Experimental Quantification of Times Needed to Comply With Air Traffic Control Advisories (FCU & MCDU Interaction)

Markus Vogel, Christoph Thiel, Hartmut Fricke
Chair of Air Traffic Technology and Logistics
Institute of Logistics and Aviation
Technische Universität Dresden
01062 Dresden, Germany
{vogel,thiel,fricke}@ifl.tu-dresden.de

***Abstract*—This work describes the experimental quantification of pilots' time requirements when making adjustments to the aircrafts navigation systems upon clearances from ATC, which was performed in order to calibrate a simulation-based safety assessment model for ATM procedures. The experiment, conducted on a A320 flight simulator with university students, was laid out as a part-task real-time HITL study following a micro-world approach. A task analysis was performed in preparation, providing an estimate for typical time demands. The 5386 measurements are WEIBULL-distributed both for device acquisition and device interaction time. Device acquisition times follow FITTS law. FCU device interaction times can be described with a parametric model dependent on input magnitude, and can be explained with a microscopic model, which segregates coarse and fine adjustments to the dial knobs with distinctly different rates and precisions. MCDU keypad interaction occurs with a typing rate much lower than computers and mobile phones (80-100 chars min^{-1}). For all devices, a trend towards earlier and more normalized responses is apparent under higher task loads. All results are summarized in one table situated at the end of the text.**

***Index Terms*—Human Machine Interaction, Human Performance Modeling, Air Traffic Management, Simulation-/Agent-Based Safety Assessment**

I. Introduction

Safety assessments of infrastructural and procedural ATM upgrades are challenging and tedious. Current methodology, both the highly formalized safety assessment techniques and complementary experimental approaches, depends strongly on expert knowledge. Consequently, the results are always expert-biased, particularly with respect to the "relevant" hazards. An integrated, simulation-based approach targeting the revelation of unknown and/or unimaginable hazards in an automated and objective fashion is therefore being developed [1]. Air traffic simulation has has a long history for the purpose of performance evaluation and industry solutions exist (e.g. RAMS plus[1]). The effects of procedural and human factors influence are, however hard to grasp and open the field of human performance / cognitive modeling [2, 3]. A classic example for the application of cognitive theory to the aviation domain is AIR-MIDAS [4], which models the operator with a high level of detail in order to optimize the crew station and according procedure design. For ATM simulations, the complexity of multiple stakeholders acting cooperatively in a so-called socio-technical system [5–7] arises, leading to distributed and inherently asynchronous event chains. The assessment of adequate models, which are mostly termed agent-based, is nontrivial due to this complexity [8, 9]. Nevertheless, numerous applications exist, with published results underlining the benefit of emergent behavior which would otherwise be hard to predict, let alone quantify [10–13]. Nevertheless, the level of detail varies greatly, depending on the application and, when risk assessment is targeted, safety case (e.g. runway incursion, midair collision, airspace congestion, scenarios etc.).

The authors have decided to focus the emergence of separation infringements from technical (actual navigation performance, ANP, [14]), procedural [1, 15] and human performance [16, 17] factors without considering the safety nets of STCA and ACAS. The approach comprises an agent-based simulation and a probabilistic safety assessment model, which identifies collision risk probabilities with safety nets disregarded.

The agent-based simulation covers flight mechanics and -management as technical factors, models current and predicted future operations according to ICAO PANS-OPS and -ATM [18, 19] and human factors on varying level of detail (the determination of which is the main objective of current research). Next to functional agent modeling, significant effort was spent on the identification and probabilistic modeling of the agents behavioral parameters (by means of literature study, interviews, task analyses, human factors experiments, sensitivity analyses, etc.). In [17], the effects of human latencies introduced by reaction time and time demand for making inputs to the aircraft's navigation systems was assessed in form of a sensitivity analysis. While technical and procedural parameters remained constant, the pilot agents' time demand for compliance with ATC, representing dead-times in terms of control theory, was varied under the hypothesis that a safety margin accommodating such human performance variations is inherent to certified procedures.

In theory, human-induced latencies form dead-times in the distributed air traffic control loop, implicating a large sensitivity. The effects of late compliance are shown schematically

[1] http://www.isa-software.com/ramsplus/

in fig. 1. Late responses to heading adjustments will lead to human-induced cross-track tolerances (XTT), while late adjustments in altitude increase vertical track tolerance (VTT). These error types are procedurally safeguarded, e.g. by the geometry of approach transitions. Speed and distance errors, however, propagate directly onto along-track tolerance (ATT), which is highly critical to classic separation and also to 4D trajectory based operations (4D-TBO). Therefore, the observed effects are as important to the future ATM generation as to the current one. The along-track error is systematically different between *vector* and *direct* advisories (fig. 1). The simulation results in [17], indicate an almost safety-neutral time buffer before of near-constant safety performance (ca. 5 s for adjustments in speed, heading and altitude, ca. 10 to 10 s for flight plan modifications). After these safety-neutral buffers, collision risk is highly sensitive to further delays in compliance.

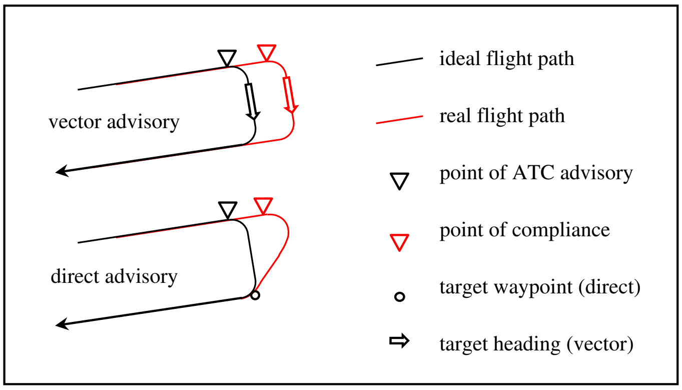

Figure 1. schematic effect of ideal (black) and late (red) compliance with a path-shortening advisory, *vector* (top) and a *direct-to* (bottom)

As the literature does not provide dependable models for time demand, a human-in-the loop (HITL, HIL) experiment was performed at a A320 research simulator. By measuring the time for full compliance, the reaction time plus device acquisition time plus device interaction time was measured. The experimental setup (described in Section II) included time-stamping the first input to the respective device, in order to segregate two figures. Action traces were recorded as well in order to understand input strategies and possible flaws (slips, mistakes, mode confusion).

II. Methodology

A. Introduction

On a formal level, a part-task real-time HITL study [20], itself closely related to the *micro-world* [21] experimental approach, was carried out [22]. Both [20] and [23] offer detailed practical insight that guided the experimental design and eventual evaluation scheme. In the terms of [21], the primary aim of the study was quantifying individual differences in performance, while (adverse) effects of system characteristics and error types and frequencies were of equal interest. Fig. 2 graphically depicts the actions and artifacts leading to the micro-world design, serving as a guideline for the remainder of this chapter.

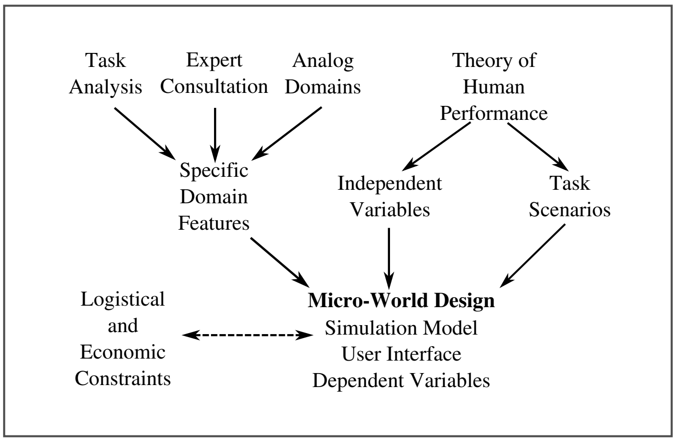

Figure 2. activity flow chart for experimental design [23]

B. Task Analysis and Theory of Human Performance

The task of ATC advisory compliance can be decomposed on an upper level into the part tasks of

1) listening and comprehension (both PF and PNF)
2) planning and implementing the cockpit input (PF)
3) confirmation (read-back) and monitoring (PNF)

For the purpose of the application (fast-time ATM simulation) the tasks of the PNF are of secondary interest and were disregarded for this study [22]. The tasks of the PF were subjected to a structured task analysis (STA, [24, 25]). Along the process, a prediction following the theory of human performance (GOMS, [26]) was investigated.

