Investigating Time Pressure for the Empirical Risk Analysis of Socio-Technical Systems in ATM

Lothar Meyer [a], *Katja Gaunitz* [b] *and Hartmut Fricke* [a]

Chair of Air Transport Technology and Logistics
TU Dresden
Hettnerstr. 1-3
01069 Dresden, Germany

ABSTRACT

Current practices of risk analysis of novel socio-technical systems rely on the subjective judgment of experts. With a view on the complex interactions between human operators and the environment in ATM, a method is needed for gaining empiric evidence directly from operations. Risk analysis that bases on Human-In-The-Loop-Simulations offer a promising approach by providing an environment in which the novel system can be applied safely. An inherent disadvantage is the effort needed to cope with the strict safety targets in ATM, e.g. 1.88E-8 accidents per operating hour in which safety metrics are subject to the statistic problem of *Right Censoring.* This paper presents our novel concept to modify conditions of the simulation for gaining a calibrated acceleration effect by which the probability of safety metrics can be estimated from a shorter experimental period. This is motivated by the methodologies of *Accelerated Life Testing*, in which the *Mean-Time-To-Failure* of products is forwarded into the experimental period by applying calibrated steps of stress-load. We developed an experimental design that applies a procedure for the induction of a calibrated time-pressure for the stimulation of human error. The results of the proof-of-concept-study show controllable stress-reactions of the test persons.

Keywords: Risk Analysis, Socio-technical Systems, Air Traffic Management, Safety Assessment, Accelerated Life Testing, Time Pressure

INTRODUCTION

Current methods for estimating the risk of socio-technical systems in ATM often rely on accident and incident reports, model-based approaches or expert judgment. In particular, the predictive estimation of risk for novel systems is traditionally performed by the subjective adaption of expert's operational experiences to the expected operation after the hypothetic startup of the target system. In this respect, the term risk complies with the following definition;

> *"Risk is defined as the probability that an accident occurs during a stated period of time" (Blom & Bakker, 2003)*

The promising model-based approaches offers the advantage of coping with enormous sample spaces, providing objective data and the statistic power to prove very little probabilities of the accident event e.g. the Target Level of Safety in ATM with a maximum of 1.55E-8 accidents per operating hour (Blom H. A., et al., 2001). The modelled a-priori assumptions, regarding the expected effects of the new design on the resulting operations, are hard to validate for all cases that might occur as there are no means of obtaining direct evidence from current operations:

> *„errors are likely to be made when designers apply error modeling techniques" (Johnson, 1999)*

https://openaccess.cms-conferences.org/#/publications/book/978-1-4951-2097-8

This might impair the external validity of the model for non-expected cases.

For the above described problem, Human-In-The-Loop Simulations (HITLS) offer an empirical approach that is often used for estimating the performance of socio-technical systems in a predictive way e.g. by means of workload measures (Kirwan, 2007). HITLS has already been used for FMEA to quantify isolated probabilities in the interaction between the operator and the working environment as well as human error probabilities that can be used for the quantification of model parameters (Stroeve, Blom, & Bakker, 2013). In contrast, a pure HITLS approach is rarely used solely for risk analysis due to the enormous efforts needed to obtain valid data as well as to the limited sample spaces that can be achieved in real time simulation (Shorrock, 2001). When using valuable experts, most studies perform not more than a few 100 hours simulation time at best (Stroeve, Blom, & Bakker, 2013), providing insufficient statistic power for the safe exclusion of rare and risk-inducing events (see ATM safety-iceberg (Blom H. A., et al., 2001)). The type I error rate would be unacceptably large when assuming an unsafe system as null-hypothesis. This error can be explained by the *Weak Law of Large Numbers*, also known as Bernoulli's theorem. It describes a decreasing difference between observable frequency and the true probability with increasing sample spaces, also known as C*onvergence in Probability*. This difference can be assumed as the type I error rate, estimated by the *Chebyshev's inequality*. When assuming one operating hour as a unit that could have the end state *accident* or *no-accident*, a *Bernoulli Distribution* can be assumed in which the error can be estimated as

$$P\left(\left|X-\mu\right|\geq k\right)\leq\frac{n\cdot p\left(1-p\right)}{k^2},$$

with the random variable X, the mean μ, the variance of the distribution $\sigma^2=n\cdot p\cdot\left(1-p\right)$ and the confidence tolerance k. Even with a sample space of 1.0E9 hours and a target safety of 1 accident per 1.0E9 operational hours, there is still a 13.5% probability to declare an unsafe system as safe when no accident has been detected in the experimental time. For instance, 1.5E9 operational hours are needed for gaining 95% assurance. Thus, empirical approaches to cope with such rare events suffer of practicability to prove the novel system by means of HITLS.

