

Usefulness of FMECA for improvement of productivity of TWR process

Jan Alexander Langlo, Aslak
Wegner Eide, Amela
Karahasanović, Lisbeth Hansson,
Hans Erik Swenggaard, Bjørn
Andersen
SINTEF ICT
Oslo, Norway
Contact: amela@sintef.no

Theodor Zeh, Stephan Kind
Frequentis
Vienna, Austria

Carl-Herbert Rokitansky, Thomas
Gräupl
University of Salzburg
Salzburg, Austria

Foreword—This paper describes a project that is part of SESAR Workpackage E, which addresses long-term and innovative research. The results of the first experiment conducted in the project are presented here.

Abstract—Improvement of productivity in the control room might help in coping with the expected growth in air traffic and requests for increased safety, predictability, and efficiency of Air Traffic Management (ATM) systems. One possible way of achieving this is to apply productivity improvement methods and techniques and validate them in the context of a tower control room. This paper describes the results of the first study we conducted to investigate the usefulness of process improvement methods in the ATM context. Visualisation of the low-level usage data collected from the simulator has been useful input for the productivity analysis. FMECA (Failure Mode and Criticality Analysis) has been useful as a means for structuring the analysis, but had some limitations when applied to the scenario we used. We plan to continue our work by investigating the potential usefulness of other methods and tools within our four-step productivity improvement process, such as heuristic methods and tools to allow for real-time and on-line response.

Keywords—Component; Air traffic management, Tower control room, Productivity improvement

I. INTRODUCTION

Meeting the increasing demand for improved Air Traffic Management (ATM) systems in general and in Europe in particular requires the development of new technologies. The importance of transport and mobility to the economy has been recognised by the European Commission [1]. The SESAR® program, a combination of the European Commission, EUROCONTROL, Air Navigation Service providers (ANSP), and industrial effort, is vital for the development of robust and future-oriented solutions. In addition, several research initiatives have been taken, and numerous technologies have been developed in recent years. The results of such inquiries have been presented to the ATM community through ATC Global and EUROCONTROL's Innovative Research Workshop, among others.

In spite of these efforts, one area continues to leave significant room for improvement: productivity in the control room itself. Different approaches to process improvement have been proven successful in mass production industries, such as the automotive industry, and are now increasingly used in other domains.

In order to improve the productivity and safety of the highly-automated ATM control room, we have proposed a four-step productivity process called the Zero Failure Management at Maximum Productivity in Safety Critical Control Room process (ZeFMaP) [2], which incorporates permanent improvement cycles.

The four steps of ZeFMaP include the following (Figure 1):

1. Modelling the target process into a production workflow and dividing it into 'production steps'
2. Optimizing the 'human machine symbiosis' for each step (outside the scope of our research)
3. Analyses of the decision points and decision content within each of the steps, with the aim of offline optimization for each decision of the overall process and the improvement of each production step through a feedback loop
4. Improvement of the target process through a feedback loop

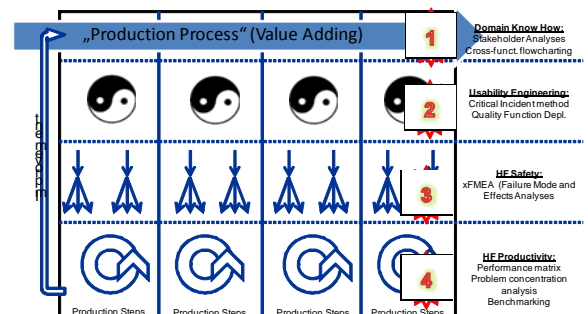


Figure 1: ZeFMaP Process

The ZeFMaP process is a four-step improvement process that will apply best practices from productivity improvement in mass production industries and adapt it to meet the challenges of ATM.

Our hypothesis is that the implementation of such a method should permanently improve the quality of the processes in the control room by optimising productivity and minimising decision failures. In this project, we are testing this hypothesis within the context of a tower (TWR) control room. Numerous improvement approaches, methods, and tools have been developed [3]. In our previous work, we identified the advantages and disadvantages of existing improvement approaches used in the ATM context, and modelled the control tower processes in a way suitable for the application of productivity improvement methods and techniques [1].