The listening and comprehension task was analyzed considering closely the regulation of standardized phraseology [27, 28]. Among highly trained aircraft operators, the text comprehension task is of subordinate significance due to the unique context of standardized phrases. Thus, ICAO phraseology resembles closely auditory *cueing* [29]. The cueing phrases are issued directly after the address (call-sign) and thus reveal the nature of the ATC advisory / clearance at an early stage (tab. I). The following changes in aircraft navigation state were part of this analysis

SPD A *speed* change (indicated airspeed) is advised. The pilot flying (PF) “SELECTS” the speed by turning the according FCU (fig. 3) dial knob, then pulling.

HDG ATC issues an advisory for a new *heading* to fly (vectoring). Cockpit input on the FCU as above.

ALT A change in (pressure) *altitude* is required. Cockpit input on the FCU as above.

DIR Alternate to the *vectoring* approach, which has the side-effect of degrading RNAV procedures, path-shortenIn IIB you discuss initial models based on highly-trained professionals, whereas your subjects are somewhat trained students... this deserves clarification.ing can be achieved with *direct*-to-waypoint advisories. The PF then “MANAGE”s the flight path via MCDU (fig. 3) by pressing the “DIR” key and following an on-screen dialog.

Figure 3. Flight Control Unit (FCU, left), Multi-Purpose Control and Display Unit (MCDU, right)

Table I
AUDITORY CUES FOR CLEARANCE TYPE AND ACTION TO BE PERFORMED

type	auditory cue	action to perform
SPD	increase / decrease	FCU dial right / left
HDG	turn right / left	FCU dial right / left
ALT	climb / descend	FCU dial right / left
DIR	(proceed) direct to	MCDU "DIR" key

Upon receiving the auditory cue, the task of moving the hand to the appropriate input device and the task of listening and comprehending the target value (airspeed, compass heading, flight level, or way point name) are to be carried out in parallel. The hand moving part task is highly trained and thus does not demand much mental capacity, but will be guided by eye. The duration of this task is predictable with FITTS law [30], which has received attention in HCI (mouse interaction) but also criticism in other domains e.g. for its one-dimensionality. Further research more directly applicable to the problem has been undertaken, most notably [31] which concludes a reaction of the eye (on a visual stimulus) within 200 ms and the succeeding hand movements 100 ms later and a duration of hand movement well according to FITTS law. The sum of eye and hand latency is commonly called reaction time (stimulus - response). Unfortunately, [31] does not list concrete values for the duration of hand movements except that "a good approximation could be achieved" with eq. 1 and $\varepsilon_H = 15\,mm$ (pointing with index finger tip) and $\theta_T = target\,distance$. It remains unclear, however, if a would represent the reaction time and what the concrete Value of K has to be. Following GOMS [26], the temporal demand should lie in the order of at least 400 ms (*home-to-keyboard*), yielding a total of at least 750 ms from auditory cue to hand on device.

$$T_{HD} = a + K * \log(\theta_T / \varepsilon_H) \tag{1}$$

As for auditory stimuli, [32] indicates that reaction times are smaller than for visual stimuli, with consistent reductions in the order of 50 ms for various *foreperiods* (alerting in advance) and stimulus intensities. With the call sign group acting as foreperiod in a transmission from ATC, the sources indicate that a reaction time in the order of 250 ms (beginning from the auditory cue) has to be expected. The duration of the hand movement has to be expected to follow FITTS law, due to a lack of reliable parameters no prediction can be made.

The task of listening and comprehending the target value to be fed into the aircraft will not be impaired by the parallel hand movement [33]. The phonological loop will hold the audio heard through the so-called rehearsal process [34] which is additionally reinforced when the PNF reads back the advisory. Furthermore, the comprehension sub-task does not lie on the critical path for task completion, as the auditory cue does not only encode the input element but also the first action to perform: grip and turn right / left in the case of the FCU, press the "DIR" key in case of the MCDU. Diligent modeling would need to include the duration of gripping or pressing, but this aspect seems to be negligible. The literature does cover aspects of force and kinesthetics (in the field of laparoscopic and robot-assisted surgery) but not temporal demands of the act itself.

Next, the input to the device needs to be considered. In case of the MCDU this can effectively be realized with the framework of goals, operators, methods and selection rules (GOMS, [26]), which states a key press rate of $\frac{1}{280\,ms}$. In case of a DIR advisory, the sequence for entering the target way point on the *scratch pad* is DIR-(W-P-1-2-3)-LS1L-LS6R. The LS keys are the line selector keys situated left and right of the CRT (compare Figure 3 on page 3). Before pressing the LS1L (uppermost left key), the PF verifies the information on the screen (correctness of way point name), which [3, p. 339] models using CPM-GOMS (cognitive perceptual motor GOMS) lasting 470 ms. LS6R triggers the activation and is always in place at this location (lowermost right key), thus not requiring any visual checking. The sequence, including visual checks denoted with $\mathbb{C}$, reads DIR-(W-P-1-2-3)-$\mathbb{C}$-LS1L-LS6R. In total, this yields $8 * 230\,ms + 1 * 470\,ms = 2310\,ms$ for the scratch pad strategy. Alternatively, the PF could opt for selecting the appropriate way point from the screen directly after pressing the DIR key, which the FMS shows according to the momentary flight plan. The resulting sequence is dependent on the flight plan and the visual scan pattern of the PF. Assuming that the FMS selection is supportive of the procedure and the PF has a good expectancy of the location of the target way point, a visual scan comprising 3 way points leads to the sequence DIR-$\mathbb{C}$-$\mathbb{C}$-$\mathbb{C}$-LSXX-LS6R and a total of $3*280\,ms + 3*470\,ms = 2250\,ms$. In summary, including the reaction time, a total of about 3 s for full compliance with the ATC advisory can be expected. It is, however, predicable that professional airline pilots will not rush through the procedure due to the safety-critical nature of the task and the equivalent training received (e.g. cross-checking the navigation display).

For FCU interaction, all inputs are made with dial knobs that further allow pulling for SELECTED mode and pushing for MANAGED mode (fixed set point vs. autonomous navigation via FMS, in Airbus terminology). For initial compliance, the sequence of pulling and turning the dial knob by a reasonable amount towards the cued direction will be sufficient. For example, after a *turn right heading…* advisory, pulling and increasing a quarter or half rotation will make the aircraft start turning immediately, while the physics of flight will leave sufficient time to enter the desired target value, which for example may be 90° relative to the previous heading. Since the literature does not yield any applicable human performance models for dial knob / jog wheel / turn and press controller / rotary input HMI, the values from GOMS may serve as an estimate: pulling and pushing are equaled with key presses, as are rotations, which are limited to 180° (16 increments) by ergonomic constraints [35, p. 978] and subject to assumed inaccuracies of ca. $\pm 10\,\%$. Initial compliance is thus reached through a sequence P-16R within $2 * 280\,ms = 760 ms$, yielding 1.3 s total, including reaction time. For full compliance, the value needs to be adjusted to the correct target value, which comprises repeated checking of the 3-5 digit 7-segment display. Following [26], the time demand for checking is lower for less complex information; furthermore, the eyes will remain fixated on the display. Hence, adjustments with visual checking are modeled with $280\,ms + 150\,ms = 430\,ms$. For the exemplary heading change, the sequence could read P-25R-27R-32R-4R-1R-1R (coarse adjustments followed by fine end adjustments), yielding a total of $1 * 280\,ms + 6 * 430\,ms = 2860\,ms$ (or 3.6 s including reaction time). It is evident that individual differences will be high and very much dependent on proficiency. Nevertheless, this kind of human performance estimation, including the randomness for inaccuracies, could be implemented in a simulation following a Monte-Carlo approach once calibrated.