Our proof-of-concept-study bases on the approach to compensate insufficient sample spaces by intensifying the probability to detect safety indicators and to hence increase the power with samples held equal. It hence follows the problem definition of Hollnagel:

> *„the problem remains of how raw data from training simulators can be modified to reflect real-world performance.“ (Hollnagel, 1993)*

Therefore we developed a concept called *Accelerated Risk AnalySis* (*AccSis*) that describes a methodology to gain an acceleration-effect when intensifying the probability for safety metrics. This acceleration effect shall practically be reached by a calibrated time pressure induction that stimulates the occurrence of human error. Concerning the time pressure induction, we developed a procedure following the *Time Budget*-principles (Bubb & Jastrzebska-Fraczek, 1999; Bubb, Human Reliability: A key to improved quality in manufacturing, 2005), which is named *Competitive Performance* (*ComPerf*) and which puts the test person under the impression of not having sufficient time to solve the problem (Chang & Mosleh, 2007). This approach is motivated by the methodologies of *Accelerated-Life-Testing* (ALT) that forwards the *Mean-Time-To-Failure* into the experimental period by means of accelerated and calibrated stress-induction during the experiment (Nelson, 2004). It explicitly addresses the occurrence of *Right-Censoring* (Cox, 1972).

In that respect, our paper presents the considered *Time-Pressure-Risk-Model* and the related conceptual framework, named *Accelerated Risk Analysis* in chapter 2. The primary subject of investigation is the problem of how to adapt the stochastic methods from ALT to the risk analysis of socio-technical systems in ATM, considering stochastic human behavior instead of stochastic processes of product aging. A HITLS experimental design is presented for the evaluation of *AccSis* and *ComPerf* following an innovative A-SMGCS for air traffic control in the scope of a proof-of-concept study in chapter 3. The results of the HITLS deliver insights on the effects of the induction procedure applying varying gradations of load indicated by means of workload measures and the detected frequency of runway incursion, all of which is outlined in chapter 4.

METHODOLOGY

The Accelerated Risk Analysis Concept

This conceptual framework has the objective of estimating the compliance of socio-technical systems with a given target probability of an accepted safety metric (e.g. the accident), expressed as an alternative hypothesis $p \leq p_{target}$, by means of HITLS-based empiric data. Facing the problem of mitigating the statistic type I error starts with analyzing *Chebyshev's inequality*. The mitigation can proceed as follows:

(1) By increasing the number of generated samples n.

(2) By modifying the simulated working conditions in the experimental design that rescales the probability by an acceleration factor a. A symmetric and linear rescaling of the target safety p_{target} and the true probability of the system p by the acceleration factor leads to $\hat{p}_{target} = a \cdot p_{target}$ and $\hat{p} = a \cdot p$ in which the alternative hypothesis is maintained. Applying the rescaling to *Chebyshev's inequality*, an effective mitigation of the type I error can be determined as follows

$$\frac{n \cdot \hat{p}(1-\hat{p})}{\hat{k}^2} = \frac{\hat{p}(1-\hat{p})}{n \cdot \hat{p}_{target}^2} = \frac{1}{a} \cdot \frac{p(1-p \cdot a)}{n \cdot p_{target}^2}$$

with $k = n \cdot p_{target}$. When defining $p << 1$ one can assume $(1-p \cdot a) \approx (1-p)$. The mitigation effect of the error can be quantified to a^{-1} and effects a virtual accumulation of the samples generated, described as $a \cdot n$.

This concept constitutes an approach to face the Safety-Iceberg problem by describing a procedure that accelerates the convergence of the type I error by modifying the boundary conditions of the HITLS, which effect a calibrated rescaling of the target and the system probability for safety relevant events.

This approach is motivated by the methodologies of *Accelerated Life Testing* (ALT), which estimates the *Mean-Time-To-Failure* (MTTF) of a physical product within a shortened experimental time. The acceleration effect is achieved by a stress-induction of e.g. thermic or mechanic stress that forwards the targeted failure event into the experimental time. In this way, the problem of *Right-Censoring* is addressed, which describes the problem of measuring the time of an event that lies beyond the experimental time (Nelson, 2004). The approach of ALT can be split into two tasks:

(1) Failure stimulation – The experiment is to be performed under varying gradations of stress that deflects the load from design stress to accelerated-stress. 3 gradations of load are recommended for capturing sufficient samples of failure events of the product.

(2) Regression analysis – The failure-distributions of each load level is fitted to analytic or non-parametric distribution models. A regression model (life-stress-model) is to be applied that extrapolates the trend of the distribution-shape to design stress (see Figure 1).

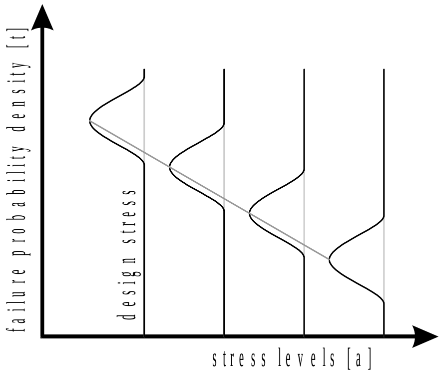

Figure 1. Stress-life-relation according to ALT concept (Nelson, 2004)

The idea to adapt this concept to accelerate the occurrence of safety relevant events in HITLS is severely impaired by the fact that human performance is a complex field that suffers of non-linearity and non-reproducibility compared to the functionality of technical products. For this reason, we identified systematic differences between the analysis of failure-events of products and the commitment of errors by operators when acting in a socio-technical system.

- The most significant difference is the stochastic that contrasts accident events of socio-technical systems and technical failure event. The product life-time is temporally limited as a result of a progress of aging which is attributed by a dependent stochastic distribution. In contrast, we assume the accident in aviation to be the result of a short term progress which hence is regarded as independent and in which a distribution cannot be modelled over time when assuming the *Bernoulli Distribution* for accidents.

