Understanding the safety-relevance of visual cue perception at a Surface Manager HMI

Lothar Meyer¹, Michael Schultz¹, Simon Schmidt-Roßleben² and Hartmut Fricke¹
Technische Universität Dresden
Chair of Air Transport Technology and Logistics
Dresden, Germany
¹{meyer, schultz, fricke}@ifl.tu-dresden.de
²simon.schmidt-rossleben@mailbox.tu-dresden.de

Abstract— Current procedures in Air Traffic Management are characterized by activities of all involved operators to obtain information required for the decision making. In particular, pilots and ATCOs decisions depend on reliable information, which, in case of failed information perception, have the highest degrees of severity. Besides voice communication, activities to perceive visual information are regarded as most critical to the provision of a reliable and authentic picture of the air traffic. Focusing on the migration from today’s concept of operations to higher levels of automation e.g. through the use of virtual control tower and A-SMGCS applications, the activities of operators visual information perception differ from proven patterns. Accordingly, the systematic identification of visual cues for safe decisions is needed for evaluation purposes of safety critical Air Navigation Systems. The availability of knowledge on required visual information is considered as pivotal to assess novel strategies and philosophies of HMI design as well as to conclude on the compliance on the Accepted Level of Safety. This paper follows a novel approach to analyze risks of design-related visual cue perception by the use of Human-In-The-Loop Simulations. It hence offers a predictive empirical approach to risk analysis for system evaluation in the system design phase. This proof-of-concept study evaluates the proposed approach at the example of a novel A-SMGCS Surface Movement Manager HMI and the probability of runway incursion. A conclusion on the most risk-inducing visual cues could be exemplarily obtained by the use of open and closed-end questionnaires and novices acting as test persons. The contribution of eye tracking data analysis is also tackled.

Keywords—safety assessment; risk assessment; hazard identification; visual cue perception; air traffic control; aerodrome control services.

I. INTRODUCTION

Every decision that needs to be taken in ATM related domains relies on the appropriate provision of information to the decision maker. Air Traffic Controller (ATCO) and Pilots represent best these operators of the ATM system that are continuously involved in monitoring and decision-making processes. Both ATCOs and Pilots historically make decisions based on the use of the direct visual view, which is nowadays complemented by voice communication and assistance systems. Surveillance instruments shall complement the adequate provision of information to support the operator in fulfilling all the relevant tasks [1]. Nevertheless, 22% of occurred runway incursions were attributable to the “failure to detect visual information” [2], which underlines the safety relevance of visual information perception. Approvingly, today’s flight deck and tower working environment is regarded as highly sophisticated. It is commonly agreed to be acceptable safe, expressed by an agreed Acceptable Level of Safety (ALoS) [3] resp. the Target Level of Safety (TLS) by ESARR 4 [4], of 1.55E-8 fatal accidents per flight hour in ATM. This TLS is the outcome of decades of evaluating the training, procedures as well as technical instruments. The evaluation contributes significantly to the fact that ATM features a level of safety that is regarded as exemplarily by safety-critical industries.

Since many concepts have been worked out in the scope of NextGen in U.S. and SESAR in the EU [5], numerous novel systems will become operable in ATM in the next decade. Consequently, a large number of startups of novel systems, e.g. in the scope of A-SMGCS, can be expected in rather short time, forming a historically challenging situation by paradigm changes both on pilots and ATCOs side. Breaking away from conventional operations might potentially impair the proven patterns of operators’ activities. Approved methods of sociotechnical design and risk evaluation that base on empirical post-startup operational data (as e.g. risk analysis) could become harder to apply due to the overlying effects of multiple parallel system and procedure changes. The retrospective identification of causes in the design of systems might then be increasingly affected by a not yet known complex interaction between multiple novel systems. Such effects might decrease the internal validity of known retrospective analysis methods when Ceteris Paribus conditions cannot be met. A method that predictively identifies safety impacts by enabling the system designer to evaluate the safety achieved under controllable and reproducible operational conditions before startup will consequently become valuable.