Here we present the work we have done to investigate the usefulness of FMECA [4,5] as a tool for optimisation of the decision-making process (step 3 of the ZeFMaP process). We conducted an experiment with five controllers using real (past time) traffic data from Hamburg TWR's control room. The experiment was conducted using the University of Salzburg's simulator NAVSIM and the electronic flight-strips tool developed by Frequentis. The collected data are analysed by using the FMECA approach.

The remainder of the paper is organized as follows. Section II describes the study we conducted. Section III describes the analysis and presents the results. Section IV discusses the results and concludes.

II. METHOD

The primary objective of the experiment reported in this paper was to collect the necessary data input into the ZeFMaP productivity improvement process. The secondary objectives of the experiment were (1) to evaluate the internal and external validity of the selected measures and experimental design in each iteration of our evaluation process (pre-tests and experiments), and (2) to increase knowledge on how human and system performance in the domain of ATM can be measured. To achieve these objectives, we conducted a set of real-time simulation exercises, where air traffic controllers were subjected to realistic work scenarios that were played out in real-time. The experiment design and the preparation of the experiment material followed the procedures described in [6-11]. The technical report with a detail description of the experiment design, analysis, results, and material can be found in [12].

A. Experiment setting

To enable a realistic simulation of the air traffic at Hamburg Airport, the experiment made use of the NAVSIM air traffic simulator at the University of Salzburg (USBG). By using this simulator, we were able to control a number of different attributes related to the simulated air traffic scenarios, which provided us with a high degree of experimental control.

The independent variables that were controlled during the experiment include, among others: traffic scenario (the air traffic simulated in the experiment was based on real traffic data from Hamburg Airport, taken from the peak hours on a specified set of dates); weather conditions; controller working position; ATC procedures and roles; and team structure. To replicate the Hamburg TWR environment, the NAVSIM simulator was set up to make use of electronic flight strip technology. The simulation involved neither simulated pilots nor participating pilots, and hence the controllers did not have to handle communication with pilots. Instead, the simulation was held as if there would be a controller-pilot data link communication (CPDLC) connection to the aircraft—performed directly from the electronic flight strips. The technical implementation of the simulation environment is presented in [13].

No other systems or frame conditions such as weather were included. The system was further limited in time and volume by the experiment. Total elapsed time of the measured session was 37 minutes and 29 seconds. One flight (DLH150) was deleted due to inconsistent data, reducing the departure flights to 27 in total. Arrival flights were 10 in total.

B. Experiment participants

The experiment involved a sample of five participants, selected by means of convenience sampling. Two participants were experienced air traffic controllers who had experience as tower controllers. One participant was an experienced air traffic controller who had no experience with working in the tower. One participant had not yet finished the full ATC education, and had no experience with working in the tower. The last participant had experience in working as a clearance delivery controller (CDC), but had no education or experience with working as a fully licensed air traffic controller.

C. Experiment procedure

The first experiment was conducted over two days in June 2012. The experiment consisted of a training session that took place on the first day of the experiment, and a measuring session that took place on the second day. Additionally, there was a preparation period prior to the experiment, where the controllers were familiarized with the TWR process and infrastructure at Hamburg Airport. The measured simulation session included several simulation runs (lasting from 10-40 minutes) where the controllers were subjected to realistic work scenarios that were played out in real-time. The task of the controllers was to control the air traffic as they would normally do while working in the TWR.

D. Measurement

1) Subjective data collection methods and tools

The data used as input to the optimization process were mainly gathered by subjective data collection methods including interviews, questionnaires, and observations. During

the simulation runs, observers took notes using a predefined observer template, registering specific events (e.g., handovers, optimum timed decisions), and timestamps for when they occurred. At the end of each run we conducted individual interviews with each participant. The interviews lasted for 20 minutes, and a set of open-ended questions was asked. Following the last run of the experiment, the participants were also asked to fill in a post-run questionnaire. The focus of the questionnaire and the interviews was on the measures given in the list below. A debriefing session was conducted at the end of the second day. The goal of the debriefing session was to give the participants an opportunity to discuss their experience and come forward with opinions regarding issues such as workload during the experiment, appropriateness of the experiment material, and stress level. The experiment also involved video, screen and audio recordings, which were analysed after the experiment. In Table I we describe the subjective measures.

