

## **Human factors specialists to the rescue for (other) systems designers – case study in railway**

Kirsten B.J. SCHREIBERS Eur.Erg., Richard VAN DER WEIDE Eur.Erg. and Colete WEEDA Eur.Erg.

*INTERGO human factors & ergonomics, Utrecht, The Netherlands*

**Abstract.** It is a challenge for human factors professionals to be seen as a colleague system designer of man-machine systems. Mostly, they are consulted for detailed user interface design. However, in this design phase fundamental questions about task allocation necessary for an optimal man-machine interface are cut short. This case describes the development of a warning system for train drivers to illustrate the role of the human factors specialist. The original question for the warning system turned from detailed design level towards a full partnership in system design by covering issues in which human factors specialists distinguish themselves from IT specialists.

**Keywords.** System design, non-perfect automation, automated decision aids, performance measurement.

### **1. Introduction**

If the human factors professional had not been present in the design team, a new warning system for train drivers, ORBIT, would not have been developed. This case could have been an example of the ‘classical’ way other systems designers view the human factors professional. At least one out of seven common reasons not to implement human factors were valid as Helander (1999) described: “first design the technical system, then consider human factors”. Another classic objections could have been “ergonomics is common sense”. However the project team allowed the human factors specialist to change level from detailed user interface design to more fundamental systems design issues. Dul et al. (2012) called for reflection on case studies and the related challenges on the two closely related outcomes of human factors: performance and well-being. Sharing this reflection aims for further exploration and implementation of specific lessons learned.

#### *1.1 Measures against incidents by signals passed at danger (SPAD).*

In The Netherlands, the most deployed automatic train protection system (ATB-EG) leaves a ‘gap’ below 40 km/h. Each year about 200 signals passed at danger (SPAD) occur. This equals on average once per ten to fifteen years per driver. Although the Dutch railway system is relatively safe, some recent incidents and accidents reinforced the need for extra measures to reduce the number as well as the effect of SPADs. One of these measures is a new warning system to fill the 'ATB gap' to a large extent. The system warns the driver when approaching a red signal at (too) high speed. Its goal is to bridge the time until nationwide implementation of the European Rail Traffic Management System (ERTMS)/ European Train Control System (ETCS).

### 1.2 A new non fail-safe warning system for train drivers, change of paradigm

The ORBIT warning system uses information present in traffic management systems, combined with GPS information and a relatively simple algorithm. The system continuously checks train position, speed, and brake characteristics against the position of the next red signal, and generates an alert when a critical speed-to-distance curve (“alarm curve”) is passed. Due to limitations of communication and GPS the system is not fail-safe.

The human factors specialist was asked to deliver specifications for the alarm ‘beep’ as it was unclear for technical designers whether it should resemble existing alarms or not. These specifications were needed at the end of the project. Anticipating, the human factors specialist inquired beforehand about the intended system characteristics. The designers described their design paradigm in progress: the train driver should better not be aware of the system at all, in order to prevent overreliance on this non fail-safe system. False alarms were to be avoided by implementation of extra time buffers to maximize reliability of the system (taking delays in traffic management systems, GSM-R/ UMTS, reaction times of brakes, etc. into account). However, in this paradigm engineers did not manage to warn the train driver in time. Although some effect on risk of reaching the point of danger was expected, the number of SPADs would not really decrease, and the business case was about to fail.

As is common in other industries, like air traffic control (Wickens et al., 2009), the human factors specialist challenged the paradigm and suggested to tolerate false alarms to a certain extent. In the revised paradigm a preliminary attention (alert), with a degree of uncertainty, informs the train driver to verify his mental model by checking for a red signal outside. If the attention has not been followed up effectively the system additionally generates an alarm – like in the original concept - some seconds later, urging the train driver to brake immediately and stop before the red signal, see Figure 1

(Van der Weide et al., 2014)