C. Analog Domains

Analog domains were researched for applicable or transferable findings, but did not offer transferable human performance models for dial knob HMI. The areas of research are clustered in human computer interaction (visualization and virtual reality [36–39], immersion [40], entertainment [41–44] and automotive [45, 46] applications), and medical engineering (rehabilitation [35], robot-assisted and laparoscopic surgery [47, 48]). The common feature of all cited publications is the presentation of concepts and solutions and thorough usability or human performance evaluation for the concrete field of application. No generalized research from cognitive psychology has been identified as relevant to this publication. As laid out above, the part task of interest also differs distinctly from other rotational inputs by closing the feedback loop by means of seven-segment numeric displays.

D. Task Scenarios

First, a set of general decisions concerning the experiment design had do be made. As laid out above, the part-task HIL experiment was developed following a minimal micro-world approach [22], which led to the decision to explicitly exclude multi-tasking and multi-crew coordination issues. Subjects were assigned to the captain's seat (left), taking the position of PF, while the first officer seat (right) remained unstaffed. The left position was chosen for dexterity, the PF role for the desired duty of implementing ATC advisories, and absence of a second subject performing FO/PNF duties was found well maintainable except for the missing read-back which would reinforce the rehearsal process of the phonological loop; meaning that a raised *mistake* rate due to aural misperceptions had to be taken into account, but not a difference in the time demand for correct actions. For maximal reproducability, ATC advisories were recorded in advance, spoken in English language following ICAO phraseology. As the subjects were of German native speakers, the temporal demand of processing the instruction could be slightly raised. To compensate for these issues, and the fact that aviation diploma and master students were recruited for the experiment, an extensive training / trial stage was scheduled [22].

Second, the actual task scenarios needed to be designed. Scenario duration, task load, variability of tasks, and realism were identified as relevant independent variables [22]. In order to limit and normalize the effects of saturation and monotony, the scenario duration was fixed to 12 to 15 min, and secondary tasks (e.g. flap setting, page changes on the EFIS) were introduced in order to add diversity. Task load and diversity (advisory types and target value changes) formed the independent variables of the experiment. The decision on the realism towards authentic (approach) procedures was ambivalent. A comprehensive dataset was crucial for the analysis; as the subjects were not trained pilots, realism towards authentic procedures was deprioritized [22]. In consequence, the experiment resorted to real-world approach procedures, but added a significant amount of intermediate advisories to increase the number of measurements. This decision demanded for intensive testing on the simulation environment (see sect. II-E) in order to prevent triggering aircraft alarms (master caution / warning, [49]). A summary of the devised scenarios is given in tab. II below. Tab. III shows the experiment's schedule for the main stage, comprising 4 sessions with 6 scenarios each. Consequently, 5h scenario time, flying intermediate approach transitions inbound to EDDM (Munich, Germany) and EDDF (Frankfurt, Germany) had to be performed in total by each subject.

Table II
TASK LOAD VARIATION WITHIN THE SCENARIOS

task load	average interval	secondary tasks
low	15 s	none
medium	10 s	few
high	5 s	numerous

Table III
TASK LOAD SCHEDULE FOR THE MAIN STAGE OF THE EXPERIMENT [49]

1st session	2nd session	3rd session	4th session
warm-up scenario			
low	variable	low	medium
variable	medium	medium	high
low	low	high	medium
break			
low	variable	low	low
variable	medium	medium	medium
low	low	high	high

Third, a software application for scenario play-back and data acquisition was designed and implemented (Java, Swing UI). For scenario play-back and time-stamping of subject responses, the internal system clock of the instructor workstation PC was used. The recorded ATC transmissions were replayed to a headphone, started according to a scenario definition file. The relevant data objects were updated by means of TCP socket communication (push updates from the simulation server, see sect. II-E). The subjects' response was then measured in relation to the occurrence of the auditory cue and the target value in the audio file (see sect. II-B, also part of the scenario definition file). No further corrections were made, as the magnitude of network jitter and computing latencies was deemed negligible. The results were saved in a set of text-based log files (raw event and aircraft state log, condensed experiment result log). A set of state machines (one for each input device) was used to track the subjects' actions in relation to the task, timing task completion and generating a condensed event chain (similar do the sequence notation in sect. II-B). Task completion was timed as soon as the values reached the target and remained unchanged for 3 s (FCU) or the flight plan was updated with the correct target way point (MCDU). Fig. 4 shows the instructor work station in a state indicating that the subject has completed a HDG advisory (green box) after inadvertently having modified and corrected the altitude (red box).

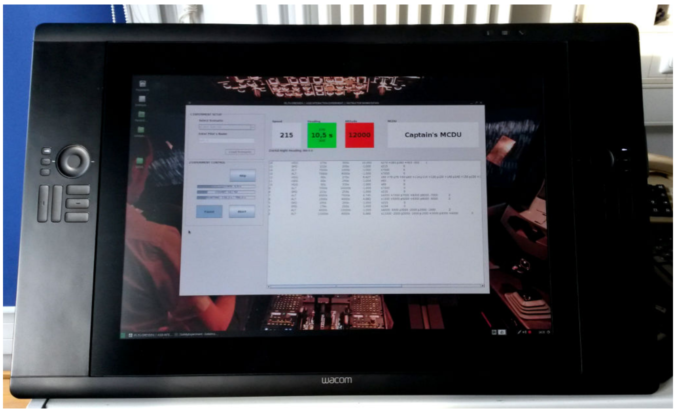

Figure 4. instructor workstation with scenario play-back and data acquisition software

E. Micro-World Design

The micro-world was largely predefined in form of the institute's the flight simulation laboratory consisting of

- FlightDeck Solutions[2] fixed-base A320 cockpit mock-up
- Faros, now ECA Group[3], MCDU and FCU Hardware
- current X-Plane[4] flight simulation server
- Qpac[5] proprietary A320 aircraft model
- ExtPlane[6] network data synchronization plug-in
- 225° outside view projection system made from commercial off-the-shelf products[7]

The simulation was initialized at the beginning (clearance limits) of approach transitions in Frankfurt (KERAX25N, PSA25S, and PSA25N) and Munich (BETOS08, ROKIL08, and BETOS26). Although instrument flight rules applied, the outside view was activated and daytime was set for realism and situation awareness. The cockpit was fully equipped and operational during the experiment, including EFIS.

[2]http://www.flightdecksolutions.com/components/a320/
[3]http://www.ecagroup.com/en/training-simulation/aviation-simulation
[4]www.x-plane.com/pro
[5]http://www.qpac.eu/index.php/research-development/aircraft-simulation
[6]https://github.com/vranki/ExtPlane
[7]curved screen, 3 close-range video projectors, software-based graphics rectification

Figure 5. flight simulator during the experiment

F. Expert Consultation

First, since the simulation setup is not a certified Flight Simulation Training Device, professional airline pilots holding a A320 type rating were consulted concerning realism. The side sticks and foot pedals were were criticized as unrealistic both in input sensitivity and force feedback and are being currently upgraded for this reason. The hardware relevant for the experiment received no critical remarks. Second, another airline pilot was consulted as part of the the experiment design and identified valuable improvements (amount and frequency of advisories, typical secondary tasks during approach procedures, [22]). Third, two airline pilots holding type ratings for B777 and A300 were able to take part in the experiment and provided both feedback and a baseline measurement for comparison.

III. Experiment

The experiment was performed over a period of six weeks with seven students of traffic engineering (6 ♂, 1 ♀, one left-handed) in the 5^{th} to 8^{th} semester (20 to 25 years of age), following the schedule in tab. III and after an initial training phase consisting of 3 sessions [49]. The two pilots (2 ♂) were in their early 30^{ies} and completed one training scenario and one measurement scenario each. Tab. IV shows the number of measurements relevant for analysis. A significant number of further measurements was acquired (non-compliant or incomplete implementations, variable task load, advisories containing combined instructions, e.g. heading and speed change, secondary tasks). Error modeling is currently in progress. Fig. 6 shows an exemplary comparison of the flight trajectories from one scenario, which exhibit divergences due to different compliance times.

