Accelerated Risk Analysis in ATM: An Experimental Validation Using Time Pressure as a Stressor

Lothar Meyer
Chair of Air Transport Technology and Logistics
TU Dresden
Dresden, Germany
meyer@ifl.tu-dresden.de

Katja Gaunitz
Student of Transport Engineering
TU Dresden
Dresden, Germany
katja.gaunitz@mailbox.tu-dresden.de

Hartmut Fricke
Chair of Air Transport Technology and Logistics
TU Dresden
Dresden, Germany
fricke@ifl.tu-dresden.de

***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. 1.55E-8 accidents per operating hour in which safety metrics are subject to the statistic problem of Right Censoring. This paper presents our novel concept which modifies the conditions of the simulation in order to create a calibrated acceleration effect with respect to error rates, allowing the estimation of statistically small safety metrics during considerably shorter experimental periods. This is motivated by the Accelerated Life Testing methodology from reliability assessment, which serves to determine the Mean-Time-To-Failure of products by fast-forwarding the experimental period with calibrated steps of increased stress-load. For ATM Human-In-The-Loop simulations, this effect is achieved with an experimental design that induces a calibrated time-pressure in order to stimulate human errors. The results of this proof-of-concept-study show controllable stress-reactions by the test persons in the scope of proving the internal validity of the concept.**

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

I. Introduction

Current methods for estimating the risk of socio-technical systems in ATM mostly rely on accident and incident reports, expert judgment or model-based approaches. In particular, the predictive risk estimation of 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 definition: *"Risk is defined as the probability that an accident occurs during a stated period of time"* [1].

The most promising model-based approaches offer the advantage of coping with enormous sample spaces, by 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 [6]. An exhausting validation of all modeled a-priori assumptions regarding the safety effects on new design in realistic operating conditions are extremely hard as there are usually no means of obtaining and transferring direct evidence from current systems and operations: *"errors are likely to be made when designers apply error modeling techniques"* [2]. This might impair the external validity of the model for unknown or unexpected 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 [3]. HITLS has also been successfully used to accompany FMEA studies, namely 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 [4]. 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 [5]. When using valuable experts, most studies perform in order of a few hundred hours of simulation time at best [4], providing insufficient statistical power for a reliable elimination of rare and risk-inducing events (see ATM safety-iceberg [6]). When applying statistical testing, 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 C*onvergence in Probability*, or more specifically, Bernoulli's theorem. It describes a decreasing difference between observable frequency and the true probability with increasing sample spaces. The difference can be assumed to describe the type I error rate, which can be estimated with the *Chebyshev's inequality*. When assuming one operating hour as a unit that could have the end state *accident* or *no-accident*, underlying a *Binomial Distribution*, the error can be estimated as

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

with the random variable *X*, the mean μ, the variance of the distribution $\sigma^2 = n \cdot p \cdot (1-p)$ and the confidence tolerance level *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% confidence. 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 addresses the problem definition of Hollnagel: "*the problem remains of how raw data from training simulators can be modified to reflect real-world performance.*" [7].

Therefore we developed a concept called *Accelerated Risk AnalySis* (*AccSis*) that describes a methodology to gain the desired acceleration-effect needed for intensifying the probability for safety metrics. This acceleration effect shall practically be reached by the induction of a calibrated time pressure that stimulates the occurrence of human error. Concerning the time pressure induction, we developed a procedure following the *Time Budget*-principles [8][9], named *Competitive Performance* (*ComPerf*). It puts the test person under the impression of not having sufficient time to solve the problem [10]. This approach is motivated by the *Accelerated-Life-Testing* (ALT) methodology that forwards the *Mean-Time-To-Failure* into the experimental period by means of accelerated and calibrated stress-induction during the experiment [11]. It explicitly addresses the occurrence of *Right-Censoring* [12].

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. This paper catches up findings identified in the results of the HITLS and delivers insights on the effects of the load induction procedure indicated by means of workload measures and the detected runway incursion frequencies, all of which is outlined in chapter 4.