TABLE I. SUBJECTIVE MEASURES

Measure	Description
Subjective workload	Measuring safety
Optimum timed decisions	Input to the optimization analysis
Time consuming decisions	Input to the optimization analysis
Effects of decisions on throughput	Input to the optimization analysis
Amount of changes of priority which led to significant negative consequences.	Input to the optimization analysis
Amount of changes of priority which led to significant positive consequences.	Input to the optimization analysis
Team performance	Input to the optimization analysis

2) Objective data collection methods and tools

The productivity aspects were mainly assessed using objective data collected by the logging functionality of the NAVSIM simulator. The recorded data were focused on the controller's interactions with the system (the actions they performed), logging of movements (taxi segments), and their timestamps. In addition, we recorded screen captures of the radar screen, and the smart strips screen of each individual controller. This data source was used to verify and explain the findings from the log files, and was particularly useful for extracting handovers of flights from one controller to another. In Table II we describe the objective measures that were obtained.

TABLE II. OBJECTIVE MEASURES

Measure	Description
Commands with timestamps	All commands issued to the system by a controller with timestamps
Total number of handovers	This is calculated based on the analysis of screen captures and audio files

Number of handovers per role/sector	This is calculated based on the analysis of screen captures and audio files
Frequency of handovers	This is calculated based on the analysis of screen captures and audio files
Time between handovers	The time between the last action done by controller i and the first action done by a controller j on the same plane
Total number of airplanes handled per exercise	Time frame is first departure from Hamburg until last arrival at Hamburg in the simulation. Only A/C arriving or departing at Hamburg are counted
Total number of commands issued per role/controller	Number of commands issued by a controller i
Taxi time from gate to runway (departures) including details on all taxi way segments.	Times for all segments from gate to runway
Taxi time from runway to gate (arrivals) including details on all taxi way segments	Times for all segments from gate to runway
Taxi distance from gate to runway (departure)	Must take into account predefined routes vs. improvisations.
Taxi distance from runway to gate (arrivals)	Must take into account predefined routes vs. improvisations.
Amount of system input per hour/sector	AIHi -number of commands issued by the controller i per hour
Number of corrections to decisions	Objective: Input to the optimization analysis
Amount of time per hour/exercise with no activity	Sum of periods longer than 5 minutes without any activity (commands issued by the controller or by the aircraft)
Delay per flight in arrival and departure	Diff between actual and scheduled time per flight

E. Post experiment assessment

Our study was exploratory and the ability to generalize from our findings (external validity) was limited. The main threat to the external validity was that the participants in the experiment might not be representative of the air traffic controller population. To minimize this likelihood, the participants were provided with training material describing the Hamburg Airport infrastructure and procedures. Further, the first day of the experiment was devoted to training.

III. ANALYSIS AND RESULTS

The overall purpose of the data analysis in this experiment was to answer:

- How suitable is FMECA as an improvement tool in ZeFMaP improvement process step 3? We start this

investigation by testing the visualization of logging data and FMECA as possible improvement tools.

- What is the quality of the conducted decisions, and how do the decisions influence the performance of the operations?

The collected qualitative data (i.e., observations, interviews, questionnaires) were coded and analysed. The coding schema for coding the screen-capture files and audio files was developed on the basis of the given metrics. For analysis of quantitative data we used different tools, including Excel, Visio, and the statistical tool IBM SPSS Statistic.

To increase the validity of our results, we used method triangulation. When possible, several data collection methods were used to collect the same data, allowing us to check them against each other. Procedures were defined for resolving any disagreements that might appear.

The analysis of the data needed for process improvement in this experiment was exploratory. The above-described subjective and objective measures were used as input to this analysis. The data from the interviews were used to understand the relation between different variables and their effects on productivity.

A. Data cleaning and preparation

Although low-level usage data provide a rich source of information on interactions between humans and technology, analysis of such data is a highly demanding task [7]. Due to the exploratory nature of this study, we applied an approach that integrates application-specific and data-driven analysis [14]. More specifically, we conducted an analysis of the logging data as a sequence of rapid iterations. In each of these iterations the level of detail and the nature of the extracted data were justified through a dialogue between researchers working on the analysis and the developer of the simulator.

The data cleaning and preparation of logging files consisted of the following steps:

- Synchronizing the logging data with the screen captures. The screen capture videos were synchronized with the simulation time (from the radar screen capture) using Kinovea (<http://www.kinovea.org/en/>). An event (a command issued by a controller on a specified flight) has been selected as a synchronization event and identified in both video files (smart strips screen captures, radar screen captures, and log files).
- Coding the handover commands. The handovers of flights were coded based on the screen capture videos of the smart strip tools. The actual handover was identified from the screen capture, and the corresponding timestamp was identified from the radar screen capture.