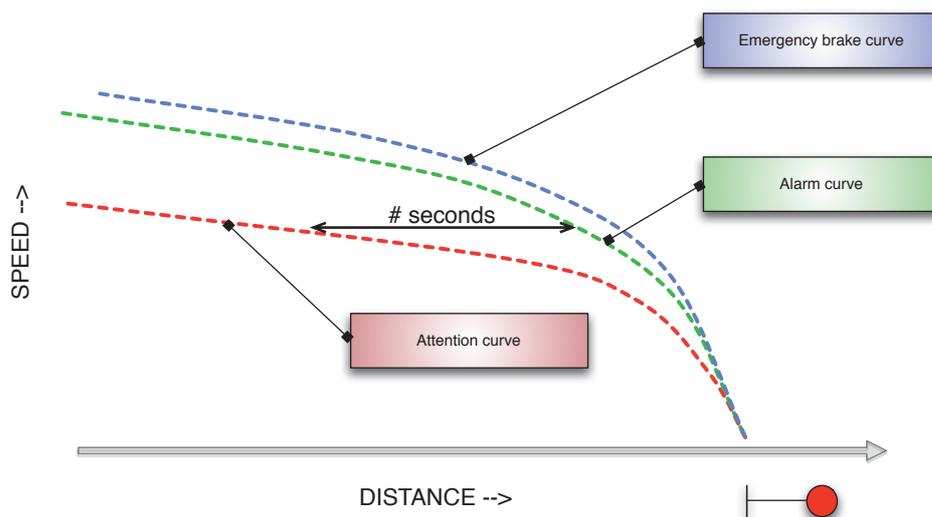


Figure 1: Schematic illustration of the warning system, consisting of train speed

*curves. The attention is followed some seconds later by an alarm if the actual train speed still exceeds the attention curve. The alarm is related to the emergency brake curve. Aim is a stop before the red signal (X-axis).*

In addition to these curves an algorithm based on positive acceleration and speed within a distance of 20 m of the red signal is used to warn at unauthorized departure. In this revised design paradigm the warning would be triggered in time for a train driver to react, resulting in a reduction of SPADs and limited impact. The business case suddenly got the right impulse. The human factors specialist changed level from detailed user interface design to system design and joined the system designers' team.

## 2. Methods

The human factors specialist focussed on alert salience by specifying the (auditory) ORBIT warning, and on the balance between reliability and number of (true and false) alerts taking into account risks of compliance and reliance. The human factors research during the development of ORBIT consisted of three steps.

A literature review had to underpin the design paradigm of tolerating false alerts. It had to point out the optimal ratio between true and false alerts, optimal signal frequency and most suitable type of signal. International guidelines and scientific publications were collected from internet (Google Scholar, Psycinfo) and through the snowball method. Search criteria were e.g. false alarm (automation), guidelines alarms, level of automation, false alarm near-perfect automation, and trust false alarms. About forty sources were included, especially on railway and road because of similarities in complexity, level of automation, signalling frequency and (limited) ways for intervention.

Simulation studies were carried out in order to fine-tune design parameters like type and timing of the 'beep' and validate driver acceptance and system efficacy. Seventeen experienced passenger train drivers from four companies were involved in half-day sessions using the ProRail MATRICS simulator. The simulated outside view was projected with a beamer. After a 15 minute test drive on a part of the experimental route, a test route of about 30 minutes was used for the evaluation of several ORBIT settings. The route considered a stopping train service in a narrow timetable, in order to meet as many red signals as possible, including crossing a complex yard (often prone to SPADs). At least one ORBIT-signal was presented in the scenario. The following conditions were tested: auditory warning: tone (attention: single tone 700 Hz, 0.2 s, 1/1 s; alarm: three-tone: 700/930/700 Hz, 0.7 s, 1/0.85 s) or a woman's voice (attention: "Attention, SPAD", alarm: "Brake now! SPAD!", translated from Dutch) and timing of attention curve (3 to 9 seconds before the alarm curve). Volume of all ORBIT warnings was about 10 dB(A) higher than simulated ambient noise of 60 dB(A). Given time constraints in users' participation, the drivers were present during other driver's tests to get maximum experience with all test conditions. Train drivers are used to the presence of colleagues in the cab for e.g. educational purposes. In order to distract drivers, as often happens during SPADs, a secondary visual and cognitive task was added. Within the primary field of view every 15 seconds a new sum was to be solved (+/- 7 or 13 to/from a two digit number between 30 and 70). Evaluation was done by loggings of reaction times (1/300 ms), number and nature of ORBIT warnings, questionnaires and debriefing semi-structured interviews. It appeared impossible to simulate TMS data propagation delays. Thus simulations studies were based on an 'optimal' ORBIT system in terms of GPS accuracy and almost zero propagation delay (Van der Weide e.a. 2014).

In a twenty days field test ten intercity trains of one type in regular service on The

Hague -Venlo route were equipped with on board ORBIT units. Regional train drivers were educated about the goal, functioning and limitations of the warning system. Beside validation of technical aspects of the system in practice by other system designers, the human factors specialists validated human behaviour and user acceptance. A remote test facilitator activated the ORBIT unit when an instructed driver was present. Loggings of distance to signal, speed, acceleration and type of ORBIT alerts were evaluated and added to data from structured interviews with train drivers involved in an ORBIT alert.