II. METHODOLOGY

A. *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 $\hat{k} = n \cdot \hat{p}_{target}$. 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.

In reliability testing, 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 [11]. 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).

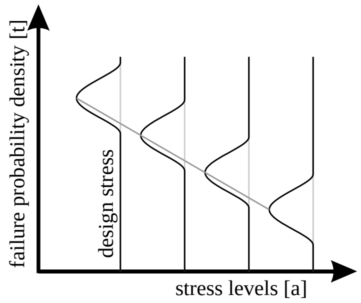

Figure 1. Stress-life-relation according to ALT concept [11].

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 failed operator's decision which hence is regarded as an independent event with limited temporal relation to operational preceding actions and in which a distribution cannot be modelled over time when assuming a 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 state-of-the-art models, although describing the relationship between human error and stress, fail to deliver a domain-specific model curve (e.g. exponential-linear)

This paper considers AccSis to be the subject of a long term validation strategy due to the reasons given above. In that scope, 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 shall be achieved. Based on this consideration, we chose the first step to be the proof of internal validity: the controlled acceleration of safety relevant events by intensifying human error.

In order to explain our choice of human errors as the key factor, we refer to 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 [14]. For a socio-technical system, the IRP can be regarded as a significant fingermark of risk in which branches of failure catenation form the resulting accident probability. One has to consider that only branches affected by the acceleration effect are taken into account for the regression-analysis.

When considering causal factors in this context, organizational, technical and human errors can be distinguished as principle accident causes. This complies with Reasons "a trajectory of accident opportunity" that models the human error propagation in the presence of corresponding hazards as Unsafe Acts [15]. The human error has been identified as the most frequent contribution to accidents and incidents in aviation with a share of 60% to 80% [16] or 75% respectively [17]. The focus on 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 must hence be covered by acceleration. Following the ceteris-paribus-principles, procedures, tasks and other boundary conditions are to be held constant during HITLS which implies a major requirement on seemingly unimportant contextual conditions of the simulation.

B. Time Pressure for the stimulation of human error

Besides uncertainty, time pressure seems to be of particular relevance when considering human decision-making processes [18]. Rastegary defines time pressure *"…as the difference between the amount of available time and the amount of time required to resolve a decision task"*.

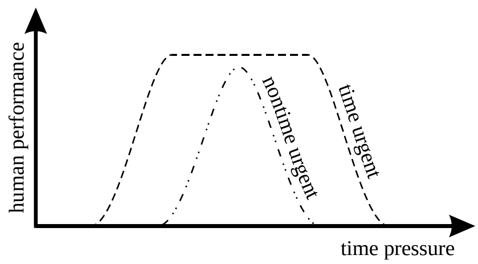

Figure 2. human performance as a function of time pressure [19].

By empiric findings, time pressure is known to significantly affect human performance [19] (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 [20].

Time pressure contributes more to Human Error Probabilities (HEP) than additional tasks when performing time critical tasks [8]. 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 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 [21].

Time pressure and human error are causally linked and can be transferred to a continuous quality metric for human actions that is ultimately classifiable in acceptable or not acceptable. Specifically, the deflection of actions below a minimum quality can be regarded as not acceptable or, in line with conventional theories, human error. Continuing, 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* [9]. 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 [15], [8] and [19].

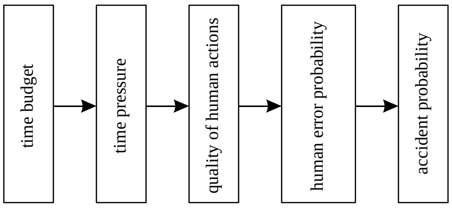

Figure 3. Causal relationship between time budget and the accident probability

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

C. Competitive Performance

Most ideas for the implementation of time pressure-induction aim at setting boundary conditions that effectively shorten the time available. Secondary tasks might, for example, shorten t_a by forcing the operator to organize task sharing and prioritization. This sharing will as well change the pattern of activities and impact the IRP picture without any control of the deflection from the design stress. The same holds true for the conventional mean of HITL calibration, namely the intensification of task load, e.g. traffic volume.