- Extending the file manually with all of the commands, using: (i) the handover commands, (ii) the 'PLANE arrived AT GATE/RUNWAY', command and the corresponding times (extracted separately from the simulator), (iii) the time between the command and a previous command issued by the same controller (calculated by the scripts we developed).
- Creating visual representations of the actions done by different controllers. Different visual presentations of the data were made in Excel. These graphs were used to explore the data and identify errors in the data set.

B. FMECA – traditional design using quantitative data

The cleaned data were analysed using failure mode, effects and criticality analysis (FMECA) [4,5]. The purpose was to reveal whether or not FMECA as a statistically-based productivity tool is suitable for ZeFMaP improvement process step 3. The analytical approach to this was to investigate whether the effect of the controllers' decisions could be graded as optimal or less than optimal.

1) Quantitative data relevant for the FMECA method

Failure mode criticality assessment may be qualitative or quantitative. For the quantitative assessment relevant for this experiment, model criticality number C_m is calculated for each failure mode of each item, and item criticality number C_r is calculated for each item.

The criticality numbers are computed as:

$$C_m = \lambda_p \alpha \beta t$$

and

$$C_r = \sum_{n=1}^N (C_m)_n$$

Model criticality number C_m is calculated for each failure mode of each item, and item criticality number C_r is calculated for each item. The criticality numbers are computed using the following values:

- Basic failure rate λ_p
- Failure mode ratio α
- Conditional probability β
- Mission phase duration t

For graphical analysis, a criticality matrix may be charted using either C_m or C_r on one axis and severity code on the other.

2) FMECA definitions for the ZeFMaP setting

System: The experiment's system is defined by the sum and effect of the five controllers' decisions. All other systems or frame conditions were excluded. The system is further limited in time and volume by the flights chosen for the experiment. Total elapsed time is 37 minutes and 29 seconds. Flight DLH150 was deleted due to inconsistent data, reducing the departure flights to 27 in total. Arrival flights are 10 in total.

Item: Each of the roles participating in the experiment: APRON1, APRON2, GND, TWR, and CDC.

Failure modes for each item are listed in the table below. These failure modes are created theoretically to form a default for the analysis in this first experiment:

TABLE III. FAILURE MODES

CDC/APRON	TWR	GND
<ul style="list-style-type: none"> Delay in strip take over Delay in push back clearance Rejected push back clearance (CTOT was valid) Departure/Push back clearance to invalid CTOT Delay in taxi clearance Rejected taxi clearance (CTOT was valid) Taxi clearance to invalid CTOT Choice of suboptimal taxi way. Delay in 'Contact TWR command' Delay in 'Contact GND command' Delay in issuing hand over to TWR Delay in issuing hand over to GND Delay in issuing hand over to TWR 	<ul style="list-style-type: none"> Delay in strip take over from APRON Invalid take off clearance (runway not clear) Rejected take off clearance although runway was clear Invalid line up clearance (runway not clear) Rejected line up clearance although runway was clear Delay in strip take over from GND 	<ul style="list-style-type: none"> Delay in strip take over from APRON Invalid taxi clearance (taxiway not clear) Rejected taxi clearance although taxiway was clear Choice of suboptimal taxi way. Delay in issuing contact TWR command Delay in handover to TWR

Basic failure rate: Number of decisions (with failure or success) per total number of decisions for each role (item)

Failure mode ratio: Ratio of one failure mode out of all failure modes for that role (item)

Conditional probability: An estimated probability describing the degree of each specific failure mode's effect on each KPI. Since this is a controlled experiment and we know the result, this factor can be either 0 or 1.

Mission phase duration: Not applicable and set to 1. It must be noted that an inclusion of environmental KPIs will require a graded scale for this factor.

Severity: Judged for each of the KPIs. Capacity = airport capacity

- Efficiency—temporal efficiency and fuel efficiency
- Flexibility—business trajectory update
- Predictability—on-time operations and service disruption effect
- Safety

Severity codes: severity is ranged on a four graded scale:

- Crises: 0
- Bad: 0,3
- Medium: 0,6

- Good: 1,0

C. Results from FMECA

We did not find any arrival-related decision that could be related to the failure modes. Within the frame of the FMECA model as used in this project, all decisions must consequently be considered optimal.