Table IV
NUMBER OF COMPLETE MEASUREMENTS FOR COMBINATIONS OF ADVISORY TYPE AND TASK LOAD

type	low	medium	high	Σ
SPD	572	916	433	*1921*
HDG	398	809	432	*1639*
ALT	472	607	359	*1438*
DIR	130	196	62	*388*
Σ	*1572*	*2528*	*1286*	*5386*

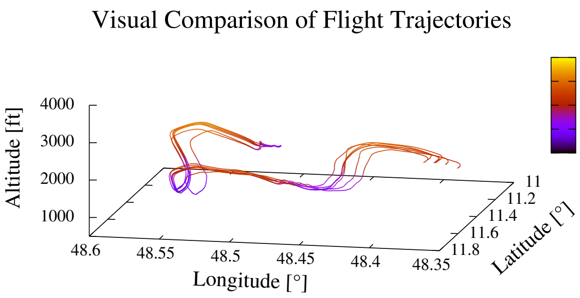

Figure 6. flight trajectories for high task load scenario (EDDM, ROKIL)

IV. Results

A. Overview

In the course of this section, the experiment results are presented and modeled. First, interpersonal differences are analyzed. Second, a basic stochastic model utilizing WEIBULL distribution functions is fitted to the data. For FCU dial knob interactions, a combined parametric model for value changes (click increments of the dial knobs for SPD, HDG, ALT) is presented, leading to a detailed discussion of the turning or twisting task including recurrent changes of grip and magnitudes of adjustment. Finally, a task load dependent model, again utilizing WEIBULL distributions, is presented.

Due to technical limitations (data acquisition by means of push-updates from the simulation server) the reaction time, the duration of hand movement towards the respective input device, and the first input to the device were measured in combination. According to the formulation of FITTS' law, the combination of reacting to a visual or auditory stimulus and pointing at target is best termed *pointing, hand movement,* or, more concrete to the task at hand, *device acquisition* time. Nonetheless, the combined time for device acquisition and the necessary first input will be referred to as *reaction* time for brevity and clarity. The following period of input device manipulation, lasting until the target value is achieved, will be referred to as *(device) interaction* time.

B. Interpersonal Differences

Interpersonal differences in performance were analyzed very early in order to evaluate the effectiveness of the learning phase [22, 49]. The analysis was then repeated for the main stage of the experiment. Fig. 7 shows the results for reaction time, fig. 8 those for device interaction time. The plot indicates usual and thus acceptable interpersonal differences among subjects. The difference between students and professional pilots is negligible for expectancy, minimum, and maximum values. The statistical spread of the inner quartile, however, is considerably higher for the pilots. As the pilots only completed one high task load scenario each, the sample size is lower. On the other hand, student no. 2 performed very similar to the two pilots over the whole course of the main stage. Individual differences for specific advisory types and value changes were also investigated, with similar results.

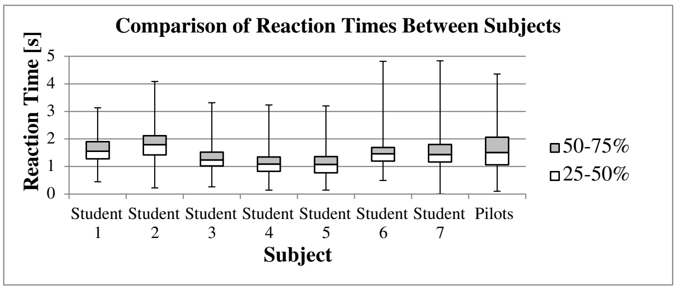

Figure 7. interpersonal differences in reaction time [49]

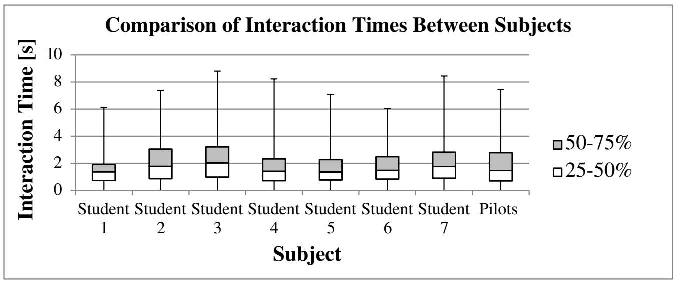

Figure 8. interpersonal differences in interaction time [49]

C. Basic Model

1) WEIBULL *Distribution:* The data for reaction and device interaction is in all cases left-bounded and skewed right. Consequently and in line with theory [50], WEIBULL distributions (denoted here with the shape parameter α, the scale parameter β, and the offset or location parameter μ, CDF in eq. 2) are valid statistical models. Alternate distributions and their specific advantages for more exact tail modeling are discussed in [51].

$$F(x, \alpha, \beta, \mu) = 1 - e^{-(\frac{x+\mu}{\beta})^{\alpha}} \tag{2}$$

The distributions were fitted by

1) subtracting the minimum (μ) from the entire dataset and performing a GUMBEL parameter estimation on α and β as a first approximation
2) using an MS Excel solver based on the Generalized Reduced Gradient (GRG) algorithm to minimize the mean squared error on classified frequencies [49].

The second step makes an implicit trade-off between the quality of fit in the core and the tail region of the distribution. The implications on risk assessment applications are to be assessed separately.

2) Reaction Time: The reaction times were processed over all datasets, including incomplete or non-compliant device interactions, varying task load and combined advisories, but separately for each input device. The results, depicted in fig. 9, graphically show the trade-off between core and tail fidelity of the distributions. The mean values qualitatively correlate with the reach distance (DIR < SPD < HDG < ALT, [49]).

As the next step, an attempt for a parametric model with the device distance as parameter is made, following the definition of FITTS Law. The difficulty index (d) was determined by distance measurements in the cockpit mock-up, using the

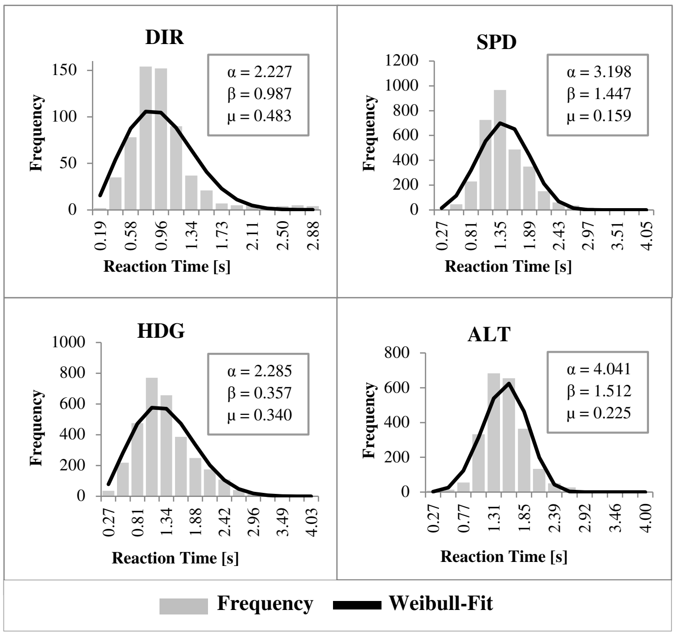

Figure 9. input device specific reaction time [49]

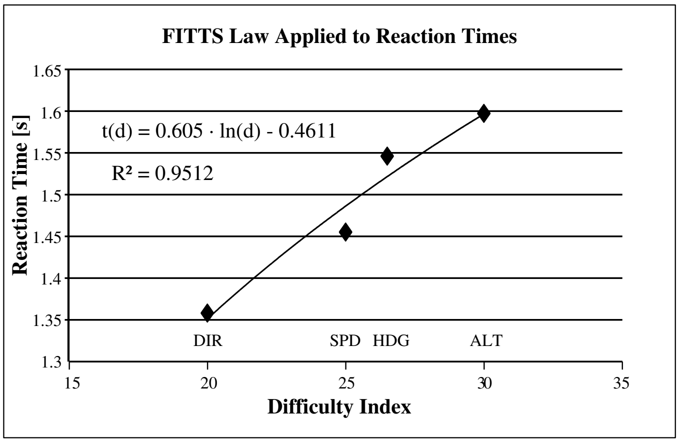

Figure 10. FITTS law fitted to the reaction time measurement

right knee location in a properly adjusted seat as the resting point of the hand and measuring distance and perpendicular width of the input elements (tab. V). Fig. 10 shows that the reaction times are generally distributed according to FITTS law. The MCDU reaction time is over 300 ms longer than expected (1030 ms vs. 1358 ms) which can be explained by a significant majority of FCU interactions in the scenarios that may have led the subjects to keep their right hand closer to the FCU than to the idealized knee resting position. Additionally, the large difference between SPD and HDG is a noticeable, as the distance between the two knobs is only 3 cm. Assuming a hand resting position somewhere between MCDU and FCU this deviation is explainable. As no video recordings were taken during the experiment, no further analysis is possible.