As time pressure is transformed from objective condition to a subjective feeling, we decided to choose the approach of "competitive arousal". Here, time pressure is generated by providing a competitive environment that triggers the desire of the operator to win [22][23]. Our concept establishes "competitive arousal" by forcing the operator to compete with a “calibrated reference operator” that operates under the same contextual conditions (cloned worlds) and is capable of acting according to a calibrated performance (see Figure 4), named Competitive Performance (ComPerf). When the human operator acts, his or her performance metric, e.g. the throughput of the system, is measured and fed back for instant comparison. The head start is the quantified indicator for the performance of the human operator compared to the Reference Operator. The Reference Operator in this instance is a model-based software agent that supports gradations of performance.

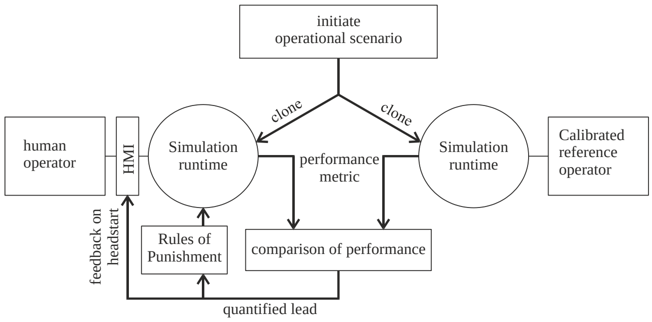

Figure 4. The concept of Competitive Performance (ComPerf).

If the lead of the human operator shrinks below a given threshold, a hard penalty applies in order to challenge the test persons to compete as hard as possible. In this instance, the effort needed to successfully finish the scenario was increased by generating additional tasks or enlarging time constraints like the scenario’s finish time. As expressed before, the implementation must carefully compensate the changed boundary conditions stemming from applied penalties in order to achieve constant and comparable contextual conditions. The advantage of controlling the time available t_a by varying the performance of the Reference Operator and thus establishing the time-budget principles in relation to the decision times of the human operator t_n is the inherent feedback loop that highly suited for automatic tuning of perceived time pressure.

III. 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 (ATCo) has been chosen as an exemplary safety critical working environment within 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 [24] . 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)…".

The hypotheses were set as follows:

- The time needed is sensitive to the target load set by *ComPerf*.
- The relative frequency of safety relevant events is sensitive to the target load 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 an 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"* [24].

The notion of precursor-ship is backed by safety management principles and the statistical understanding that the occurrence of collision accidents relates to Runway Incursions in a ratio of 1:100, which would, by the way, imply a runway collision accident rate of one every 3.7 years [25].

A. 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 [26], 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 [27] and ICAO PANS-ATM doc. 4444 [24] 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 (ICAO code: EDDF).

The generated traffic consists of inbound and outbound a/c traffic movements at Frankfurt airport on the 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 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 initially distributes 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. The routes of the ground movements are initialized by the *Floyd und Warshall algorithm*, which optimizes routes according to a given operational concept and ensures a similar task load for all experimental executions. The software aircraft/pilot agents are capable of self-separating on taxiways and to solve taxi obstruction and crossing conflicts autonomously according to the rules laid out in ICAO Annex 2 [28]. The execution scenario demands that the test persons work both ground and tower positions in parallel, i.e. control the whole airport at a severely increased task load (160 movements at 2 x real time).

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 human 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.

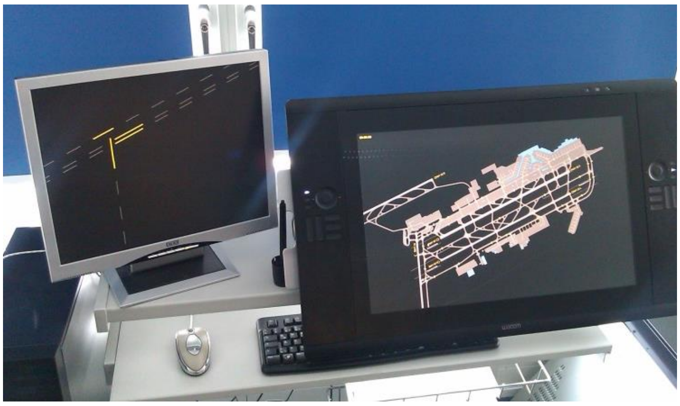

Figure 5. The Surface Movement Manager HMI consists of a ground surveillance of the airport and a secondary surveillance radar of the vicinity of the airport.