- The second difference, which is that the procedures of applying stress are completely incompatible with socio-technical systems, is related to the first one,

- The third difference is the missing accident-stress-model for human behavior, since the state-of-the-art models, although describing the relationship between human error and stress, do not cover a domain-specific model curve (e.g. exponential-linear)

The concept *AccSis* comprises the application of ALT in the context of risk analysis of socio-technical systems. This paper considers the concept to be incomplete due to the reasons given above. Our current research follows a step-wise validation strategy to overcome the mentioned differences and in which finally a full compliance of *AccSis* with the requirements of risk analysis of socio-technical systems may be given. Based on this consideration, we chose the first step of the concept for proving internal validity: the acceleration of safety relevant events by intensifying human error.

In order to explain our choice of human error as the key factor, we took into account the *Integrated Risk Picture* (IRP), which describes the contribution of human errors to accidents in the combination of causal factors by means of a *Fault Tree Model* (Eurocontrol, 2006). For a socio-technical system, the IRP can be regarded as a significant fingermark of risk in which branches of failure catenation forms the resulting probability of accident. One has to consider that only branches that are affected by the acceleration effect are taken into account for the regression-analysis.

When considering causal factors in this context, one can distinguish between organizational, technical and human errors as principle causes of accidents. This complies with the model describing “a trajectory of accident opportunity” that defines the human error in the presence of a corresponding hazard as an *Unsafe Act* (Reason, 1990). The human error is today identified as the most contributing cause to accidents and incidents in aviation by a factor of 60% to 80% (Shapell & Wiegmann, 1996) or 75% (Müller, 2004). The choice of human error thus addresses a causal key factor of socio-technical systems: The major contribution of human error to risk. A vast amount of causal branches is hence expected to be covered for acceleration. Following the *ceteris paribus* principles, procedures, tasks and other boundary conditions are to be held equal during HITLS what implies a major requirement to maintain equal contextual conditions according to the conditions of design stress.

Time Pressure for the Stimulation of Human Error

Besides uncertainty, time pressure seems to be of particular relevance when considering human decision-making processes (Rastegary, 1993). Rastegary defined time pressure

> *"…as the difference between the amount of available time and the amount of time required to resolve a decision task"*.

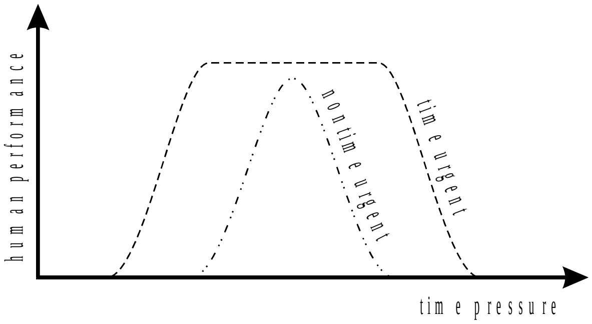

Figure 2. human performance as a function of time pressure

By empiric findings, time pressure is known to affect the human performance significantly (Freedman, Edwards, & McGrath, 1988) (see Figure 2).

This relation points to the significant impact of time pressure on human performance, i.e. on acting correctly according to the procedures. This influence can be explained by the fact that cognitive information processing is a function of time pressure that effects a minimization of cognitive effort in a cost/benefit frame of reference. It is reported that an increased selectivity of information is observable. Under time pressure, more pieces of information are used but in a shallower way (Edland, 1993).

Time pressure contributes more to Human Error Probabilities (HEP) than additional tasks when performing time critical tasks (Bubb & Jastrzebska-Fraczek, 1999). Therefore, a time budget (TB) was defined, which puts the *time available* t_a into relation to a *time needed for decision* t_n, as follows

$$TB = \frac{t_n}{t_a}.$$

An increased error probability was measured by a factor of 14 to 0.43 under the condition of time pressure. This observation corresponds to the assumptions of the *Human Reliability Assessment* THERP, which considers a factor of 10 under stress conditions (Swain & Guttman, 1983).

Time pressure and human error, in regard to which human actions can be assumed as a quality statement that is classifiable in *acceptable* or *not acceptable*, are causally linked. The deflection of the action from a minimum quality can be regarded as *not acceptable* or as human error. Continuing, the quality is linked to performance as follows

$$P = \frac{Q}{t}$$

with the human performance P, the quality of human action Q and the time given t (Bubb, 2005). Thus, time pressure affects Q, divided by time. We identified the definition of TB as an inherent advantage for the stimulation of human error for two reasons:

(1) it induces a calibrated time pressure by setting t_a

(2) human performance is sensitive to time pressure.

To summarize, the concept of accelerating the occurrence of accidents unifies many theories about accident causation and human error to a comprising causal catenation, as shown in Figure 3, with each of the links being already empirically validated by the elementary findings (Reason, 1990), (Bubb & Jastrzebska-Fraczek, 1999) and (Freedman & Edwards, 1988).