In parallel of these steps, system decision makers were early united in an advisory group, having identified hazards that should be managed. Current and future stakeholders were following progress, considering participation in future, and also occasionally looking for feedback modes on driving skills of train drivers.

### 3. Results

#### 3.1 *An adequate though not ironclad paradigm - literature review*

A review of about forty literature sources revealed classification of false alarms, valuable existence of some types of false alarms, relevance of trust, guidance for amounts of alarms and – finally as originally asked for – specifications for the kind of ‘beep’.

Automated systems affect operator’s behaviour. Complying with automation can result in adverse effects like ignoring alarms if alarms are often false and thus result in disuse or misuse of the system (Dixon et al., 2007; Lees & Lee, 2007). On the other hand trust in automation can lead to overreliance on automation: the operator does not stick to his own perception or sources, even though the system is wrong (Wickens et al., 2009).

Most sources on alarms in traffic did not explicitly distinguish between levels of alarms. In accordance to other domains an alarm is defined to activate the user to act upon an abnormal condition or a process malfunction with urgency. An alert (here ‘attention’) has a lower priority notification than an alarm. Few sources discerned between types of false alarms, situations when there is no relevant change in the state of world but automation does say so (e.g. Wickens & McCarley, 2008).

In non-perfect automated systems high rates of ‘false alarms’ may lead to ignoring the system and a cry wolf effect with late or no reaction to alarms (e.g. Parasuraman & Riley, 1997; Wickens et al., 2009). To a certain extent ‘false alarms’ appear to be acceptable or even informative for the user (Wickens et al., 2009). Based on technical characteristics of the system in this project the human factors specialist identified three types of false alarms as more or less acceptable. The classification refers to performance of the technical system (utility), process (predictability of the alarm) and general purpose of the alarm (intent) (Lees & Lee, 2007). Correct alarms are intended by the system, predictable and helpful for the train driver. First unnecessary alarms occur when the train driver feels in full control during braking but an alarm shows up anyway. Unnecessary alarms are intended by the designer, predictable, but not helpful for the train driver. They can contribute to understand system’s goals but can lead to annoyance if they occur at high frequency. Secondly, inadvertent or overdue alarms occur when the system still thinks a line side rail signal shows a red aspect while in the real world the aspect has changed to yellow or green. The system did not receive this relevant change due to long propagation times in traffic management systems, GSM-R or UMTS systems. Inadvertent alarms can contribute to understanding the system, but can also lead to undermining the system’s credibility. Thirdly, really wrong, as random experienced, alarms of the system occur while in the real world no red signal is being approached. These wrong alarms may occur because of incorrect GPS positioning of train or signal and affect trust adversely.

For non-perfect automated systems, aiming for an as perfect as possible system appears

to be useless. It has been suggested that a system reliability of about 0.80 will generally assist human performance, meaning eight out of ten accurate alarms and two out of ten misses or ‘false alarms’ (Wickens & McCarley, 2008). Much more important in non-perfect automated systems is a high level of trust and useful sources for the user to calibrate trust (e.g. Parasuraman et al., 2000). Trust is a dynamic process and compensates for feelings of uncertainty. Consistency and predictability of (mal-) functioning of the system is very important. Trust itself is influenced by propensity to automation and internal implicit attitude towards the system. In case of automation failure, users with a higher propensity show greater reduction of trust than users with lower propensity. In case of ambiguous automation, users interpret information according to their internal implicit attitude of that system (Meritt, 2012).

Considering the ‘kind of beep’, the use of an audible alarm was confirmed, as the visual channel is occupied during train driving. Furthermore, it was precondition for the system that no thorough modifications of the cab were to be made. In case of a high-expected warning frequency, e.g. daily, a sequence of beeps would be suitable. For lower frequencies (e.g. weekly) short, instructive spoken messages would be more appropriate because of other ‘beeps’ present in the cab and the unclear association of the beep.