The absolute number of traffic movements, which depart from the simulated airport or reach their designated stand, 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 lead of the human operator by comparing these metrics to the autonomous software competitor. Presuming the test person would refuse any action, a time can be calculated for which the lead 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.

The countdown was visually and acoustically fed back 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 lead time, graded from 300 to 180, 30 and 10 simulated seconds, accompanied by increasing playback volume. A lead of zero was accompanied by an unpleasant alarm noise, indicating the Time Error condition that results in penalty. 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.

The penalty was implemented as an increase of the aircraft-queue by two additional movements. This consequently increased the duration of the experiment indirectly by the time necessary for handling and finalizing the movements. As the simulated world of the agent is synchronized with the test person's world the duration of the experiment effectively lies in the test person's hand. This mechanism is regarded as a sufficient measure of motivation for winning the competition, since we presume that all test persons are not only motivated to successfully compete with the controller-agent but moreover to finish the simulation in time (and be done).

Figure 6. The clock on the ground surveillance display feedbacks the lead to the human operator.

B. Test persons and training

For the empiric study, we acquired three 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 indicated whether the rules of runway separation can be mastered according to the trained procedures.

C. Measurements

The measurements consisted of three metrics, namely the time necessary, the frequency of Runway Incursion and the frequency of Time Errors, which fulfilled our requirements to capture reactions to the gradations of load according to our hypotheses.

Firstly, we recorded runway incursion events as the principle safety-metric during the experiment. Runway incursion were automatically detected as soon as rules of the reduced runway separation minima and parallel runway operations described in ICAO PANS-ATM Doc. 4444 [24] were violated. Secondly, the time necessary t_n is regarded as a correlating dependent of the cognitive decision-time and is the measured time period from requesting the clearance by the aircraft till granting the aircraft by the human operator. Thirdly, the frequency of Time Error (TE) was recorded, quantifying the number of penalties applied when the lead is zero and the time budget is hence >1.

D. Calibration of target load

The calibration procedures were performed prior to the experiment and consisted of a trajectory that varies t_r over time through a predefined bandwidth between 0s and 150s. The calibration procedure is explained in more detail in [13]. Two target load levels were quantified as parameters for the controller-agent, defining two experimental configurations (Table I).

TABLE I. TARGET LOAD PARAMETERS

ComPerf A/B	t_r[s]
A	30
B	20

E. Executive planning

The experiments were conducted according to a sequence plan that varies the configurations and its target load in a systematic order. The necessity for a comparison of the results for learning effects of the human operator can thus be expected.

IV. RESULTS

A. Time necessary

According to the stated hypothesis for correlation, it is expected that decreasing reaction times of the controller agent (increasing target load) effect an accelerated working speed of the human operator (hence, decreased time needed t_n). With respect to this expectation, the three test persons show unclear reactions in the time needed to grant clearances. This is indicated by the measurements (n > 1000), illustrated in Figure 7, which contrasts t_n as box-plots according to the selected target loads for each test person. The measurements of test persons A and B indicate tendencies of accelerated working speed. In contrast, test person C shows a tendency to maintain his or her working speed.

TABLE II. MANN-WHITNEY-U RANK-SUM TEST

Test person	p-value %
A	1.11
B	9.89
C	5.09

For testing these observations objectively, the Mann-Whitney-U test provides a probability (p-value) for two independent non-parametric samples on its central tendency to belong to the same population.