Table V
MEASURED DIFFICULTY INDICES FOR INPUT DEVICES

device	distance	width	width measurement	difficulty index
DIR	0.30 m	0.015 m	DIR key diagonal	20
SPD	0.50 m	0.020 m	knob base diameter	25
HDG	0.53 m	0.020 m	knob base diameter	26.5
ALT	0.60 m	0.020 m	knob base diameter	30

3) FCU Interaction Time: Before modeling FCU dial knob interaction times following the same approach, the influence of left versus right turns of the dial knobs was evaluated and found unremarkable [49]. Mean values differed less than 5 % in varying directions, and standard deviations differed less than 10%, also in various directions. Therefore, the datasets for left and right turning directions were combined. Furthermore, as the haptic feedback and value increment angular resolution was the same for all three dial knobs (32 mechanical clicks per revolution), the dataset as transformed using click increments as sole metric [49]. Some datasets were combined by this, as the value changes were not unique to the input device (SPD: *10*, 20, *30* KIAS; HDG: *30*, 50, 90 °; ALT: 2, 4, 6, *10* · 1000 ft).

Fig. 11 shows exemplary results for the WEIBULL approximation (value changes of 10, 30 and 90 increments). A steady, nonlinear relation between value change magnitude and duration of device interaction is apparent when observing the peaks. Fig. 12 makes this relation more clear with the well-interpretable GAUSSIAN distribution parameters (raw data in gray, 3[rd] order polynomial trend lines in black). Further analysis of the fitted distributions parameters yielded remarkably exact approximations for the WEIBULL parameters, shown in fig. 13.

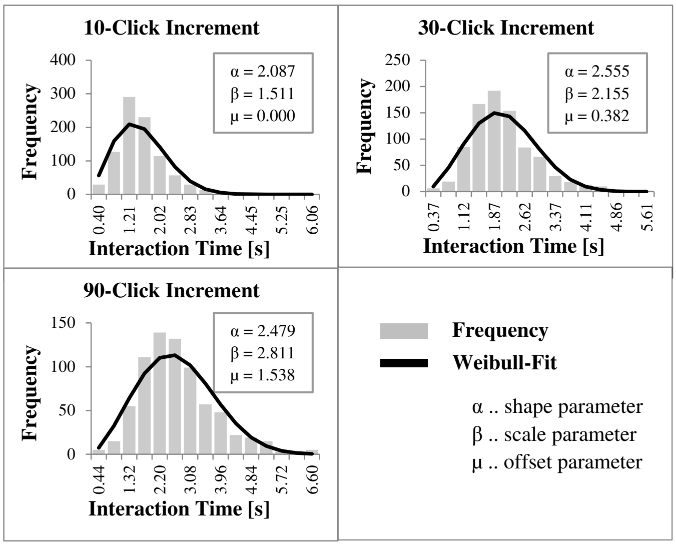

Figure 11. exemplary WEIBULL distributions showing the influence of the value change magnitude on FCU interaction times [49]

4) Dial Knob Manipulation: Next, the manipulation of the FCU dial knobs was analyzed in detail, observing angular velocity, re-adjustments of hand position, and accuracyfrom the logged interaction sequence [49]. The results indicate that coarse adjustments (>10 increments) are performed with a rate of ca. 60 increments per second while fine adjustments have a rate of ca. 10 increments per second. A parametric model could be established, showing a linear relation between value change and coarse adjustments and a logarithmic relation for fine

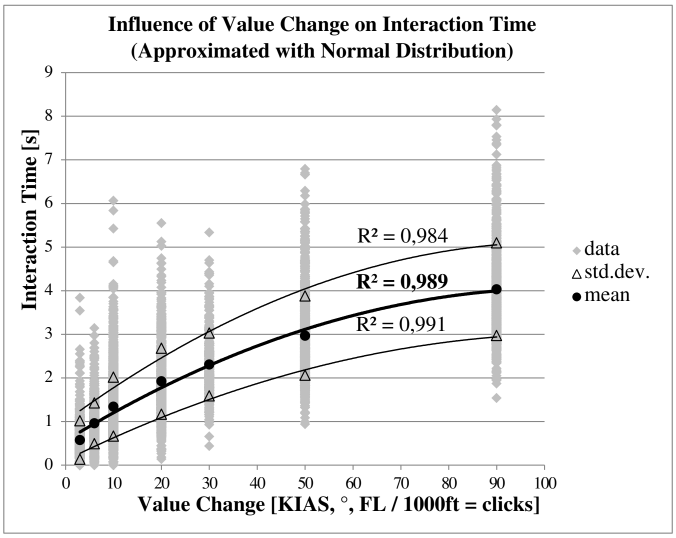

Figure 12. parametric GAUSSian model for the influence of the value change magnitude on FCU interaction times [49]

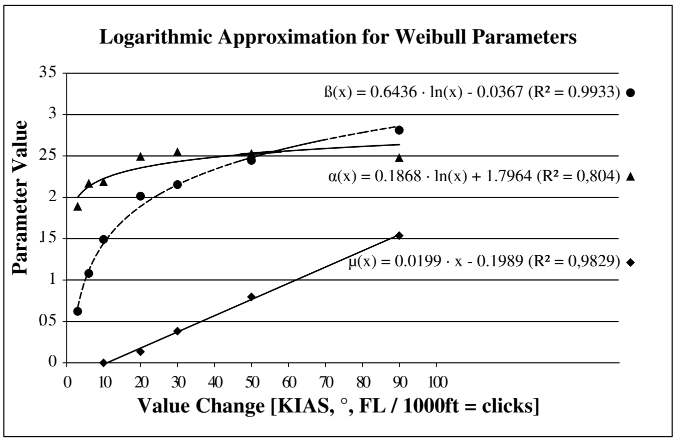

Figure 13. parametric WEIBULL model for the influence of the value change magnitude on FCU interaction time [49]

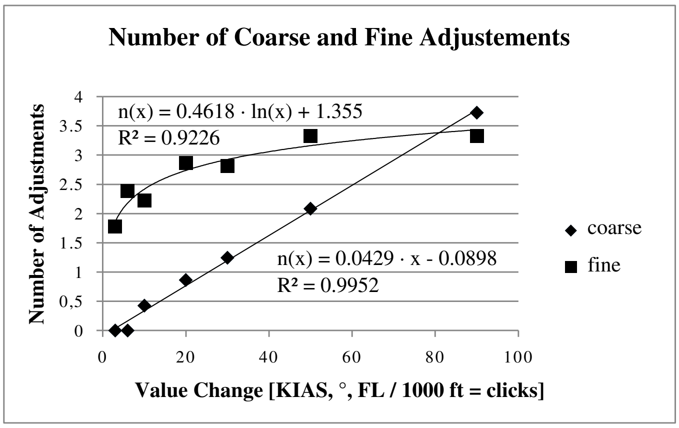

Figure 14. influence of the value change magnitude on number of coarse and fine adjustments [49]

adjustments (fig. 14) [49]. With respect to accuracy, it could be determined that the percentage of actions overshooting the target value is roughly proportional to the magnitude of change, whereas no significant difference in time demand is detectable between subject responses that did and did not overshoot. This leads to the conclusion that the final coarse grained adjustments are over- and undershooting the target value by the similar amounts and probabilities.

