A Post-simulation Assessment Tool for Training of Air Traffic Controllers

Aslak Wegner Eide, Stian Støer Ødegård, and Amela Karahasanović

SINTEF, Oslo, Norway
{aslak.eide,amela.karahasanovic}@sintef.no,
stiancid@outlook.com

Abstract. This paper proposes a post-simulation assessment tool that aims to improve the training of air traffic controllers (ATCOs) by visualizing their performance. The tool helps the controllers to identify bottlenecks in flight traffic and find alternative solutions that might improve traffic throughput. The usefulness of the tool was evaluated in a study involving benchmark tests and interviews with five experienced ATCOs. The results from the study indicate that the tool can help ATCO students to (1) identify irregularities in their work, (2) find possible underlying causes of these irregularities, and (3) find alternative solutions preventing these irregularities. Visual feedback consisting of workflow graphs and radar replays might generate valuable insights that enable self and peer assessment during ATCO training. Our results might be interesting both for the practitioners working with ATCO training and for researcher investigating the effects of visualization in education.

Keywords: visualization, air traffic control, training, real-time simulation.

1 Introduction

The continuously growing demand in air transport has heightened the need to improve productivity in air traffic control. Coping with this challenge requires not only new automation tools and enhanced procedures, but also a rethinking of air traffic controllers (ATCOs) training [1-2]. In order to achieve the required learning effects, existing training programs need to integrate hands-on training with knowledge acquisition and skill development [3]. Training should not only teach users how new tools should be used, but also help overcome resistance to change.

In today's practice, real-time simulation (RTS) of work scenarios is regularly used as a cost-effective way of training new and experienced ATCOs. One of the great advantages of RTS, compared with other learning aids, is the ability to freeze and replay scenarios directly, enabling instructors to provide timely feedback on a given traffic situation and on the quality of the decisions made by the trainee [4]. However, the delivery of such feedback requires the full attention of an instructor, and may also cause disruptions in the internal planning process of the trainees, making it difficult for them to progress and positively reinforce their learned skills [5]. Furthermore, shortage of time, inadequacy of the feedback, fear of failure, and negative environments might also cause problems during the ATCOs training [5].

S. Yamamoto (Ed.): HIMI 2014, Part I, LNCS 8521, pp. 34–43, 2014.

In light of the abovementioned challenges, this paper proposes a post-simulation assessment tool that aims to improve the training of ATCOs by allowing them to review their own work, and the work of others, without the need for an instructor to be present. The tool achieves this by providing visual feedback that help ATCOs identify bottlenecks in their work and alternative solutions that could result in higher traffic throughput. Our study positions the tool as a useful supplement to RTS, which can help mitigate the challenge of obtrusive feedback during ATCO training.

The training tool is described in Section 2, followed in Section 3 by a description of the study conducted to evaluate the tool. The results from the study are presented in Section 4. A discussion of the results in the context of current research is given in Section 5. The final conclusions of the paper are given in Section 6.

2 Post-simulation Training Tool

To improve ATCO training, we developed a post-simulation assessment tool that help ATCOs review their workflow and performance immediately following an RTS. The tool was developed through an iterative and user-centered design process involving requirements engineering, prototyping, and end-user evaluations. Throughout this process, the tool evolved from paper-based prototypes, into the functional software application presented here. The final version of the tool consists of 3 components: (i) a workflow graph, (ii) a radar image, and (iii) a set of timeline controls (see Fig. 1). The workflow graph shows the flights that the ATCOs handled during the simulation, distributed along a global timeline. Each flight is visualized by a bar in the graph.