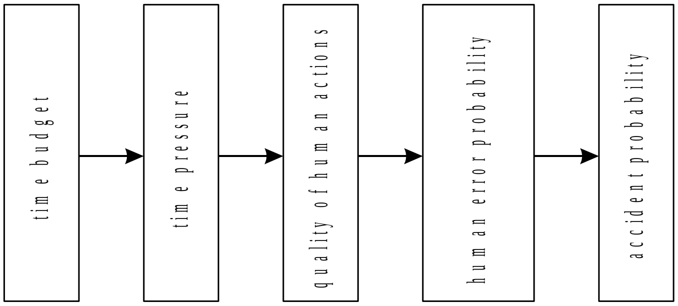

Figure 3. Causal relationship between time budget and the accident probability

This concept for utilizing the *time budget*-principle to stimulate human errors and thus to intensify the probability of accidents, is just a generic description of effect mechanism. It is necessarily a domain-specific challenge to develop a procedure that produces time pressure by means of this principle.

Competitive Performance

Most ideas for the implementation of time pressure- induction aim at setting boundary conditions that shorten the available time significantly. Secondary tasks might, for example, effect a shortening of t_a by forcing the operator to organize task sharing and prioritization. This sharing might change the pattern of activities and impacts the IRP picture without any control of the deflection from the design stress. The same can be said about the idea to intensify the task load, e.g. traffic volume.

As time pressure is a subjective feeling, we decided to choose an approach that emerges "competitive arousal". The time pressure is generated by providing a competitive environment that triggers the desire of the operator to win (Malhotra, 2010; Kerstholt, 1994). Our concept establishes the "competitive arousal" by forcing the operator to compete with a ‘calibrated reference operator’ that operates under equalized contextual conditions (cloned worlds) and is capable to act according to a calibrated performance (see Figure 4), named Competitive Performance (*ComPerf*). When the operator acts, a performance metric is measured for comparison, e.g. the throughput of the system. The head start shall be a quantified indicator for the performance of the human operator compared to the Reference Operator. The Reference Operator might be a model-based agent that supports gradations of performance.

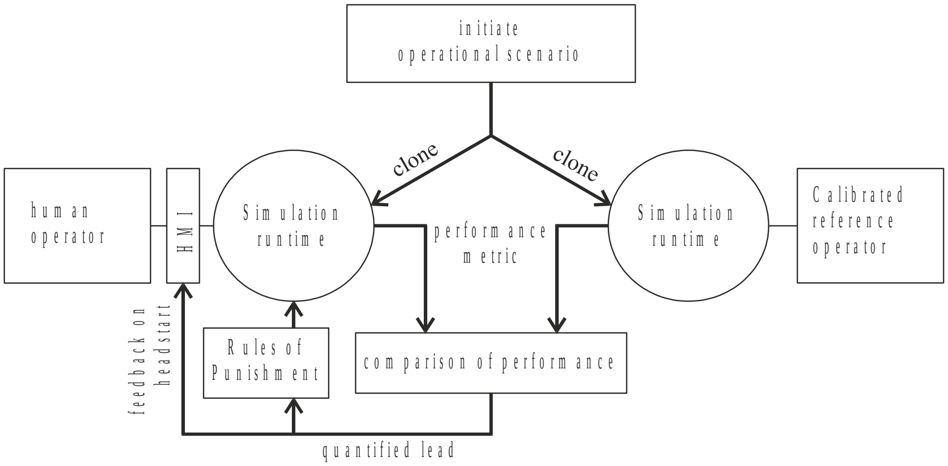

Figure 4. The concept of Competitive Performance *ComPerf*

When the head start falls back to a given threshold, the simulation is feed backed by a rule of punishment. For punishment we propose to extend the efforts necessary to finish the simulation by e.g. the generation of additional tasks or enlarging time constraints like the finish time of the simulation.

The concept provides for the fact that the boundary conditions of the simulation are kept equal even when the punishment is active. The advantage is to control the time available t_a by varying the performance of the Reference Operator and thus to establish the time-budget principles when put into relation to the decision times of the human operator t_n.

EMPIRIC STUDY

The introduction to the conceptual methodology of risk analysis by means of *AccSis* and the approach for time pressure induction with the help of *ComPerf* were both deduced to an experimental design, in which the internal validity of the risk model, as shown in Figure 3, should be the subject of investigation. The Controller Working Place (CWP) of the Air Traffic Controller (ATOC) has been chosen as an exemplar safety critical working environment in Air Traffic Control. The related task is to control traffic at the airport in the function of a tower controller according to procedures defined by ICAO PANS-ATM Doc. 4444. The principle tasks of the ATCO are defined as follows:

> *“Aerodrome control towers shall issue information and clearances to aircraft under their control to achieve a safe, orderly and expeditious flow of air traffic on and in the vicinity of an aerodrome with the object of preventing collision(s)…” (ICAO, 2007).*

The hypotheses were set as follows:

- Time pressure is sensitive to the load of the time budget set by *ComPerf*.
- The relative frequency of safety relevant events is sensitive to the load of the time budget set by *ComPerf*.

These hypotheses set the focus on two major causal relationships of the risk model (see Figure 3).

We decided to choose the Runway Incursion as the target safety relevant event instead of the accident event. In the present context of aerodrome traffic control, the Runway Incursion (RI) is a precursor of an accident event and is as such selected as a risk indicating event, defined by ICAO Doc. 4444 as following:

> *„Any occurrence at an aerodrome involving the incorrect presence of an aircraft vehicle or person on the protected area of a surface designated for the landing and take-off of aircraft” (ICAO, 2007)*

This is a valid assumption due to the statistic fact that the occurrence of accidents and Runway Incursion comply with a ratio of 1:100. This implies an accident rate of one every 3.7 years (EUROCONTROL, 2006).