With this motivation, we consider an approach basing on Human-In-The-Loop Simulations (HITLS) as most promising to analyze the risk of sociotechnical systems predictively. Its principle advantages are the ability to provide an environment that controls initial and boundary conditions during the simulation and the ability to conclude from the observable. In this context, the A-SMGCS Surface Movement Manager HMI [6] has been chosen as a representative novel sociotechnical system for a proof-of-concept study in the scope of our current

The authors would like to thank Tobii for supporting us with the Mobile eye tracking – Tobii Glasses hardware during the experimental conduction.
methodological research. The HMI is an assistance tool that shall support the ATCO during all tasks related to the aerodrome control by means of semi and full automation support. Its ground surveillance display visually provides traffic information, enabling the operator to act in accordance to his procedures. These procedures include safety critical clearances e.g. take-off, lineup and landing, clearing traffic via data-link in accordance to the definitions of the implementation levels by Eurocontrol [7].

While the role of voice communication diminishes in this computerized environment, the operator’s visual sensory input, in contrast, will be increasingly stressed and is assumed to become more safety relevant. Spanning the bridge to the safety context, the case of failed visual perception might cause flawed conclusions on the expected development of the traffic situation. The resulting consequences offer the highest degrees of severity [8] [9]. Concluding, when proposing a novel sociotechnical system for startup, we assume the predictive identification of safety-relevant visual cues and the related quantification of a design dependent contribution to risk by perceptual error occurrences as most important for the analysis of risk.

This paper bases on an approach in risk analysis that considers failed activities with regards to the perception of visual information at the example of an experimental Surface Manager HMI application. Visual cues already have been subject for research in ATC and flight crew operations [10] [11]. With a view on the adaption of proven research methods with the objective to quantify the risk of failed perception, a method is proposed that shall identify visual cues and quantify its safety impact by means of a Human-In-The-Loop Simulation (HITLS). The chosen method consists of a questionnaire on the importance of visual cues combined with empiric data on runway incursion events provided by a HITLS-based study with novices as test persons. The conducted online questionnaires and the causal identification will deliver findings about the system designs contribution to the probability to cause runway incursion when the visual cue perception is compromised. Finally the most important findings are presented and discussed with regards to the achieved validity of the results.

II. METHODOLOGY

Our methodology was developed at the example of an ASMGCS Surface Manager HMI in the context of tower ATCOs tasks to control the surface movements of an airport. The methodology follows an empiric approach, concluding from the observable events on risk by conducting HITLS with systems that are not in operation yet. The term risk is used in accordance to the following definition:

“Risk is defined as the probability that an accident occurs during a stated period of time” [12].

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“ [13].

The relation between the visual perception event and the RI is modeled in figure 1, showing an event tree that divides the causal path from the initial decision situation to three possibilities cases, with the RI among other cases.

Assuming n possible causative visual cues, a fault tree, associated with statistic probabilities, models the visual perception events of these visual cues as initial events and the RI as the end event (figure 2). The probability of a perception event of a specific visual cue \( k \in n \) is defined as \( p_{VFK} = P(VFK) \) and the related PF, quantified by an estimated probability of occurrence \( p_{PFk} = P(PF_k) \).

The principle objectives that are investigated within this paper are addressed as follows:

1. How can safety critical visual information that might induce risk in the case of failed perception be identified (Hazard identification)? Our identification approach bases on a closed-end questionnaire that identifies the contributing cues to RI that has been caused by granting conflicted clearances. This is a questionnaire-based approach, tracing backwards from the detected RI event within the HITLS. It relies on the subjective estimation of the test person on the possible causative Perception Failure PF and the related visual cue (failure questionnaire). This shall provide a relative frequency of PF, defined as \( h_{PFk|RI} = H(PF_k|RI) \) that indicates a safety-relevance of the visual cue \( k \) when satisfying \( h_{PFk} > 0 \). The proposed visual cues for closed-end questionnaire are derived from a ‘Visual Cue Model’ that will be introduced in the following modeling section.

![Figure 1. The event tree model “visual perception” and the related cases 1 to 3.](image)

![Figure 2.](image)
(2) How can the probability of the RI be estimated, when a specific visual cue is used for decision making, expressed as \( P(\text{RI}|VP_k) \)?