Failure modes were found among the departure-related decisions. The following analysis is therefore limited to the departure flights. The total number of decisions taken by all roles in the experiment related to departures is 244.

Divided according to roles:

- APRON (1 and 2): 76
- Ground: 9
- Tower: 78
- CDC: 81

TABLE IV. OBSERVED DECISIONS AND FAILURE MODES

ID	Flight	ROLE	Failure mode	Description of observation	KPI effected	Severity code
P01	KN1744	CDC	Departure clearance to invalid CTOT	Releases several flights at 15:32:28. KN1744 is affected with ATOT at -5,30min pre slot time	Predictability	P0,6
P02	AFR2211	CDC	Departure clearance to invalid CTOT	Releases AFR2211 at 15:37:12, affected with ATOT -6,23min pre slot time	Predictability	P0,6
P03	DLH053	CDC	Departure clearance to invalid CTOT	Releases DLH053 at 15:37:49, affected with ATOT -9,65 min pre slot time	Predictability	P0,6
P04	HLX8HD	CDC	Departure clearance to invalid CTOT	Releases HLX8HD at 15:39:20, affected with ATOT -5,3 min pre slot time	Predictability	P0,6
P05	BER503	CDC	Departure clearance to invalid CTOT	Releases BER503 and the remaining 15 flight at 15:41:36. BER503 is affected with ATOT 17,63 min pre slot time	Predictability	P0,3
P06	DAT42WU	CDC	Departure clearance to invalid CTOT	DAT42WU is affected with ATOT -40,15 min pre slot time	Predictability	P0,3
P07	DLH055	CDC	Departure clearance to invalid CTOT	DLH055 is affected with ATOT -11,43 min pre slot time	Predictability	P0,6
P08	DLH087	CDC	Departure clearance to invalid CTOT	DLH087 is affected with ATOT -25,12 min pre slot time	Predictability	P0,3
P09	DLH3WY	CDC	Departure clearance to invalid CTOT	DLH3WY is affected with ATOT -8,88 min pre slot time	Predictability	P0,6
P10	BER905	CDC	Departure clearance to invalid CTOT	BER905 is affected with ATOT -12,83 min pre slot time	Predictability	P0,6
P11	HLX4W	CDC	Departure clearance to invalid CTOT	HLX4W is affected with ATOT -37,07 min pre slot time	Predictability	P0,3
P12	DLH3RT	CDC	Departure clearance to invalid CTOT	DLH3RT is affected with ATOT -36,13 min pre slot time	Predictability	P0,3
P13	BER80W	CDC	Departure clearance to invalid CTOT	BER80W is affected with ATOT -24,83 min pre slot time	Predictability	P0,3
P14	BER66Z	CDC	Departure clearance to invalid CTOT	BER66Z is affected with ATOT -8,23 min pre slot time	Predictability	P0,6
P15	DLH3FP	CDC	Departure clearance to invalid CTOT	DLH3FP is affected with ATOT -7,55 min pre slot time	Predictability	P0,6
P16	DLH6UJ	CDC	Departure clearance to invalid CTOT	DLH6UJ is affected with ATOT -11,77 min pre slot time	Predictability	P0,6
P17	DARIUS	CDC	Departure clearance to invalid CTOT	DARIUS is affected with ATOT -11,30 min pre slot time	Predictability	P0,6
P18	BER724	CDC	Departure clearance to invalid CTOT	BER724 is affected with ATOT -20,15 min pre slot time	Predictability	P0,3
P19	HBVNV	CDC	Departure clearance to invalid CTOT	HBVNV is affected with ATOT -8,73 min pre slot time	Predictability	P0,6

Flights not considered in the table above have been found to be free of failure modes and must consequently be regarded as optimal decisions. However, due to the limited number of flights in the experiment, there are no indications of proximity to any bottlenecks. We anticipate that there is a considerable 'grey area' of optimal decision that could have turned out differently in a more intense scenario. This correlates with the general assumption that computerized optimization does not make any difference until a system is running at 80–100% of its theoretical maximum performance.