Experimental Tasks and Simulation Scenarios

The chosen HITLS consists of test persons that operate a Surface Manager HMI as the primary working device (Figure 5). The device complies with the Eurocontrol A-SMGCS Implementation level 3 (Adamson, 2005), with the functional exception of a missing device that prevents RI (Runway Incursion Prevention and Alerting Systems, RIPAS) automatically. Tasks to be performed by the test persons are defined by ICAO Annex 11 (ICAO, 2001, S. 3-1) and ICAO PANS-ATM doc. 4444 (ICAO, 2007) for tower and ground control services. The Surface Manager HMI allows for the selection of a target aircraft by pen strokes, as well as granting pushback, taxi, lineup or take-off clearances on an airport surface surveillance radar screen presenting the entire traffic situation at Frankfurt airport.

The generated traffic consists of inbound and outbound a/c traffic movements at Frankfurt airport (ICAO code: EDDF) on three active runways (RWY) in direction 25, operating 25L as a landing only runway, 18 as take-off only runway and 25R in mixed mode. This complies with the old operational concept before runway north started for operations. Runway dependencies can be found between RWY 18 and RWY 25R, as well as between RWY 18 and RWY 25L. The dependency between 25R and 25L was considered according to the reduced runway separation and semi-mixed parallel runway operations. The random traffic generator distributes initially 160 movements over 240 simulated minutes per execution run according to a given set of stochastic parameters with uniformly distributed destination routes or departure gates (including north and south area stands) and runways. We accelerated the simulation speed by a factor of two, hence the simulation time is double the real time. The routes of the movements are initialized by the *Floyd und Warshall* algorithm, optimizing routes, which according to a given operational concept, ensure a similar task load for all experimental executions. The simulated a/c agents are capable of separating from one another on taxiways and to solve taxi obstruction conflicts as well as taxi crossing situations autonomously according to ICAO Annex 2 (ICAO, 1990). The execution scenario assumes one controller for both ground and tower controller tasks controlling the whole airport and inducing a higher task load than realistic scenarios would do.

The concept of *ComPerf* was adapted to the experiment by the application of a simple controller-agent that is capable to act as an ATCO who is allowed to grant clearances. The evaluation of the agent's decisions by a traffic-movements predictor effects the resulting operation to be verifiably conflict-free. No prioritization is implemented, since the agent handles all movements simultaneously and independently. The agent is configurable by a reaction time t_r per clearance that calibrates the performance concerning the number of aircraft handled per time. By setting t_r, the decision-making of the controller-agent gives a controllable advantage to the human operator in the context of the performance comparison of *ComPerf*. The humans operators time necessary for decision making t_n is hence set into competition with t_r by which the time-budget principles are established when defining t_r=t_a. Setting a desired rapidness t_r of decision-making can be consequently assumed as a target load for the human operator.

The absolute number of traffic movements, which leaves the simulated operation, has been chosen to be the principle performance metric for *ComPerf*. Leaving the system is defined by the moment of (1) granting the clearance for take-off for outbound movements or (2) granting the last taxi clearance before entering the aircraft stand. The comparison calculates the performance head start by comparing these metrics. Presuming the test person would refuse any action, a time can be calculated for which the head start becomes zero. This can be regarded as a quantified head start, calculated on the basis of a fast-time simulation of the controller-agents world that establishes the complete agents- timeline, including timestamps of all operational events, in very little time.

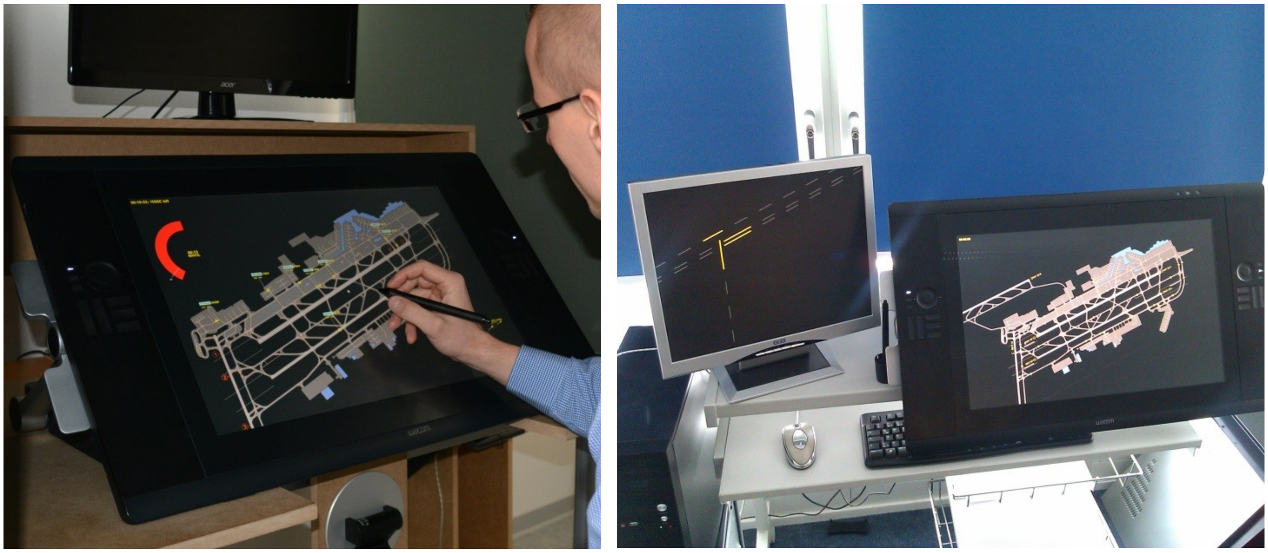

Figure 5. The Surface Movement Manager HMI consists of a ground surveillance of the airport and a secondary surveillance radar of the vicinity of the airport.