Assuming a visual cue \( k \) as being safety relevant, its contribution to \( p_{\text{RI}} \) can be quantified by the use of \( p_{PF_k|RI} \), \( p_{RI} \) and \( p_{VP_k} \) as follows:

\[
P(\text{RI}|VP_k) = \frac{p_{PF_k|RI} \cdot p_{RI}}{p_{VP_k}},
\]

with an estimation of the probability using the failure questionnaires results \( h_{PF_k|RI} \approx p_{PF_k|RI} \) and assuming \( P(PF_k|VP_k) = 1 \).

Following this analytic approach, the failure questionnaire \( p_{PF_k|RI} \) and the regular questionnaire on ‘importance’ \( p_{VP_k} \) is assumed to provide sufficient data for the identification of safety relevance and the quantification of visual-cue related risk.

A. Safety Assessment Methodology

Our methodological approach is based on the Safety Assessment Methodology (SAM) provided by Eurocontrol [14] to reflect the best practices for safety assessment of ANS. The SAM development in the context of the EATMP (European Air Traffic Management Programme) provides an aviation standard procedure to reflect the high potential of damage for both the aircraft and involved third parties. An obligation on the providers of Air Traffic Management services to ensure the safety of air traffic is demanded in ICAO Annex 11 [3]. The “burden of proof” for ATC is to satisfactorily demonstrate safe procedures and systems by obtaining an ALoS [15]. In prior research projects we gained sustainable experiences on safety assessment and we further proposed a model-based approach to improve in standard assessment methods to ensure, that safety relevant results will significantly enhance the system development process [8]. As the first step of the SAM, the Functional Hazard Assessment (FHA) outputs safety objectives for identified hazards respecting system definitions, functional requirements and the TLS during system definition phase.

B. Visual Cue Model

The initial proposed visual model of tower ATCOs tasks includes demanded visual cues captured of the task analysis [16, 9] and a study of the manual of operations of the German ANSP DFS [17]. Taking into account the cognitive information processing of human, visual cues can be composed and derived of lower level visual cues, e.g. the distance between two aircrafts on a radar screen is a conclusion from two single position cues. Visual cues can also be the result of an interpretation by operator's experiences e.g. optical indicators for wind shearing in the final approach airspace.

Summarizing, a visual cue tree can be found representing a cue hierarchy that fundamentally bases on “elementary cues” in the lowest level of each branch. The proposed demand model is illustrated in figure 3 at the example of issuing a take-off clearance.

A. Experimental Environment

The primary objective of the experimental design is to assess the proposed model, to identify cues that need to be considered and to exclude cues that are not demanded. Secondly, cues that were contributory to the occurrence of RI are to be identified. The chosen HITLS consists of test persons that operate a Surface Manager HMI as the primary working device. The device complies with the Eurocontrol ASMGCS Implementation level 3 [7], with the functional exception of a missing automat that prevents RI (Runway Incursion Prevention and Alerting Systems, RIPAS). Tasks to be performed by the test persons are defined by ICAO Annex 11 [3, pp. 3-1] and ICAO PANS-ATM doc. 4444 [13] for tower and ground control services. The Surface Manager HMI allows for the selection of target a/c 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 (figure 4).
As one part of the experimental environment the test persons wear a mobile eye tracking device. These glasses allow for recording the individual eye movement during the experiment. The mapping of the data from the included scene camera with the display information results in reliable statements about visual areas of interest.

**B. Test scenarios**

The generated traffic consists of inbound and outbound a/c traffic movements at Frankfurt airport (ICAO code: EDDF) on the four active runways (RWY) in direction 25, operating 25L and 25R as landing only runways, 25C and 18 as take-off only runways. Runway dependencies can be found between RWY 18 and RWY 25C, as well as between RWY 18 and RWY 25L. The dependency between 25C and 25L was not considered due to simulation function limitations. The random traffic generator distributes 42 movements over 45 minutes per execution run with uniformly distributed destination or departure gates (including north and south area stands) and runways. The 42 routes of the movements are initialized by a database providing 42 predefined route proposals that ensure similar task load for all experimental executions. The simulated a/c agents are capable to separate from each other on taxiways and to solve taxi obstruction conflicts as well as taxi crossing situations autonomously.