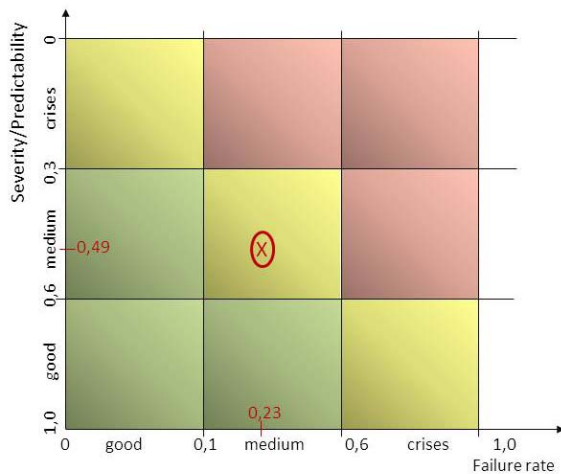
The apron controllers' queuing of flights following the massive release from CDC can be summarized as approximately 45 minutes of unnecessary engine idle time for

28 departing flights. It is recommended to include the environmental KPIs in the next experiments. By its nature, KPI flexibility cannot be 'computerized' or automated. It must be noted that the apron controllers demonstrated flexibility by 're-queuing' the flights to a feasible take-off sequence.

Calculation: The analysed non-optimum decisions are all related to one 'item', namely the CDC. All decisions were the same failure mode, which was 'Departure clearance to invalid CTOT'. The basic failure rate is $19 \div 81 = 0.23$. There is no theoretical or empirical information on possible failure modes per item. This experiment could theoretically have revealed several modes per item, but that was not the case. Consequently there is only one failure mode, giving a failure mode ratio of 1. The item (CDC) criticality number is therefore

0.23. Given the single item and single failure mode, the severity codes for predictability are averaged to 0.49.

TABLE V. SEVERITY/PREDICTABILITY GRAPH



IV. DISCUSSION AND CONCLUDING REMARKS

It has been stated that FMECA is not particularly suited for multiple-failure scenarios or unplanned cross-system effects [15]. This indicates some weak aspects of the method when applied to ATM, which is complex and multi-dependent by nature. However, there are some factors that deserve to be discussed. The most important are probably:

- Statistical significance and validity of the collected data
- Grading and scales
- Statistical optimum

Statistical significance and validity of the collected data

Our limited scenario with 37 movements in 37 minutes and 29 seconds was obviously a fairly easy task for the ATCOs participating in the experiment to perform. There were only 19 non-optimal departure-related decisions observed, consequently making the remainder of 225 optimal decisions for the entire departure sequence. This ratio is obviously questionable, as remarked above.

The approach to 'isolating' the non-optimal decisions requires a challenging scenario that leaves a relatively small number of assumed optimal decisions for closer investigation. One alternative is of course to study all decisions made by a team of ATCOs experienced with conditions at Hamburg Airport. This is very labour intensive, but could reveal data sufficient to make an 'inverse' FMECA to separate the good, the very good, and the optimal decisions. Without having made a sensitivity analysis, it is reasonable to state that one must be prepared to investigate a significantly higher population than 244 decisions to find significant and relevant data. FMECA rests upon inductive reasoning, a kind of reasoning that constructs or evaluates general propositions that are derived

from specific decisions. In order to find generally valid and computable data, huge samples related to every hour of operation, every season, and every variant of throughput and timetables must be investigated.

Grading and scales

The scales for criticality number and severity are created by judgment to form a default for the analysis in this first experiment. Given a number of failure modes and severities, they would be separable and computable even by such a default scale. But if the results were to be implemented in a system and thus formed a vital factor for decision support, more comprehensive work must be done on grading. It is obvious that a criticality number of 0.23 and a predictability severity code of 0.49 do not necessarily reflect the reality.

Statistical optimum

The limited scenario with 37 movements in 37 minutes and 29 seconds did not put any observable stress on the ATCOs, leaving a relatively high number of decisions as optimal. As described above, one might investigate this more closely. With some effort spent, one would most likely find a statistical optimum.