The countdown was visually and acoustically feed backed to the human operator by the visualization of a clock on the ground surveillance display (Figure 6) and by an alarm noise. The noise indicated the head start remaining, graded from 300 to 180, 30 and 10 simulated seconds, accompanied by increasing loudness. The head start of zero was accompanied by an unpleasant alarm noise, indicating the Time Error and the activation of the punishment rule. The visualization of the head start consisted of a circle-like clock that covered 6 Minutes as a full circle with a logarithmic time-axis.

Figure 6. The clock on the ground surveillance display feed backs the head start to the operator.

The activation of the punishment rule was implemented as growth of the aircraft-queue by two. This consequently increased the corresponding duration of the experiment indirectly by the time necessary to handle them. Further, the simulated world of the agent is synchronized with the test person's world to reset the boundary conditions of the simulation after punishment. Consequently, the duration of the experiment lies in the test person's hand. This mechanism is regarded as a sufficient measure of punishment, since we presume that all test persons are motivated to finish the simulation in time and to successfully compete with the controller-agent, according to the hypothesis.

For varying gradations of load, we defined a t_r-trajectory, the purpose of which was to calibrate the time pressure by the excitation of a step-response and to stepwise decrease t_r in various steps (Figure 7).

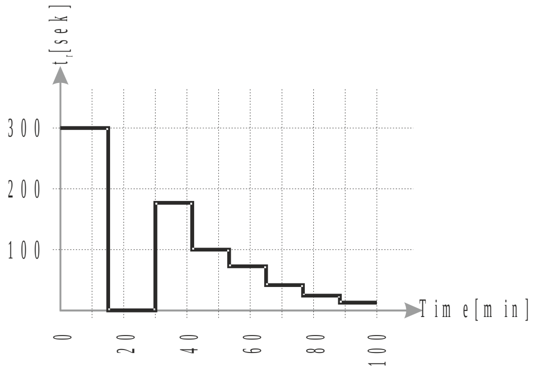

Figure 7. The calibration scenario varies the reaction time t_r following a target trajectory.

Test Persons and Training

For the empiric study, we acquired 3 students of the study program "Transport engineering" in the 4th year of their diploma as novice test persons. We educated them according to the tasks described above and trained them by means of the test setup. Every test person successfully completed a training consisting of 10 hours and final tests that indicate whether the rules of runway separation can be mastered according to the trained procedures.

Measurements

The measurements consisted of three metrics, namely work load, time pressure and the frequency of Runway Incursion, which fulfilled our requirements to show reactions to the gradations of load according to our hypotheses.

Firstly, we measured the current work load and time pressure by frequently interrogating the test persons (3 minutes interval). For interrogation, we used the Integrated Workload Scale (IWS) (Pickup, Wilson, Norris, Mitchell, & Morrisroe, 2005) in a cut down version of 8 of its gradations (Table 1).

Table 1: Used scales for the estimation of the work load and time pressure

scale	work load	time pressure
7	too demanding	too high
6	extreme effort	extreme
5	high effort	High
4	acceptable effort	Acceptable
3	moderate effort	Moderate
2	little effort	Slight
1	minimal effort	Minimal
0	no effort	no TP

Secondly, we recorded runway incursion events as the principle safety-metric during the experiment. The Runway Incursion was automatically detected as soon as rules of the reduced runway separation minima and parallel runway operations described in ICAO PANS-ATM Doc. 4444 (ICAO, 2007) were violated.

RESULTS

Work Load and Time Pressure

The data from the work load and time pressure measurements were analyzed regarding their sensitivity to the reaction time t_r. The answers of the test persons indicated a reaction on the time set in which the work load as well as the time pressure increased when increasing the load (Figure 8). In all executions, the work load increased at the beginning of the experiment, what can be assumed as a transition phase in which the aircrafts start to enter into the system. The work load converges to an average value that corresponds to a balanced throughput of aircraft handled by the test person. In contrast, the time pressure followed the steps of load accompanied by a relaxation time of between 10 and 15 minutes. With increasing load, the test persons stated indeed an increasing time pressure that motivated to accelerate decision making as possible.