The test results (Table II) show no clear rejection of the null-hypothesis for all test persons. Only the distribution of test person A exhibits a significant increase in reaction time, indicated by a value of $p < 5\%$. Test person B shows the same tendency. The reaction of test person C is contrary to A and B.

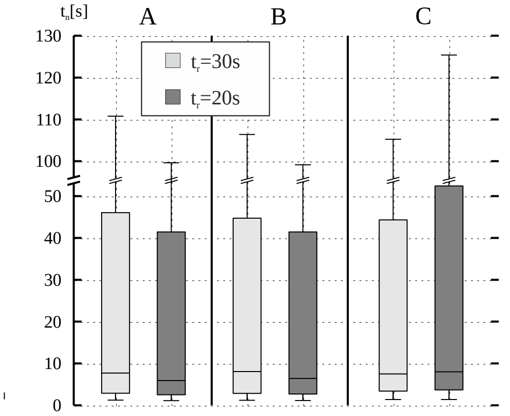

Figure 7. Reaction times t_n over variing target loads t_r.

B. Runway Incursion

Runway Incursions (RI) were measured as an absolute frequency per target load and test person. The frequency was divided by the number of take-off clearances granted by the human operator. This shall compensate for varying periods of the execution scenarios due to the extensions by the applied penalties. A common tendency can be found through all measurements (Figure 8). This confirms a sensitivity of the target load on the resulting frequency of safety relevant events. Therein, test person B shows the largest increase and C the lowest, indicating degradation on the quality of decision making.

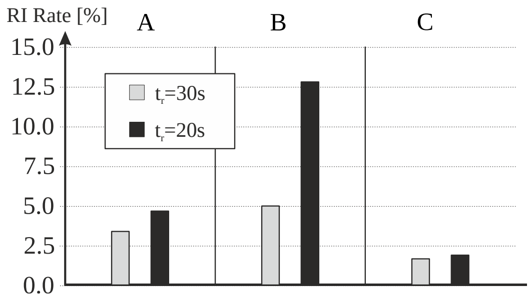

Figure 8. Runway incursion rate.

The measured frequency of RI was subject to a learning curve, indicating an increase in competence and hence quality over the course of the experiment.

C. Time error and time budget

The absolute frequency of TE indicates the compliance of the working speed to the given target load. This is also a measure that shows the ability of the human operator to respond to the time pressure induced. The test persons show a two-track reaction on increasing load (Figure 9). Test persons A and C showed less reactions on increased target load than test person B, while test person C shows a smaller overall susceptibility to the induction procedure. As the human operator, permitting a higher frequency of time errors might be an attractive mean to effectively extend the time available t_a while accepting that the experimental period is extended by the penalty. Thus, a correlation between the frequency of TE and the mean TB (Table III) can be expected.

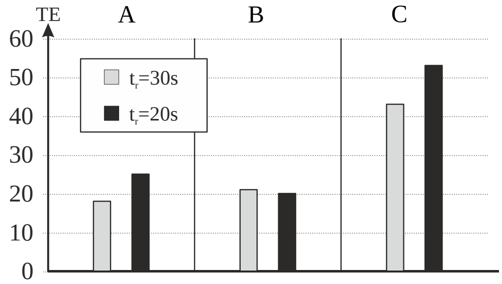

Figure 9. frequency of time errors.

Dividing the samples of t_n by the reaction time t_r delivers the time budget samples whose mean values are summarized in Table III.

TABLE III. MEAN TIME BUDGETS

	A		B		C	
t_r[s]	***30***	***20***	***30***	***20***	***30***	***20***
Mean TB	1.30	1.82	1.29	1.76	1.37	2.20

The Spearman-correlation rank coefficient was quantified as 60% (p-value: 20.8%). Even when no significance can be proven, it provides an indication of a strong relation between TE and TB.

V. CONCLUSION

In summary, the results show clear reactions of increased stress and lowered quality for all test persons. An increased uncertainty during decision making can be concluded from the data observed. It also shows a strong dependency to the individual, forcing the human operator to decide between quality and working speed in a subjective speed-accuracy-trade-off that is in line with the definition of *Fitts' law* [29]. This decision is exemplarily illustrated in Figure 10 which shows individual operating points on the RI over TE scale, effectively balancing the available performance between the two claims of the task definition.