5) MCDU Interaction Times: As laid out in sect. II-B, two strategies exist for implementing a direct-to advisory: entering the way point name into the scratch pad (8 key presses) and selecting the way point from the list (3 key presses). The subjects were trained for both strategies and did not receive instructions which strategy to employ. Quite surprisingly [49], only 6 % of the measurements follow the strategy of list selection, which resulted in a device interaction time of 3.691 s (0.81 s s.d.). The predominant strategy of entering the way point name resulted in a considerably longer device interaction time of 5.548 s (1.25 s s.d.). It is assumed that the strategy decision was dominated by the robustness against varying flight plans and according MCDU modes. Consequently, the statistical spread would have to be lower for the latter strategy, which it is not.

With respect to the theory, the interaction times are much longer than predicted (compare sect. II-B). The GOMS key press duration is defined for touch-typing at a computer keyboard at a speed of $\frac{60\,s}{280\,ms} = 214$ characters per minute. The keys on the MCDU keypad, however, are ordered alphabetically, not the familiar *qwerty* order (fig. 3), and are operated with only the right index finger and without a palm rest, requiring visual attention in addition. Thus, the key press duration for mobile phones, which is in the range of 250 ms to 400 ms [52, p. 13], can be assumed to still underestimate the results. In fact, solving the equations of sect. II-B results in a key press duration between 600 ms and 750 ms (80-100 characters per minute). This overhead which is equally attributed to the MCDU keypad design and the cognitive task of cross-checking the display. The data was equally modeled with a WEIBULL distribution, yielding the parameters α=1.820 s β=2.914 s and μ=3.334 s, resulting in a satisfactory approximation.

D. Task Load Dependent Model

Upon scenario completion, subjects were asked to assess *strain* on a 5-point LIKERT scale [22]. The results for task load are 2.03 (s.d. 0.71) for low, 2.76 (s.d. 0.63) for medium, and 3.71 (s.d. 0.56) for high task load scenarios [49]. The analyses described above were repeated with the dataset dissected by scenario task load.

Reaction times were consistently found to be shorter with rising task load and also less diverse [49], indicating higher alertness and professionalism which may also be attributed to the experiment's schedule because high task load scenarios occurred towards the end. The parameters determined by WEIBULL fitting are summarized in tab. VI. A parametric model for the self-assessed task load investigated, but without satisfactory results.

The FCU device interaction time trends towards lower values and lower variance with rising task load. The data was also WEIBULL fitted, and the resulting parameters were approximated with logarithmic functions [49]. The results of this step, shown in fig. 15 to 17, become increasingly less predictable with rising task load, which is explainable by individually varying stress responses between subjects. The shape parameter α is most affected by this. Interpreted in terms of reliability engineering, the shape parameter α's more constant behavior indicates a more predictable (human: normative) behavior under higher loads. The approximations of the scale (β) and location (μ) parameters is qualitatively more acceptable and trends towards reduced statistical spread and a stronger dependence on the magnitude of the value change (slope of linear μ approximation). In summary, higher task loads tend to

- narrow the distribution of compliance times (earlier and less diverse response times),
- making the interrelation between time demand on the work to be performed more clear.

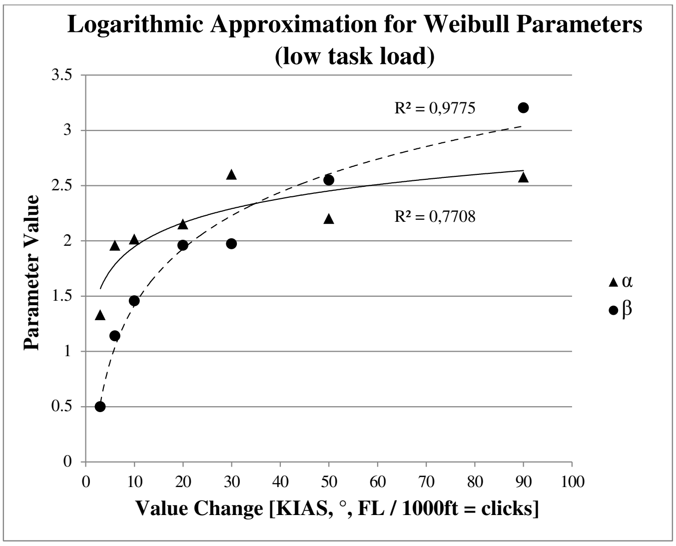

Figure 15. parametric WEIBULL model for low task load FCU interactions [49]

Finally, the time demand for direct-to implementations using the MCDU was analyzed depending on task load [49]. The trends observed above apply to the observed results as well, indicated by the decreasing mean values 9.054 s, 8.198 s, 6.992 s and standard deviations 2.419 s, 1.643 s, 1.514 s for low, medium, and high task loads. WEIBULL fitting was performed, but caution is advised as the number of measurements was comparatively small.

The results of the experiment are presented in summary in tab. VI.

V. Summary and Conclusions

The experimental quantification of pilots' time requirements when making adjustments to the aircrafts navigation systems upon clearances from ATC was conducted on a fixed base A320 flight simulator with students (20-25 years of age, 6 ♂, 1 ♀, one left-handed) and 2 airline pilots (both ♂). In total, 5386 reactions to pre-recorded ATC advisories arranged in low, medium, and high task load scenarios were captured.

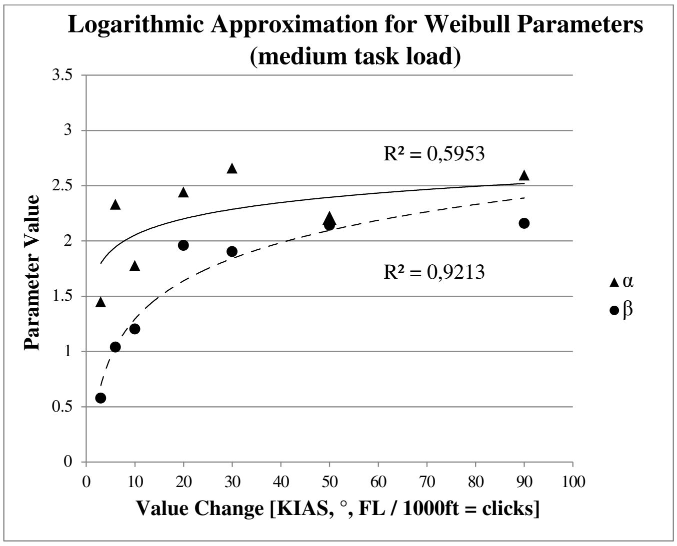

Figure 16. parametric WEIBULL model for medium task load FCU interactions [49]

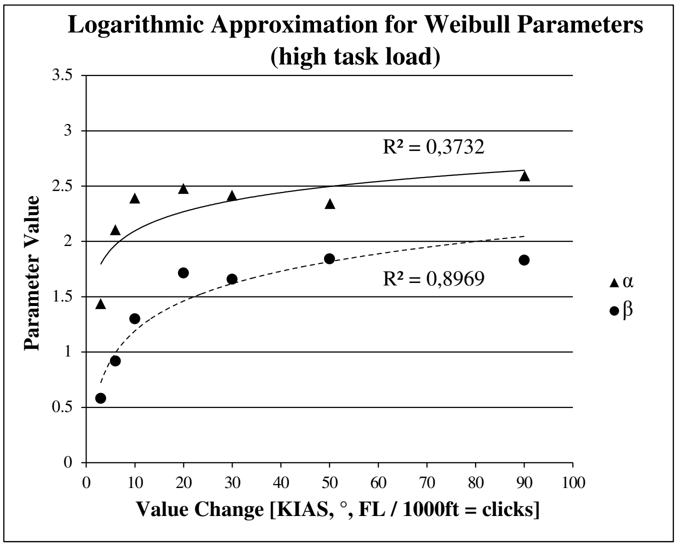

Figure 17. parametric WEIBULL model for high task load FCU interactions [49]

The advisories demanded speed (1921), heading (1639), and altitude (1438) adjustments.