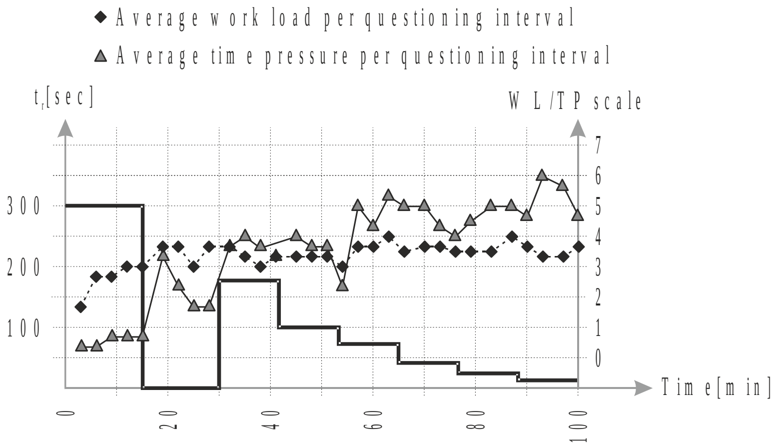

Figure 8. Work load and time pressure measurements over the experiment

The regression analysis shows a dependency between t_r and the resulting work load and time pressure (Figure 9). Fitting the measurements linearly, the rise of time pressure is 10 times higher than the work load and quantifies the strong relation between time pressure and the reaction time set in the controller-agent. With the help of the test persons' statements, a change of the decision strategy could be identified. Firstly, this addresses the order in which aircrafts were handled when test persons tried to minimize handling times with the surface. The observed strategy followed increasingly short-term objectives instead of using options to optimize the long-term throughput. While trying to increase the throughput, the acceptance of the possibility to commit errors increased.

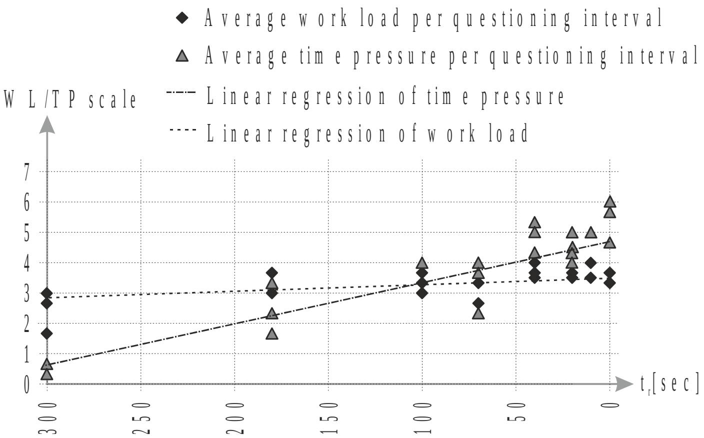

Figure 9. Work load and time pressure over t_r in a linear regression analysis

The test persons stated the importance to see with their own eyes how the controller-agent performs. A demonstration of the capabilities was regarded as fundamental for the motivation of the test person by the assurance to compete under equal conditions.

Runway Incursion and Time Error

Besides the subjective questionnaire, the objective measurements showed a two-fold picture. This can be observed by the reaction of the frequency of Runway Incursion and the time error that occurs when the head start falls off to zero (Figure 10). For this reason, the diagrams show the test persons reactions individually. Test person B showed an enormous increase in the runway incursion caused by errors in the line-up and take-off clearance. By comparison, the corresponding time error is in contrast little during the experiment. The test persons A and C showed a contrary behavior in which the Runway Incursion remains in little relations to the time error.

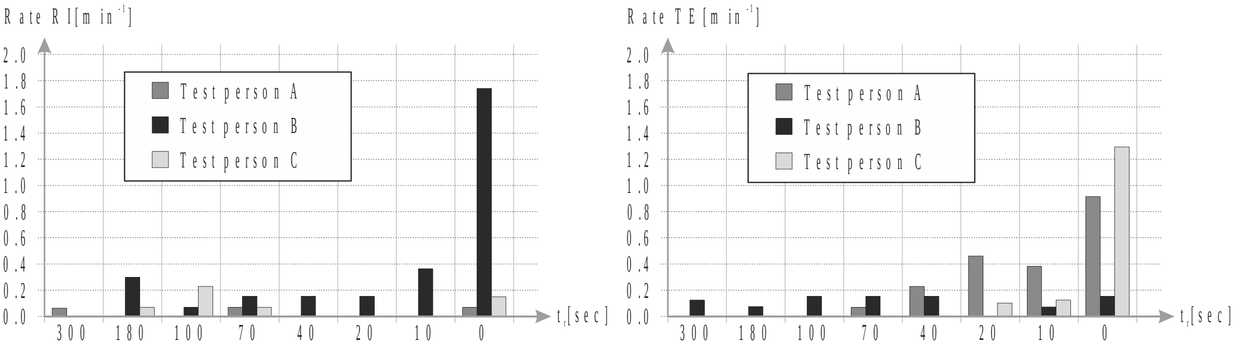

Figure 10. Rates of Runway Incursion over t_r (left) and Rates of Time Error over t_r (right)

This clearly points to the prioritization of the test persons, to decide in favor of conflict-free operations or to successfully compete with the controller-agent. This can be a sign of a personal attribution of each test person whether sensitivity to time pressure is given or not. Another explanation takes into account the contribution of the test person’s inability to link the simulated reality to real operations, which have the potential for causing serious damage. This is known as missing awareness for consequences of novice test persons.