### *3.2 A second paradigm shift – necessity of user participation during simulation studies*

The simulation studies revealed a second design paradigm shift, coming from participating train drivers. Contrary to the instruction for simulation it was observed that all drivers actuated the brake after the attention warning, instead of checking the lineside signal outside first followed by braking if necessary. They declared this behaviour being their natural reaction to auditory warnings in general as other auditory warnings refer to dangers like doors open, too high speed, alarm call, etc. They experienced the meaning of the warning system, notwithstanding being not a hundred percent reliable, as important as the other auditory warnings. Also, despite the positive attitude to the warning system they felt any attention or alarm like ‘a slap on the wrist’. They preferred the warnings to be sparse, leaving just enough time to react properly. It was concluded that the four seconds interval between attention and alarm was best, compared to longer intervals like six to nine seconds.

Train drivers preferred spoken warnings to tones, mainly because of many other tones in the train cab and mobile devices competing for attention. Also in theory spoken messages would suit best from a theoretical point of view, given the lower frequency and high-perceived urgency of the message.

The train drivers showed a high propensity to automation, a high level of trust and an implicit internal attitude for human operation. These parameters confirmed the theoretical buffer capacities needed for possible future adverse effects of the non-perfect system.

### *3.3 Setting down - field tests*

During the field tests 5800 red line side signal approaches were covered. Seventy-eight warnings were triggered, mostly attention warnings occasionally followed by an alarm. Out of those, twenty-one warnings were inadvertent/ overdue and eight were wrong (experienced as ‘random’), according to the classification set for this project. The larger part, forty-two warnings, appeared to be unnecessary. A slight seven warnings appeared to be necessary out of which one attention was actually followed by an alarm, recapturing and guiding the train driver’s attention towards the red signal. Train drivers’ acceptance was good, even for unnecessary warnings. However, human factors analyses revealed that some algorithmic parameters needed adjustment in order to reduce unnecessary alarms and prevent nuisance in the future. Other types of alarms could be banned out by technical

adaptations, like delays in the automation.

Taking future measures into account, the results indicated that this new, fairly simple warning system could result in a large-scale reduction of SPADs and a considerable improvement of impact. Given the feasibility of future measures the human factors specialist agreed with the 'go', for nationwide implementation. Additional requirements were set for future analysis of possible adverse effects on reliance and compliance, like changes in the amount of attentions and alarms.

#### 4. Discussion and Conclusion

In accordance with Dul et al. (2012) it remains a challenge for a human factors specialist to be recognized as a system designer who is also contributing to system performance. Vital moments where human factors obviously paid off were developing guidance for the design team on the design of warnings, input of appropriate reaction times, the systematic user participation of train drivers and management during simulation studies and field tests, and last but not least systematic and structured analysis of data. The human factors specialist also contributed to user participation of system decision makers, by satisfactorily settling all identified human factors hazards. Now it is our challenge to continue the desirable position as a system designer during implementation and in design of other man-machine systems.

#### References

- Dixon, S.R., Wickens, C.D., McCarley, J.S. (2007). On the independence of compliance and reliance: Are automation false alarms worse than misses? *Human factors*, 49, (2), 564-572.
- Dul, J., Bruder, R., Buckle, P., Carayon, P., Falzon, P., Marras, W.S., Wilson, J.R., Van der Doelen, B., (2012). A strategy for human factors/ergonomics: developing the discipline and profession. *Ergonomics*, 55:4, 377-395.
- Helander, M.G., (1999). Seven common reasons not to implement ergonomics. *International journal of industrial ergonomics*, 25, 97,101.
- Lees, M.N. & Lee, J.D. (2007). The influence of distraction and driving context on driver response to imperfect collision warning systems. *Ergonomics*, 50, (8), 1264-1286.
- Meritt, S.M., Heimbaugh, J., LaChapell, J., Lee, D. (2013). I trust it, but I don't know why: Effects of implicit attitudes toward automation on trust in an automated system. *Human Factors*, 55, (3), 520-534.
- Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors*, 39, 230–253.
- Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics. Part A: Systems and Humans*, 30, 286-297.
- Van der Weide, R., Schreibers, K.B.J., Weeda, C. (2014). To beep or not to beep: Developing a non fail-safe warning systems in a fail-safe train protection environment (in press). In Ahram, T., Karwowski, W., Marek, T., (Eds.). *Proceedings of the 5<sup>th</sup> International Conference on Applied Human Factors and Ergonomics AHFE 2014 Krakow*.
- Wickens, C.D. & McCarley J.S. (2008). *Applied attention theory*. Boca Raton: CRC Press.
- Wickens, C.D., Rice, S., Keller, D., Hutchins, S., Hughes, J., Clayton, K. (2009). False alerts in air traffic control conflict alerting system: is there a "cry wolf" effect? *Human Factors*, 51, (4), 446-462.