The execution scenario assumes one controller for both ground and tower controller tasks controlling the whole airport and inducing higher taskload than realistic scenarios would do. The scenarios are consciously adjusted to induce a task overload during a peak time in the scenario provoking sufficient samples of safety critical events to record and to conclude from; this is called the overload scenario.

**C. Test persons and tasks**

For experiment execution, 6 student novices took the role of test persons with each of them conducting the experiment within 10 hours divided into 4 days. The non-expert experience was compensated by a training course and a benchmark of RI-frequencies and reaction time in accordance to the known performance of novices [18]. With increasing hours of training, the behavior during decision situations is assumed to become more constant, which will contribute to the quality und reproducibility of observed visual activities and questionnaire results. The test persons are students in their 4th year of the diploma study programme of Transport Engineering at the Technische Universität Dresden. The students had been instructed in accordance to the tasks of tower and ground control services, as well on the provided A-SMGCS functions of the HMI. Head on conflicts were to be prevented by the test person through routing and guiding anticipatory. The most safety critical task was assumedly clearing movements for take-off on RWY 25C and RWY 18 due to the runway dependencies that involves movements on RWYs 25C, 25L and 18 at the same time.

**D. Eye tracking**

The data gathered from the Tobii Glasses mobile eye tracking system consists of the scene video data, eye fixation point data and microphone data. A static cue code map of the surrounding working environment assisted to map fixation coordinates to instrument information that were probably percepted (e.g. [19, 11]). Following this strategy, the coordinate transformation from fixation coordinates to screen coordinates promises to deliver “elementary cues” of the Surface Manager HMI that were probably detected (figure 3).

To map the recorded video scene to the displayed scenario the optical flow of the background (experimental environment, e.g. monitors) and foreground (individual objects, e.g. hand, pen) have to be calculated and separated. The optical flow of the backgrounds provides information about the head movements and has been used to classify the position of the frame in the global coordinate system. The initial frame sequence is exemplarily shown in figure 5 followed by figure 6, where the frames are relocated considering the individual head movement.

Using the final aggregate picture of the scene camera, the local coordinates of the tracked eye movements was transformed to the global coordinate system and mapped to the displayed scenario. During the eye tracking process each 33 ms the location of the focal point is determine and rated as a fixation or not. These fixations were aggregated, if the radius of the display location correlates with the frequency of the stored fixations (figure 7).


E. Questionnaire method

The questionnaire method is designed to freeze the simulation run immediately after randomly selected clearance situations with 6 till 8 interceptions per execution run (figure 2). The monitors are turned full black during the brake (black freeze) and the test person shall fill out the questionnaires form, namely the “regular questionnaire”. Related task was to estimate the importance of the given set of visual cues (see figure 3) on a scale from 1 (no importance at all) to 4 (very important) that were used for the last clearance decision according to the definition of $p_{\text{VQ}}$. The questionnaires form is limited to elementary cues (cf. visual cue model) to assure comparability to the eye tracking data.

IV. RESULTS

A. Visual cue questionnaire

The scenario execution concluded with the assessment of the demanded visual cues sampling 153 taxi clearances and 74 take-off clearances by regular questionnaires. The test persons used the open part of the questionnaire to propose visual cues that were useful to complement the situational picture during all or individual situations.

Figure 6. Correct localization of frames regarding the global view

Figure 7. Aggregated localization of eye fixations overlayed the scene camera recordings

Figure 8. Results of the questionnaire on the taxi clearance situation

The results of the closed end questionnaire show a varying degree of importance of a visual cue depending on the type. The error bar chart, figure 8, shows selected results of the regular questionnaire for the taxi clearance situation. The information about the selected route is assumed as the most important one, allowing the operator to forecast potential head on conflicts with surrounding a/c and contribution to traffic congestions at the airport bottlenecks. In contrast, all test persons agreed that the altitude of the target a/c is never taken into account when deciding for a taxi clearance and is hence identified as a candidate for exclusion from the visual cue quantity.