Conclusions and future work

The results and discussion show that the FMECA method is of limited usefulness in ZeFMap improvement process step 3. FMECA requires statistically representative data for each scenario. A large number of scenarios should be included to address variables such as changing weather conditions, changes in timetables, changes in operating procedures related to different wind directions, etc. As a consequence, the use of FMECA in this context will require an extensive effort to collect reliable data. Thus, FMECA has a limited ability to separate optimal from non-optimal decisions; however, it has been useful for structuring the overall analysis. The results should thus be treated cautiously. The traffic scenario we used was limited (37 movements within 37 minutes and 29 second). Although this allowed us to carefully test the software that was used (NAMSIM simulator and electronic flight strip tool), it had a negative impact on the usefulness of the data we collected. The tasks were too easy for the controllers, which resulted in a limited number of non-optimal decisions.

Future studies should be conducted with higher intensity traffic samples, preferably in the range of 80–100% of an airport's theoretical maximum capacity. Statistical optimisation methods should be included to analyse data and results. To provide computerized support for optimal decisions in a dynamic and complex context, heuristic optimization methods and tools should be utilized.

ACKNOWLEDGMENT

This work is co-financed by EUROCONTROL acting on behalf of the SESAR Joint Undertaking (the SJU) and the EUROPEAN UNION as part of Work Package E in the SESAR Programme. Opinions expressed in this work reflect

the authors' views only, and EUROCONTROL and/or the SJU shall not be considered liable for them or for any use that may be made of the information contained herein. All authors contributed equally to the paper. Stephan Kind has recently left Frequentis. The authors are grateful to Anna Wennerberg from EUROCONTROL and Tobias Wolff from Avinor for their help with the experimental design. We also thank Stian Stør Ødegård from the University of Oslo for testing the experimental material, Barbara Jandl, Elias Pschernig, and Martin Mayr from the University of Salzburg and Volker Grantz from Frequentis for helping with the conduct of the experiment. We are particularly grateful to the controllers who participated in the experiment for their valuable effort and patience.

REFERENCES

- [1] Zeh, T., Grantz, V., Kind, S., Rubenser, R., Karahasanović, A., Andersen, B., Hansson, L., Langlo, Swengard, H.E., (2011) Zero Failure Management at Maximum Productivity in Safety Critical Control Room, in Schaefer, D. (ed) Proceedings of the SESAR Innovation Days 2011, EUROCONTROL. ISBN 978-2-87497-024-5.
- [2] EC White Paper – Roadmap to a Single European Transport Area – Towards a competitive and resource efficient transport system, COM (2011) 144, March 2011.
- [3] B. Andersen, Business process improvement toolbox, 2nd ed., Milwaukee: ASQ Quality Press, 2007.
- [4] Mikulak, R.J., McDermott, R., Michael Beauregard, M., The Basics of FMEA, 2nd Edition, Taylor and Francis Group, USA, 2009.
- [5] Nahman, J.M., Dependability of Engineering Systems: Modeling and Evaluation, Springer, 2002
- [6] Shadish, W.R., T.D. Cook, and D.T. Campbell, Experimental and quasi-experimental designs for generalized causal inference. 2002: Houghton Mifflin.
- [7] Lazar, J., J.H. Feng, and H. Hochheiser, *Research methods in human-computer interaction*. 2009: Wiley.
- [8] Wohlin, C., Experimentation in software engineering: an introduction. 2000: Kluwer Academic.
- [9] Baber, C. (2005). Human Factor Methods, A practical guide for engineering and design. Ashgate Publishing
- [10] Rothaug, J., Age, Experience and Automation in European Air Traffic Control – Survey in the ECAC Area. 2004: Eurocontrol.
- [11] Hart S., Staveland, L. Development of NASA-TLX (Task load index): Results of empirical and theoretical research. In Hancock, P. and Meshkati, N. (eds.) Human mental workload, Holland, 1988, pp. 139-183.
- [12] Eide, A.W et al., E.02.16-ZeFMaP-D2.2-Evaluation plan and results (Phase I), SESAR WP-E, technical report.
- [13] T. Gräupl, B. Jandl, and C.-H. Rokitsky, Simple and Efficient Integration of Aeronautical Support Tools for Human-In-the-Loop Evaluations, in Proc. ICNS'12, 2012.
- [14] Thomas, R., Karahasanovic, A., Experience with an extensible workspace of low-level usage data, Software Practice and Experience, 2009, vol. 39, pp 1185-1213.
- [15] Associate Administrator for Commercial Space Transportation (AST), 'Destination FAA', 2004.
http://www.faa.gov/about/office_org/headquarters_offices/ast/about/media/032504.pdf