The device acquisition times were found well in line with FITTS law. The device interaction time for FCU can be modeled for all dial knobs and turning directions combined, resulting in a parametric model for the WEIBULL shape, scale and location parameters. The process of manual adjustments to the dial knobs was analyzed in detail, resulting in a valid segregation of coarse and fine adjustments at distinctly differnet rates and precisions, allowing for microscopic modelling of human performance. The interaction time for the MCDU is largely dependent on the chosen input strategy, resulting in different numbers of key presses (3 vs. 8). The key typing rate was found to be much lower than for computers and mobile phones, at approximately 80-100 characters per minute. Reasonable explanations for this difference could be

Table VI
SUMMARY OF RESULTS: WEIBULL PARAMETERS FOR REACTION AND INTERACT. TIMES. INPUT DEVICES. AND TASK LOADS (TL)

	TL	SPD	HDG	ALT	DIR
reaction time	all	α = 3.198 β = 1.447 μ = 0.159	α = 2.285 β = 1.358 μ = 0.340	α = 4.041 β = 1.512 μ = 0.225	α = 2.227 β = 0.987 μ = 0.483
	low	α = 2.278 β = 1.064 μ = 0.558	α = 2.045 β = 1.326 μ = 0.428	α = 3.747 β = 1.349 μ = 0.439	α = 1.783 β = 1.119 μ = 0.483
	med.	α = 2.811 β = 1.256 μ = 0.300	α = 2.583 β = 1.453 μ = 0.200	α = 3.604 β = 1.442 μ = 0.225	α = 1.563 β = 0.576 μ = 0.776
	high	α = 3.349 β = 1.303 μ = 0.159	α = 2.183 β = 1.112 μ = 0.462	α = 3.13 β = 1.169 μ = 0.425	dataset insufficient
interaction time	all	$\alpha(x) = 0.1868 \cdot ln(x) + 1.7964$ $\beta(x) = 0.6436 \cdot ln(x) + 0.0367$ $\mu(x) = 0.0199 \cdot x - 0.1989$			α = 1.820 β = 2.491 μ = 3.334
	low	$\alpha(x) = 0.3141 \cdot ln(x) + 1.2234$ $\beta(x) = 0.7384 \cdot ln(x) - 0.2831$ $\mu(x) = 0.0178 \cdot x + 0.0128$			α = 2.671 β = 6.792 μ = 3.016
	med.	$\alpha(x) = 0.2122 \cdot ln(x) + 1.5648$ $\beta(x) = 0.4889 \cdot ln(x) + 0.1949$ $\mu(x) = 0.0229 \cdot x - 0.2166$			α = 2.840 β = 4.855 μ = 4.073
	high	$\alpha(x) = 0.2521 \cdot ln(x) + 1.507$ $\beta(x) = 0.3828 \cdot ln(x) + 0.3264$ $\mu(x) = 0.0257 \cdot x + 0.2676$			dataset insufficient

established. The quality of WEIBULL fits and parametric metamodels is lower for the task-load dependent model due to the reduced sample size and individual stress responses. Nevertheless, an understandable trend towards more early and less variant responses is observable, and also a strengthening of the relation between work to be performed and time demand.

In conclusion, the analysis delivered valid and reliable results for the question of time demands for compliance with the ATC advisories and quantitative evidence for the influence of task load upon human performance. The models will be incorporated into the agent-based simulation and evaluated following a Monte-Carlo approach using high performance computing hardware. With respect to crew station design, two simple shortcomings stand out. First, the dial knob for speed should be switchable between 1 and 10 KIAS increments since the latter is what ATC usually advises. Second, the keypad layout of the A320 MCDU is highly unergonomic (no palm rest, alphabetic key order, both corrected at recent Airbus models). With respect to procedural design, it is understandable that currently, *vectors* are preferred over *directs*, as compliance times are inherently different. While follwing a *vector*, however, the airraft's FMS is "out of the loop". Future 4DT procedures will have to resort solely to *directs* for path shortening, which results in a strong demand for more efficient HMI, or preferably datalink.

ACKNOWLEDGMENT

The presented research is part of the project *Combined Micro- and Macroscopic Simulative Aviation Risk Assessment Based on Technical, Procedural, and Human Performance Parameters with Minimized Parameter Set*, which is funded by Deutsche Forschungsgemeinschaft (DFG). The authors wish to thank Nicolette Györy who contributed greatly to the experiment design with her final seminar paper [22] and Michael Werner, who supervised the experiments on contract and analyzed the results as part of his student research project [49].

REFERENCES

[1] M. Vogel, C. Thiel, and H. Fricke. "A quantitative Safety Assessment Tool based on Aircraft Actual Navigation Performance". In: *Proceedings of the International Conference on Research in Airport Transportation (ICRAT)*. Budapest, Hungary, 2010.

[2] John R Anderson. "ACT: A simple theory of complex cognition." In: *American Psychologist* 51.4 (1996), p. 355.

[3] Bonnie E John and David E Kieras. "The GOMS family of user interface analysis techniques: Comparison and contrast". In: *ACM Transactions on Computer-Human Interaction (TOCHI)* 3.4 (1996), pp. 320–351.

[4] K.M. Corker and B.R. Smith. "An architecture and model for cognitive engineering simulation analysis: Application to advanced aviation automation". In: *Proceedings of the AIAA Computing in Aerospace 9 Conference*. San Diego, California, USA, Oct. 1993.

[5] Robert P. Bostrom and J. Stephen Heinen. "MIS Problems and Failures: A Sociotechnical Perspective Part I: the Cause". In: *MIS Quarterly* 1.3 (Sept. 1977), pp. 17–32. ISSN: 0276-7783. DOI: 10.2307/248710. URL: http://dx.doi.org/10.2307/248710.

[6] Brian Whitworth. "The Social Requirements Of Technical Systems". In: *Handbook Of Research On Sociotechnical Design And Social Networking Systems* 3 (2009). URL: http://moderntimesworkplace.com/good_reading/GRNewTech/Social.Requirements.of.TechSystems.Whitworth.pdf.

[7] Jean-Marie Flaus. *Risk Analysis: Socio-technical and Industrial Systems*. 1st ed. Wiley-ISTE, July 2013. ISBN: 9781848214927. URL: http://amazon.com/o/ASIN/1848214928/.

[8] R.H. Bordini et al. "Automated Verification of Multi-Agent Programs". In: *Proceedings of the 23rd IEEE/ACM International Conference on Automated Software Engineering*. L'Aquila, Italy, Sept. 2010, pp. 69 –78.

[9] T.H. Lacey and S.A. DeLoach. "Verification of Agent Behavioral Models". In: *Proceedings of the International Conference on Artificial Intelligence*. Las Vegas, Nevada, USA, June 2000.

[10] D. Sislak, P. Volf, and M. Pechoucek. "Large-scale Agent-based Simulation of Air-traffic". In: *In Proceedings of the Twentieth European Meeting on Cybernetics and Systems Research (EMCSR)*. Vienna, Austria, 2010.

[11] Henk AP Blom and GJ Bakker. "Safety of advanced airborne self separation under very high en-route traffic demand". In: *Proc. SESAR Innovation Days, ENAC, Toulouse* 29 (2011).

[12] M.H.C. Everdij et al. "Agent-Based Safety Risk Analysis of Time-Based Operation in Future TMA". In: *Proceedings of the Air Transport and Operations Symposium*. Delft, The Netherlands, 2012.

[13] Soufiane Bouarfa et al. “Agent-based modeling and simulation of emergent behavior in air transportation”. In: *Complex Adaptive Systems Modeling* 1.1 (2013), pp. 1–26.

