same weight). This kind of traffic is generally regulated by the frequency intervals between trains (named "constant headway"). In order to generalise our experiments to main railway lines, we introduced new equipment allowing one to regulate the traffic in time and in space, indeed several kinds of trains (different speeds, sizes and weights) could be regulated. This equipment includes two supplementary tracks that allow low speed trains to be overtaken by fast trains. We developed, in collaboration with ALSTOM, the necessary engineering tasks in the ATS and the traffic simulator to obtain this equipment. The simplified view of the track plan is presented in figure 1.

The Man Machine Interface is composed of two elements. There is a general mimics display called Schematic Control Display SCP that can be seen by every operator in OCC. The second element corresponds to the operator's computer that presents several views of the railway track plan. Figure 2 shows the SPICA-RAIL platform and the general MMI of the track plan.

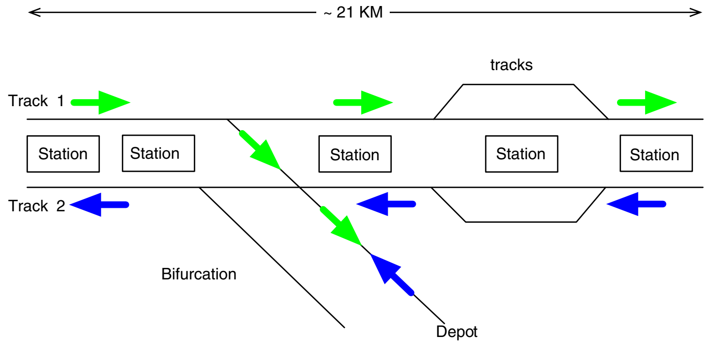

Figure 1: SPICA-RAIL track plan.

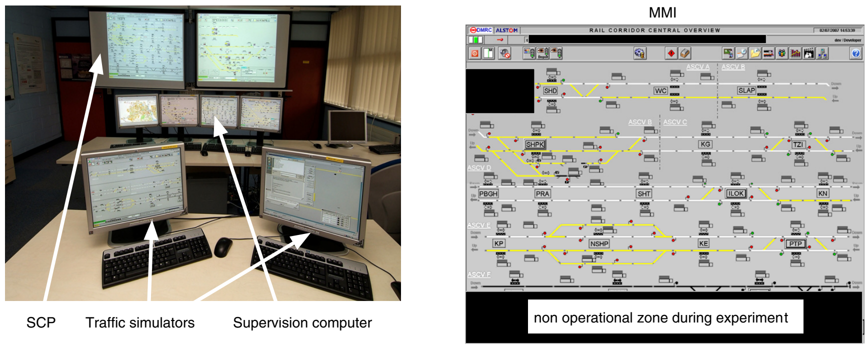

Figure 2: SPICA-RAIL platform.

3.2 Experimental protocol

Three voluntary novices, assimilated to the ATS operator in formation, participated in the experiments. Each of them was formed and trained to the detection of three types of incidents:

- Switch point uncontrolled. This problem comes from the position sensor that cannot indicate the real position of the point. Thus, the signal protecting this point goes to absolute stop. This incident is not safety critical, but implies serious disturbances on the traffic and stresses the activity;
- Signal failed to open. This incident is not safety critical, but as a precedent seriously stresses the traffic;
- Signal failed to close. This incident is safety critical, it has the same consequences as a signal passed at danger. It is thus of prime importance to detect this default in order to ensure safety.

Four variables have been introduced in the protocol in order to evaluate the monitoring performance in different traffic supervision configurations.

The type of incident is the first variable, because there are more indices on the MMI for an uncontrolled switch point, we assume that the time detection would be lower for this type of incident than the case of signal failure.

The time of the incident occurrence is the second variable. In order to simplify the sessions, each scenarios was 30 minutes long. We assume that the detection would be faster at the beginning of the session.

The third variable concerns the presence of a train around the incident. We assume that human operator focus his attention periodically on dynamic elements of the MMI. Consequently, the detection should be lower when trains circulate near the incident.