Figure 9. Results of the questionnaire on the take-off clearance situation

The relative large standard deviation of visual cues addressing information of other a/c is attributable to individual varying habits of the test person, who monitor the environment more or less intense, which is indictable by lower standard deviation when analyzing test person wise. An individual bias in the estimation is also observable. The experimental environment supported only the clearance by the pen stroke and no voice communication. In deviation to real operations,
there also were no flight strips to be handled where the callsign could have been relevant (callsign of target a/c M=1.61; SD=0.88, callsign of other a/c M=1.27; SD=0.44).

Compared to the taxi clearance results, the take-off clearance showed clear differences regarding the importance of information on the surrounding traffic (figure 9). Other Position (M=3.24; SD=0.89), Speed (M=1.75; SD=1.05), Altitude (M=1.62; SD=1.05) became more important by the circumstance to adjust the take-off clearance with the traffic on the dependent runways. The importance of the cue route decreased by almost half. Obviously, there are no routing activities to be done for an a/c waiting on the take-off position. The test person assumed the point of time “when to clear”, involving dependencies to other runways and expected time of a/c unblocking the runway for a non-conflicted take-off, as the most important cues. A significant difference can be quantified on other a/c altitude when the take-off on runway 18 (M=2.1) and 25C (M=1.3) is separately analyzed. The test person used the altitude and position information of the landing aircraft to anticipate the expected occupancy of 25L.

The open part questionnaire identified a runway coloring function, indicating the runway occupancy, to be a proper function to support the test person during take-off decision making.

B. Failed visual perception detection

The RI event was triggered 43 times in total during take-off and lineup decisions.

The biggest contribution to the occurrence of a hazard can be allocated to the position of other a/c (figure 10). This addresses in particular the position of departing and landing a/c when clearing for take-off. For both the importance and the failure rate, the contribution of the visual cue other a/c plays a dominant role, which is obvious by its safety relevance to represent environmental traffic information.

C. Eye Tracking

To evaluate the potential of visual cues and their relevance to prevent hazardous situations, the application of the eye tracking equipment will be demonstrated at the following example (figure 11). To give the take-off clearance to a/c #1, the test person has to check three different areas of interests to be clear of other aircraft:

- Area 1: inside runway protection zone and climb-out area of runway 25C,
- Area 2: facing intersection of runway 25C and runway 18,
- Area 3: departing from runway 25C.

The progress of two scenarios table 1 and table 2 showed exemplarily the situations instantly before a RI occurs.

![Figure 10. Visual cues and the related relative frequencies of perception failures.](image)

![Figure 11. Area of interests (1, 2, 3) for giving the take-off clearance of a/c #1 on runway 25C at EDDF.](image)

<table>
<thead>
<tr>
<th>TABLE I. SCENARIO A</th>
<th>Action of test person</th>
<th>Time [s]</th>
</tr>
</thead>
<tbody>
<tr>
<td>notices a/c #1 requesting clearance</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>checks for other a/c lined up for take-off on runway 25C</td>
<td>0.5</td>
<td></td>
</tr>
<tr>
<td>checks for departing a/c on the radar, no other departing a/c detected on radar for landing on 25L</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td>clicks clearance symbol at Surface Manager HMI</td>
<td>2.5</td>
<td></td>
</tr>
<tr>
<td>confirms take-off clearance for a/c #1 (failure mode D)</td>
<td>3.7</td>
<td></td>
</tr>
<tr>
<td>causes RI, a/c #1 and a/c #2 performing take-off run on runway 25C at the same time</td>
<td>4.2</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TABLE II. SCENARIO B</th>
<th>Action of test person</th>
<th>Time [s]</th>
</tr>
</thead>
<tbody>
<tr>
<td>notices a/c #1 requesting clearance</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>checks climb-out area of runway 25C, no departing a/c detected</td>
<td>1.1</td>
<td></td>
</tr>
<tr>
<td>scans runway 25C for other a/c, a/c #2 not perceived at all (failure mode C)</td>
<td>1.9</td>
<td></td>
</tr>
<tr>
<td>checks Radar for departing a/c, no other departing a/c detected on Radar</td>
<td>3.0</td>
<td></td>
</tr>
<tr>
<td>clicks clearance symbol</td>
<td>3.5</td>
<td></td>
</tr>
<tr>
<td>scans end of runway 25C for other a/c</td>
<td>5.2</td>
<td></td>
</tr>
<tr>
<td>confirms take-off clearance for a/c #1</td>
<td>5.6</td>
<td></td>
</tr>
<tr>
<td>causes RI, a/c #1 and a/c #2 performing take-off run on runway 25C at the same time</td>
<td>6.7</td>
<td></td>
</tr>
</tbody>
</table>
The following table 3 summarizes the progress with regards to the essential checks for giving the take-off clearance and the performed checks. The eye tracking data clearly indicates the lack of information, since the test person has not focused the relevant areas of interests at the Surface Manager HMI and radar environment.