[14] C. Thiel and H. Fricke. “Collision Risk on final Approach - a radar-data based Evaluation Method to assess Safety”. In: *Proceedings of the International Conference on Research in Airport Transportation (ICRAT)*. Budapest, Hungary, 2010.

[15] M Vogel, C Thiel, and H Fricke. “Simulative Assessment of Reduced Lateral Separation Minima for the Terminal Airspace”. In: *Proceedings of the International Conference on Research in Air Transportation (ICRAT). San Francisco, USA*. 2012.

[16] Markus Vogel and Hartmut Fricke. “Investigating the Safety Impact of Time Buffers in Current TMA Arrival Operations”. In: *Aeronautical Sciences (ICAS 2012), Brisbane* (2012).

[17] Markus Vogel, Christoph Thiel, and Hartmut Fricke. “Assessing the air traffic control safety impact of airline pilot induced latencies”. In: *Proceedings of the 2nd International Conference on Application and Theory of Automation in Command and Control Systems*. IRIT Press. 2012, pp. 38–45.

[18] ICAO. *Procedures for Air Navigation Services - Air Traffic Management (PANS-ATM, Doc 4444)*. 14th ed. International Civil Aviation Organization (ICAO). Montreal, Canada, June 2001.

[19] ICAO. *Procedures for Air Navigation Services - Aircraft Operations (PANS-OPS, Doc 8168)*. 5th ed. International Civil Aviation Organization (ICAO). Montreal, Canada, Oct. 2006.

[20] Andrew Harvey et al. *Best Practices for Human-in-the-Loop Validation Exercises*. Tech. rep. EUROCONTROL Experimental Centre, Federal Aviation Administration William J. Hughes Technical Center, Titan Systems, Feb. 2003. URL: dia.

[21] Berndt Brehmer and Dietrich Dörner. “Experiments with computer-simulated microworlds: Escaping both the narrow straits of the laboratory and the deep blue sea of the field study”. In: *Computers in Human Behavior* 9.2 (1993), pp. 171–184.

[22] Nicolette Győry. *Entwicklung des Experimentaldesigns für HiL-Experimente am A-320 Simulator hinsichtlich Bedienreaktionszeiten der FCU und MCDU*. Tech. rep. TU Dresden, 2014.

[23] J Sauer, DG Wastell, and GRJ Hockey. “A conceptual framework for designing micro-worlds for complex work domains: a case study of the Cabin Air Management System”. In: *Computers in Human Behavior* 16.1 (2000), pp. 45–58.

[24] Barry Kirwan and Les K Ainsworth. *A Guide To Task Analysis: The Task Analysis Working Group*. CRC press, 1992.

[25] Dan Diaper and Neville Stanton. *The Handbook Of Task Analysis For Human-computer Interaction*. CRC Press, 2003.

[26] Stuart K Card, Allen Newell, and Thomas P Moran. *The Psychology Of Human-computer Interaction*. L. Erlbaum Associates Inc., 1983.

[27] ICAO. *Aeronautical Telecommunications (Annex10), Volume II Communication Procedures including those with PANS*. 6th ed. International Civil Aviation Organization (ICAO). Montreal, Canada, Oct. 2001.

[28] ICAO. *Manual of Radiotelephony (Doc. 6432)*. 4th ed. International Civil Aviation Organization (ICAO). Montreal, Canada, 2007.

[29] SA Mudd and EJ McCormick. “The use of auditory cues in a visual search task.” In: *Journal of Applied Psychology* 44.3 (1960), p. 184.

[30] Paul M Fitts. “The information capacity of the human motor system in controlling the amplitude of movement.” In: *Journal of experimental psychology* 47.6 (1954), p. 381.

[31] C Prablanc et al. “Optimal response of eye and hand motor systems in pointing at a visual target”. In: *Biological cybernetics* 35.2 (1979), pp. 113–124.

[32] Pekka Niemi. “Stimulus intensity effects on auditory and visual reaction processes”. In: *Acta Psychologica* 43.4 (1979), pp. 299–312.

[33] Christopher D Wickens. “Multiple resources and mental workload”. In: *Human Factors: The Journal of the Human Factors and Ergonomics Society* 50.3 (2008), pp. 449–455.

[34] Alan Baddeley. “The episodic buffer: a new component of working memory?” In: *Trends in cognitive sciences* 4.11 (2000), pp. 417–423.

[35] Ludovic Dovat et al. “A Haptic Knob For Rehabilitation Of Stroke Patients”. In: *Intelligent Robots and Systems, 2006 IEEE/RSJ International Conference on*. IEEE. 2006, pp. 977–982.

[36] Michael Chen, S Joy Mountford, and Abigail Sellen. “A Study In Interactive 3-d Rotation Using 2-d Control Devices”. In: *ACM SIGGRAPH Computer Graphics*. Vol. 22. 4. ACM. 1988, pp. 121–129.

[37] Shumin Zhai. “Human performance in six degree of freedom input control”. PhD thesis. University of Toronto, 1995.

[38] Shumin Zhai. “User performance in relation to 3D input device design”. In: *ACM Siggraph Computer Graphics* 32.4 (1998), pp. 50–54.

[39] Beverly L Harrison et al. “Squeeze me, hold me, tilt me! An exploration of manipulative user interfaces”. In: *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM Press/Addison-Wesley Publishing Co. 1998, pp. 17–24.

[40] Jens Ziegler and Leon Urbas. “Begreifbare Interaktion mit Distributed Wearable User Interfaces.” In: *Mensch & Computer Workshopband*. 2013, pp. 207–212.

[41] Hugh McLoone, Ken Hinckley, and Edward Cutrell. “Bimanual interaction on the microsoft office keyboard”. In: *INTERACT’03*. 2003, pp. 49–56.

[42] Wanjoo Park et al. “Design of haptic interface for brickout game”. In: *Haptic Audio visual Environments*

and Games, 2009. HAVE 2009. IEEE International Workshop on. IEEE. 2009, pp. 64–68.

[43] Laehyun Kim et al. "A Haptic Dial System For Multimodal Prototyping". In: *18th international conference on artificial reality and telexistence (ICAT 2008)*. 2008.

[44] Laehyun Kim et al. "A Universal Remote Control With Haptic Interface For Customer Electronic Devices". In: *Consumer Electronics, IEEE Transactions on* 56.2 (2010), pp. 913–918.

[45] John H Day. "SUCCESS STORY-Can BMW's iDrive Pass Its Road Test Now? After a catastrophic debut and a major makeover, BMW's iDrive infotainment system is gaining acceptance with critics, competitors, and". In: *Electronic Design* 52.14 (2004), pp. 48–55.

[46] Bernhard Niedermaier et al. "The new BMW iDrive–applied processes and methods to assure high usability". In: *Digital Human Modeling*. Springer, 2009, pp. 443–452.

[47] Russell Taylor et al. "A steady-hand robotic system for microsurgical augmentation". In: *The International Journal of Robotics Research* 18.12 (1999), pp. 1201–1210.

[48] H Xin, JS Zelek, and H Carnahan. "Laparoscopic surgery, perceptual limitations and force: A review". In: *First Canadian student conference on biomedical computing*. Vol. 144. Citeseer. 2006.

[49] Michael Werner. *Bestimmung von cockpitseitigen MCDU/FCU Bedienreaktionszeiten für Agentensimulationen mittels Echtzeitexperimenten*. Tech. rep. TU Dresden, 2014.

[50] Gordon D Logan. "Shapes of reaction-time distributions and shapes of learning curves: a test of the instance theory of automaticity." In: *Journal of Experimental Psychology: Learning, Memory, and Cognition* 18.5 (1992), p. 883.

[51] Fermin Moscoso del Prado Martin. "A theory of reaction time distributions". In: (2008).

[52] Miika Silfverberg, I. Scott MacKenzie, and Panu Korhonen. "Predicting Text Entry Speed on Mobile Phones". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '00. New York, NY, USA: ACM, 2000, pp. 9–16. ISBN: 1-58113-216-6. DOI: 10.1145/332040.332044. URL: http://doi.acm.org/10.1145/332040.332044.