Routes are pieces of track where trains are authorised to circulate. There is two way to trace a route from ATS. The permanent way traces the route only once, and trains do not clear the route after crossing. The automatic destruction way trace the route for only one train, the route is cleared after crossing. Because their are more graphical indices in case of equipment failure on permanent routes, we assume that the time detection should be lower in this case.

Twenty scenarios were developed in order to vary the modalities of these variable. Ten distractors scenarios have been introduced in order to prevent subject to make inference on their evaluation.

Each session has been performed individually. Subjects have been asked to detect equipment failures and to diagnose the failures. Data collected are time detection and the number of correct and false detections.

4 Results

Ten hours of recorded video per subject have been collected. In general, the majority of equipment failures have been detected. Non detection rate and false detection rate are 1/30 for each subject. This result indicates that the formation of the subjects was efficient.

Because of high variability inter and intra subject and the presence of extreme values, means and variances could not be used. In consequence, statistical analyses are performed with medians and ranges. The median detection time (T_D) is of $11, 96$ with a range of 299. Analyses of detection times on the basis of scenarios tests reveal a high variability between subjects. Medians and ranges of the three subjects are presented in table 1.

This primer analyses implies that detection times are relatively high considering the favourable environment of the experiment:

- Subjects knows that there is one failure to detect in each scenario (this is not the case in real situations);
- There are only three kinds of failure to detect. In real situations events could be more varied;
- Duration of scenarios was very short (30 mn) in comparison to 6 hours in real ATS;

Table 1: Detection times: medians and ranges.

Subject	T_D **median (sec.)**	**Range**
1	5,48	299
2	22,42	160,9
3	11,57	298,96

- Detection was the only task to perform, in ATS several activities could be perform in parallel;
- The track plan was simplified to a short zone with a single bifurcation.

The subject number two seems to be less effective than the two others. His result indicates that 60% of his detection times are greater than 20 sec. This rate fall to 20% for the subject 1 and 3. It put forward that detection time depends on the strategy used by operators to explore the MMI. Moreover, the failure information given by MMI are not sufficiently striking to be perceived in the same manner by all subjects.

Data collected does not fulfil the normality and homogeneity conditions, as a result of which statistical significance for every comparisons between subject and variable modalities have been realised with non-parametric test for paired observations. Significance threshold was fixed at $\alpha = 0,05$ for every comparisons.

The Mann and Witney statistic have been performed to compare subject together. This statistic makes the sum of the number of observations of the first subject that are greater than those of the second. Table 2 shows the p-values, that represent the probabilities of false reject for the null hypothesis that the two samples are the same.

Table 2: Comparison between subjects. Mann and Witney, $\alpha = 0.05$.

Subjects	p	H_0: $T_D^1 = T_D^2$
Subject 1 - Subject 2	0,00041	Rejected
Subject 1 - Subject 3	0,21	Accepted
Subject 2 - Subject 3	0,03	Rejected

This result confirms the primer descriptive statistic. Detection times of the subject number 2 are significantly different from the two others.

In order to evaluate the impact of the traffic configuration variables on the detection times, the Wilcoxon signed-rank test was employed. This test is usually used for two related samples or repeated measurements on a single sample. It involves comparisons of differences between measurements.

Unfortunately, these tests are not as much significant as required to involve conclusions on the impact of the modalities of the four variables on the detection time. Sample size seems to be responsible of this lack of significance. Nevertheless, a non parametric paired rank test such as Wilcoxon test is applicable for these data. In each case, the number of pair is greater than 8, furthermore this kind of test could give results for few pair [18].

However, the effect of the variable "presence of train around incident" give significant result. It was assumed that the detection time should be lower when a train circulates around the incident. The Wilcoxon statistic reveals the opposite effect with significance. This result is quite surprising. Indeed, it was reasonable to suppose that the strategy used by the human operator consists of following periodically the evolution of trains. Complementary investigations should be performed to understand this effect.