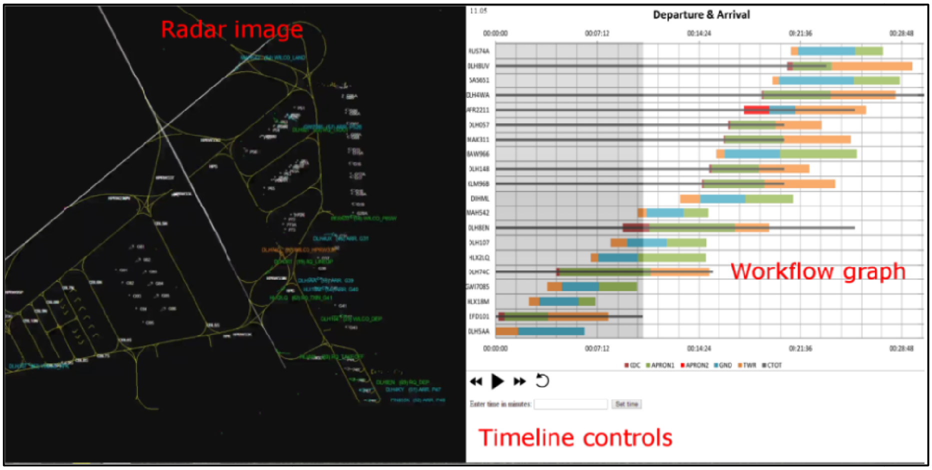

Fig. 1. Overview of the training tool

For departing flights, the bar runs from the point in time when the flight first requested departure clearance, to the point in time when the flight is no longer touching the runway (see Fig. 2). The end of the black horizontal line shows the calculated take-off time (CTOT) of departing flights, which is the point in time when the flight should (according to the flight plan) leave the runway. For arriving flights, the bar runs from the point in time when the flight was given clearance to land, to the point in time when the flight has arrived at gate (see Fig. 3). The bar is divided into blocks with different coloring, depending on which controller managed the flight at the time. In a control tower, responsibilities are often divided among the following roles: Clearance Delivery (CDC), Apron (APRON1 and APRON2 in our case), Ground, and Tower.

Fig. 2. Visualization of a departing flight in the workflow graph: The bar should be read from left to right. The colored blocks in the bar show which controller that managed the flight at the given time periods. The first (leftmost) block represents the timespan when the flight was managed by the CDC controller. Block two, three and four represents the timespans when the flight was managed by the APRON2 controller, the GROUND controller and the TOWER controller.

Fig. 3. Visualization of an arriving flight: The bar should be read from left to right. The colored blocks in the bar show which controller that managed the flight at the given time periods. The first (leftmost) block represents the timespan when the flight was managed by the TOWER controller. Respectively, block two and three represents the timespans when the flight was managed by the GROUND controller and the APRON1 controller.

The radar image shows a video playback of the movements of aircrafts during the given timeframe. A trainee can pan and zoom the radar image, and use the timeline controls to move back and forth along the timeline. The timeline controls include a play/pause button for playing and pausing the time, fast forward and fast rewind buttons for moving back and forth in time, and a numeric field for jumping to an exact point in time. The timeline and the radar screen are synchronized according to the selected point in time. The time that has passed is colored gray in the workflow graph. In combination, these features help the trainee to identify non-optimal flights, and reason about the underlying causes. The standard use-case of the tool is given below:

1. The ATCO starts the tool and selects the training simulation or the work period he/she wants to review. The tool shows the corresponding graph and radar image.
2. The ATCO studies the graph to identify bottlenecks and potential for improvement (e.g., delayed flights, flights that waited too long to get departure clearance).
3. The ATCO moves the timeline to the point in time when the flights identified in step 2 were handled. The radar image shows the airport state at the time.
4. The ATCO reviews the chain of events for the given flights using the radar screen image and the graph, and tries to understand the causes of the irregularities.
5. The ATCO identifies the causes of the irregularities, and tries to come up with an alternative solution for how the irregularity could have been avoided.