A plausible explanation can be given by the relation between quality, performance and time needed, introduced in section II B. Assuming the RI rate as a reciprocal metric of quality and the frequency of TE as a valid measure of the time budget (cf. correlation test), the relationship can be expressed as

$$\frac{t_n}{Q} = \frac{1}{P} \Leftrightarrow RI_{rate} \cdot TB \cdot t_r = \frac{1}{P} \Leftrightarrow \boxed{RI_{rate} \cdot TE = \frac{1}{P \cdot t_r}}.$$

This term describes the product of t_r and P as the reciprocal function of a surface, with the measured factors forming the dimensions of the related rectangle. With a view on Figure 10, the shift of the operating points of the test persons A and C approximately follow this relationship, leading to a measure of performance that takes into account the mentioned trade-off effect and the target load t_r. Test person B obviously switched the priority when the target load increased, indicated by the proportions.

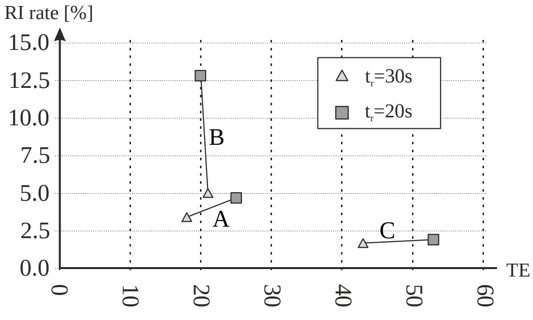

Figure 10. Speed-accuracy-trade-off

The success of the time pressure-induction was investigated regarding the internal validity of the relations shown in Figure 3. The common tendency indicates an increase of the probability of safety relevant events and human error when increasing the target load (Figure 8). The related sensitivity strongly depends on the trade-off effect between quality and working speed. Assuming the performance to be sufficiently constant, the operating point potentially shifts into an insensitive interval which does provide less metrics regarding an Accelerated Risk Analysis (cf. test person C). On the other hand, being insensitive for time pressure arousal is an essential mark of quality of the air traffic controller. Thus, permitting TE complies fully with the completed training, qualifying the test person to act corresponding to his role model. The claim for quality seems to be at odd with the principles of Accelerated Risk Analysis. The term ‘Acceleration’ means in this case that the method accelerates an uncertainty of decision quality which characterizes human behavior fundamentally. As a part of the human character, the trade-off is to be considered an elementary part of the authentic working environment. As the trade-off is closely linked with time pressure, it becomes subject of the acceleration as well. This can be best observed in the time needed of test person C (Figure 8), expressing robustness against the induction by maintaining best the working speed. This complies with Figure 2, classifying test person C as a time urgent individual whose breakdown point provides a constant performance over a larger interval. The opposite is represented by test person B which shows strong reactions to the induced time pressure.

Concluding, our further investigation will focus on the stabilization of the stress response and taking into account the trade-off effect.

REFERENCES

[1] H.A.P. Blom, and G. J. Bakker, "Stochastic analysis background of accident risk assessment for Air Traffic Management," IFAC Conference on Analysis and Design of Hybrid Systems. 2003.

[2] C. Johnson, "Why human error modeling has failed to help systems development," Interacting with computers 11.5 (1999): pp. 517-524.

[3] B. Kirwan, "Technical Basis for a Human Reliability Assessment Capability for Air Traffic Safety Management," Eurocontrol EEC Note No. 06/07, 2007.

[4] S.H. Stroeve, H.A.P. Blom, and G. J. Bakker. "Contrasting safety assessments of a runway incursion scenario: event sequence analysis versus multi-agent dynamic risk modelling," Reliability Engineering & System Safety 109 (2013): pp. 133-149.

[5] S.T. Shorrock, "Assessing human error in air traffic management systems design: methodological issues," Le travail humain 64.3 (2001): pp. 269-289.