<table>
<thead>
<tr>
<th>Area of interests</th>
<th>Source</th>
<th>Scenario A</th>
<th>Scenario B</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Surface Manager</td>
<td>NO</td>
<td>YES, a/c not detected</td>
</tr>
<tr>
<td>2</td>
<td>Surface Manager</td>
<td>NO</td>
<td>YES</td>
</tr>
<tr>
<td>3</td>
<td>Radar</td>
<td>YES</td>
<td>YES</td>
</tr>
</tbody>
</table>

V. CONCLUSION

The main purpose of this paper was to investigate a method that identifies safety critical visual cues and that allows for the quantification of the causal relation to the probability of RI in the scope of a risk analysis. A candidate method has been nominated that consists of a HITLS and an representative scenario, promising best conclusions about RI and the related causative failed perceptions. The questionnaire and eye tracking data recorded the critical moment during the decision-making of the novice test person whether a safe clearance can be granted or not. Therein, the regular questionnaire was designed to gather the test person’s estimations of proposed visual cues concerning its importance for decision making. The results showed a diverse mix of the importance with the position information of the target and other a/c as the most important cues. Few visual cues could be obtained by an open part of the questionnaire. This is due to the fact that novices use the provided information by the solution uncritically and without any experienced reference. Visual cues that are to be excluded from the proposed amount were identified according to our expectations (e.g. altitude of taxiing target a/c). It was clearly observable that novices used the application impartial with an observable high degree to learn from the errors. Even simulation bugs were taken into account for clearance decision-making during training. The major drawback of novices is the missing awareness regarding consequences of their decisions. The use of novices as test persons in an overload scenario provoked a large amount of sampled RI events. This is beneficial to collect a surrounding set of failed perception that covers a large bandwidth of cases for analysis for causing pattern that might be symptomatic for the situational development of a RI. However, the internal validity of the results is seriously affected by the use of novices.

The collected subjective and objective data provides the calculation of the contribution of a visual cue to risk $P(R|V_{P_k})$, expressed by a conditional probability for a RI. The parameters for the application of the related formula, section 2, are obtainable by assuming the relative frequency as representative probabilities. The results show the visual cue position (other a/c) by far as the most risk inducing. Beyond this qualitative statement, a problem in the quantification remains the probability-estimation of the visual cue frequency $p_{V_{P_k}}$ by subjective questionnaires on the ‘importance’. Novices have little claims on the absolute accuracy and provide non-calibrated data that allow for a qualitative and relative estimation of the usage. For this reason, we consider the eye tracking camera to be a candidate to gather objective frequencies of visual cue perception for decision making. The camera provided valuable information about visual scan pattern of the test person on the HMI surface. The order of picking up cues during decision making gives detailed information how information was misperceived. The eye tracking camera alone didn’t allow conclusions on the identification of visual cues, since the pure fixation data provides multiple possible cues at a time by one coordinate. However, the combination with questionnaire-based visual cue identification allowed for the limited conclusion and interpretation on the probable intention of the test person and hence the visual cue to perceive. Therefore, our next steps will include further steps for the evaluation of methods that reliably quantify risk by means of HITLS-supported empirical data.

VI. REFERENCES