CONCLUSIONS

We have developed a concept for risk analysis that aims on modifying conditions of the HITLS to accelerate the occurrence of safety-relevant events by means of a calibrated time pressure induction. The experimental setup represents an adaption of the concept in the scope of a proof-of-concept study. The results deliver insights on the relation between the time pressure induced, human performance and the resulting frequency of Runway Incursion. The results do not confirm the risk model (Figure 3) completely, due to the expected increase of the frequency of Runway Incursion to the time pressure induced, which has not been observed clearly. In contrast, the subjective reactions matched our expectations by the demonstrated ability to control the time pressure independent from the work load. The causes can be very practical ones and might be found in the impact of novices’ behavior in a simulated environment or in insufficiencies in the concept of *ComPerf* which has become a rather complex mechanism. The cost/benefit balance of the test persons is regarded as a good explanation for the relations between the occurrence of Runway Incursion and Time Error in which humans might be fully aware of committing rule violations in order to comply with a system of reward. Additionally, we observed decreasing error rates even under highest level of load, what indicates the ability of test persons to optimize their behavior even after 30 hours of simulated hours.

A valuable outcome of the concept is the possibility to provide an effective training environment for optimizing controller activities by setting controllable gradations of stress. The *ComPerf* comparison quantified the time necessary t_n of the human for decision-making to between 15 and 25 seconds.

REFERENCES

Adamson, P. (2005). Definition of A-SMGCS Implementation Levels. Brussels, BE: Eurocontrol.

Blom, H. A., Bakker, G. J., Blanker, P. J., Daams, J., Everdij, M. H., & Klompstra, M. B. (2001). Accident risk assessment for advanced air traffic management. Progress in Astronautics and Aeronautics, S. 463-480.

Blom, H., & Bakker, B. (2003). Stochastic analysis background of accident risk assessment.

Bubb, H. (2005). Human Reliability: A key to improved quality in manufacturing. Human Factors and Ergonomics in Manufacturing & Service Industries 15.4, S. 353-368.

Bubb, H., & Jastrzebska-Fraczek, I. (1999). Human Error Probability depending on Time Pressure and Difficulty of Sequential Tasks. Safety and Reliability, S. 681.

Chang, Y. H., & Mosleh, A. (2007). Cognitive modeling and dynamic probabilistic simulation of operating crew response to complex system accidents. Part 2: IDAC performance influencing factors model. Reliability Engineering & System Safety, S. 1014-1040.

Cox, D. R. (1972). Regression models and life-tables. Journal of the Royal Statistical Society - Series B (Methodological), S. 187-220.

Edland, A. (1993). Judgment and Decision Making Under Time Pressure. In Time Pressure and Stress in Human Judgment and Decision Making. Plenum Press.

EUROCONTROL. (2006). Final Report on the Generic Cost Benefit Analysis o. Brussels, BE.

Eurocontrol. (2006). MAIN REPORT FOR THE: 2005/2012 INTEGRATED RISK PICTURE FOR AIR TRAFFIC MANAGEMENT IN EUROPE. Brussels,BE.

Freedman, J. L., & Edwards, D. R. (1988). Time pressure, task performance, and enjoyment. The social psychology of time: New perspectives, S. 113-133.

Freedman, J. L., Edwards, D. R., & McGrath, J. E. (1988). Time pressure, task performance, and enjoyment. The social psychology of time: New perspectives, S. 113-133.

Hollnagel, E. (1993). Human reliability analysis: Context and control. London, UK: Academic Press.

ICAO. (1990). Rules of the air - 9th edition. Montreal,CA: ICAO.

ICAO. (2001). Air Traffic Services - Annex 11. Montreal,CA: ICAO.

ICAO. (2007). PANS-ATM Air Traffic Management - Doc 4444. Manual doc4444. Montreal, CA: ICAO.

Johnson, C. (1999). Why human error modeling has failed to help systems development. Interacting with computers 11.5, S. 517-524.

Kerstholt, J. H. (1994). The effect of time pressure on decision-making behaviour in a dynamic task environment. Acta Psychologica 86.1, S. 89-104.

Kirwan, B. (2007). Technical Basis for a Human Reliability Assessment Capability for Air Traffic Safety Management. Paris,FR: Eurocontrol.

Malhotra, D. (2010). The desire to win: The effects of competitive arousal on motivation and behavior. Organizational Behavior and Human Decision Processes, S. 139-146.

Müller, M. (2004). Risk and risk management in aviation. Zeitschrift für ärztliche Fortbildung und Qualitätssicherung, S. 559.

Nelson, B. (2004). Accelerated testing: statistical method, test plans, and data analysis. New Jersey: Wiley.

Pickup, L., Wilson, J. R., Norris, B. J., Mitchell, L., & Morrisroe, G. (2005). The Integrated Workload Scale (IWS): A new self-report tool to assess railway signaller workload. Applied Ergonomics, 36(6), S. 681-693.

Rastegary, H. (1993). The Interactions among Time Urgency, Uncertainty, and Time Pressure. In Time pressure and stress in human judgment and decision making. Plenum Publishing Corporation.

Reason, J. (1990). Human Error. New York,USA: Cambridge University Press.

Shapell, S., & Wiegmann, D. (1996). U.S. naval aviation misshaps 1977-92: Differences between single- and dual-piloted aircraft. Aviation, Space, and Environmental Medicine, S. 65-69.

Shorrock, S. T. (2001). Assessing human error in air traffic management systems design: methodological issues. Le travail humain, S. 269-289.

Stroeve, S. H., Blom, H., & Bakker, G. (2013). Contrasting safety assessments of a runway incursion scenario: Event sequence analysis versus multi-agent dynamic risk modelling. Reliability Engineering & System Safety, S. 133-149.

Swain, A. D., & Guttman, H. E. (1983). Handbook of human reliability analysis with emphasis on nuclear power plant applications. NUREG.