5 Conclusion

In order to perform human reliability analysis, a real automatic train supervision product was integrated and coupled to a traffic simulator. This environment called SPICA-RAIL allows to recreate safety critical scenarios that could not be observed directly from real OCC because the low frequency of these scenarios and the disturbance generated by observers on the activities of operators. First experiments realised in collaboration with specialists of psychology have been performed with SPICA-RAIL platform. The efficiency of the joint system formed by the ATS-MMI and the operator was evaluated by the measure of time detection of three kinds of incidents.

The statistical result involves that it will be useful to extend this experiment with the study of the strategy used by human operator to monitor the traffic. A meaningful difference was recorded within one subject. Thus this difference could be provided by the monitoring strategy. However, more subjects should be evaluated in order to confirm this difference. Moreover, the contradiction with the hypothesis put forward relative to the presence of trains around the incident is a supplementary clue to extend the study in this way. If it will be possible to compare different monitoring strategies and classify them by their performance, several enhancement of ATS MMI would be specified.

References

[1] Belmonte, F., Berkani, K., Boulanger, J. & Schön, W., Taking into account human factors in railway supervision. *Ninth International Symposium of the ISSA Research Section: Design process and human factors integration: optimising compagny performance.*, ed. I.S.S.A. (ISSA), Nice (France), 2006.

[2] Meister, D., *Reliability Handbook*, McGraw-Hill, New York, chapter Human factors in reliability, section 12, 1966.

[3] Bieder, C., *Les facteurs humains dans la gestion des risques, évolution de la pensée et des outils*. Hermes, 2006.

[4] Swain, A.D. & H-E., G., Handbook on human reliability analysis with emphasis on nuclear power plant application. Technical Report NUREG/CR-1278, USNRC, 1983.

[5] Hollnagel, E., *Cognitive reliability and error analysis method*. Oxford: Elsevier Science Ltd, 1998.

[6] Dougherty, E.M., Human reliability analysis - where shouldst thou turn? *Reliability Engineering & System Safety*, **29(3)**, pp. 283–299, 1990.

[7] Strater, O., Considerations on the elements of quantifying human reliability. *Reliability Engineering & System Safety*, **83(2)**, pp. 255–264, 2004.

[8] Apostolakis, G., On the assessment of human error rates using operational experience. *Reliability Engineering & System Safety*, **12(2)**, pp. 93–105, 1985.

[9] Moray, N., Objective & subjective estimates of human error. *IEEE Transactions on Reliability*, **38(3)**, 1989.

[10] Kim, B.J. & Ram, R.B., Uncertainty of human error and fuzzy approach to human reliability analysis. *International Journal of Unertainty, Fuzziness and Knowledge-Based Systems*, **14(1)**, pp. 111–129, 2006.

[11] Fujita, Y. & Hollnagel, E., Failures without errors: quantification of context in HRA. *Reliability Engineering & System Safety*, **83(2)**, pp. 145–151, 2004.

[12] Hannaman, G.W., Spurgin, A.J. & Lukic, Y.D., Human cognitive reliability model for pra analysis (epri rp 2170-3). Technical report, Electric Power Research Institute, Palo Alto, CA, USA, 1984.

[13] Moieni, P., Spurgin, A.J. & Singh, A., Advances in human reliability analysis methodology. part i: frameworks, models and data. *Reliability Engineering & System Safety*, **44(1)**, pp. 27–55, 1994.

[14] Rasmussen, J., Skills, rules, knowledge; signals, signs, and symbols, and other distinctions in human performance models. *IEEE Transactions on Systems, Man and Cybernetics*, **13**, pp. 257–266, 1983.

[15] Lord-Cullen, The ladbroke grove rail inquiry, part 1. Technical report, HSE Books, Her Majesty's Stationery Office, Norwich, 2000.

[16] Lord-Cullen, The ladbroke grove rail inquiry, part 2. Technical report, HSE Books, Her Majesty's Stationery Office, Norwich, 2001.

[17] Belmonte, F., Berkani, K., Boulanger, J. & Schön, W., Safety enhancement of railway traffic by modern supervision systems. *Seventh World Congress on Railway Research.*, ed. WCRR, Montreal (Canada), 2006.

[18] Siegel, S. & Castellan, N., *Nonparametric statistics for the behavioral sciences*. McGraw-Hill, New York, 1988.