[6] H.A.P. Blom, G.J. Bakker, P.J.G. Blanker, et al. "Accident risk assessment for advanced air traffic management," Progress in Astronautics and Aeronautics 193 (2001): pp. 463-480.

[7] E. Hollnagel, "Reliability analysis and operator modelling," Reliability Engineering & System Safety 52.3 (1996): pp. 327-337.

[8] H. Bubb, and I. Jastrzebska-Fraczek, “Human error probability depending on time pressure and difficulty of sequential tasks,” Safety and Reliability 10th Conference, 1999, pp. 681-686.

[9] H. Bubb, "Human Reliability: A key to improved quality in manufacturing," Human Factors and Ergonomics in Manufacturing & Service Industries 15.4, 2005, pp. 353-368.

[10] Y.H.J. Chang, and A. Mosleh, "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 92.8, 2007, pp. 1014-1040.

[11] W.B. Nelson, “Accelerated testing: statistical models, test plans, and data analysis,” Vol. 344. John Wiley & Sons, 2009.

[12] D.R. Cox, "Regression Models and Life-Tables," Breakthroughs in Statistics. Springer New York, 1992, pp. 527-541.

[13] L.Meyer, K. Gaunitz, and H. Fricke, “Investigating time pressure for the empirical risk analysis of socio-technical systems in ATM,” Advances in Human Aspects of Transportation: Part I, 2014, pp. 33-45

[14] J. Spouge, and E. Perrin, “2005/2012 Integrated Risk Picture for Air Traffic Management in Europe,” Eurocontrol EEC Note 2006/05, 2006.

[15] J. Reason, “Human error,” Cambridge university press, 1990.

[16] S.A. Shappell, and D.A. Wiegmann, "US naval aviation mishaps, 1977-92: differences between single-and dual-piloted aircraft," Aviation, Space, and Environmental Medicine 67.1, 1996, pp. 65-69.

[17] M. Müller, "Risk and risk management in aviation," Zeitschrift für ärztliche Fortbildung und Qualitatssicherung 98.7, 2004, pp. 559-565.

[18] H. Rastegary, and F.J. Landy, "The interactions among time urgency, uncertainty, and time pressure," Time pressure and stress in human judgment and decision making. Springer US, 1993, pp. 217-239.

[19] J.L. Freedman, and D.R. Edwards, “Time pressure, task performance, and enjoyment,” The social psychology of time: New perspectives, 1988, pp. 113-133.

[20] A. Edland, and O. Svenson, "Judgment and decision making under time pressure," Time pressure and stress in human judgment and decision making, Springer US, 1993, pp. 27-40.

[21] A.D. Swain, and H.E. Guttmann, “Handbook of human-reliability analysis with emphasis on nuclear power plant applications,” Final report, Sandia National Labs., Albuquerque, NM (USA), 1983.

[22] D. Malhotra, "The desire to win: The effects of competitive arousal on motivation and behavior," Organizational Behavior and Human Decision Processes 111.2, 2010, pp. 139-146.

[23] J.H. Kerstholt, "The effect of time pressure on decision-making behaviour in a dynamic task environment," Acta Psychologica 86.1, 1994, pp. 89-104.

[24] ICAO, “Procedures for Air Navigation Services - Air Traffic Management (PANS-ATM) - Doc 4444,” 15th edition, Montreal (CA), 2007.

[25] M. Birenheide, “Generic Cost-Benefit Analysis of A-SMGCS Levels 1 and 2 ,” Eurocontrol Brussels (BE), 2006.

[26] M. Birenheide, “Definition of A-SMGCS Implementation Levels,” Eurocontrol Brussels (BE), 2010.

[27] ICAO, “Air Traffic Services - Annex 11,” 13th edition, Montreal (CA), 2001.

[28] ICAO, “Rules of the Air – Annex 2,” 9th edition, Montreal (CA), 1990.

[29] P.M. Fitts, “The information capacity of the human motor system in controlling the amplitude of movement,” Journal of Experimental Psychology, 47, 1954, pp. 381-3