3 Research Method

The usefulness of the training tool was assessed in a study at Gardermoen airport tower. The study made use of individual benchmark testing [6] where participants used the training tool to review a flight traffic scenario from Hamburg airport, and open-ended interviews. A sample of five experienced ATCOs took part in the study, selected by means of convenience sampling [7]. The participants differed somewhat in their level of experience. Three participants were certified ATCOs at Gardermoen tower and two participants were currently undergoing training to achieve such certification. All participants were experienced ATCOs, with working experience ranging from 7 to 14 years (median 10). The study lasted approximately 35 minutes per participant and consisted of three parts that were conducted individually for each participant. In the first part we introduced the participant to the training tool, explaining its purpose and functionality. The participant was also given an introduction to the airport layout in the scenario that was to be used during the benchmark testing, and was given an explanation of which position is responsible for which area at this airport.

In the second part, the participant was seated in front of a computer running the high fidelity prototype of the training tool, loaded with the predefined scenario from an earlier RTS training session involving Hamburg airport. As soon as the participant was ready, he/she was then asked to solve a set of predefined benchmark tasks by using the training tool (see Table 1 for an overview of the tasks). The benchmark tasks were designed to assess the tool's ability to generate insight, enabling ATCOs to identify non-optimal flights and reason about them. Due to the complexity of the scenario, there were several correct answers for the second and third task. For data collection we used the think-aloud protocol [8]. The participants were encouraged to say what they were looking at, thinking, doing, and feeling. This was audio recorded. In addition, we collected screen captures and observer notes.

Table 1. Overview of benchmark tasks

#	Benchmark task	Expected solution
1	Find the flight with the largest delay.	The flight with the largest delay was flight DLH8UV (we refer to a flight by using its unique flight identifier number). This task had only one correct answer.
2	Find a possible cause for the delay of this flight.	One main cause of the delay was that flight DLH4WA was given departure clearance before flight DLH8UV (even though DLH4WA had a later CTOT than DLH8UV), causing DLH8UV to have to wait in line behind DLH4WA.
3	Find an alternative solution where this delay could have been avoided.	One main alternative solution was to prioritize flight DLH8UV before flight DLH4WA, and thus allow DLH8UV to take off before DLH4WA.

The last part of the study consisted of individual interviews with each participant. During the interviews, the participant was asked open-ended questions designed to trigger subjective reflection on the usefulness of the tool, particularly in terms of its ability to facilitate learning in tower control rooms. The participant was also encouraged to come up with suggestions for tool improvement, and was asked about his/her experience with similar tools. The interviews were also recorded by audio.

To assess whether the training tool was capable of generating valuable insight, we analyzed the correctness of the solutions that the participants gave to each of the three benchmark tasks. We transcribed the audio recordings from the think-aloud protocol and registered the participants' interactions with the training tool by replaying the screen capture recordings. The transcriptions and interactions of each participant were then sorted chronologically in a spreadsheet, allowing the researchers to take both the thoughts and the interactions of the participants into account when assessing their solutions to benchmark tasks. For the first task, which only had one correct solution, the participants' solutions were compared with the correct solution. For the second and third task, which had several possible solutions, the researchers judged the correctness of the solution.

The spreadsheet containing transcriptions and interactions was also used to determine which strategies the users made use of to solve the various benchmark tasks. This was achieved by comparing the participants' interactions with the tool during each of the benchmark tasks, while looking for patterns in their style of interactions.

The audio recordings from the interviews were transcribed. The researchers then summarized the answers from the participants for each interview question.

4 Results

The results from the study indicates that the proposed training tool helped the participating ATCOs to identify non-optimal flights, to reason about possible underlying causes, and to find alternative solutions. The detailed results from the benchmark tests and the interviews are described in the sections below.

4.1 Benchmark Tests

Figure 4 gives a summary of the benchmark results, showing the number of plausible solutions identified by each single participant for each of the three benchmark tasks. Due to a corrupted screen-capture recording for participant 3 that could not be replayed, we were unable to adequately interpret his proposed solutions to tasks 2 and 3, forcing us to omit these data from the results, and from the overview in Figure 4.

As shown in Figure 4, all five participants managed to find the correct answer to the first benchmark task (identification of the flight with the largest delay). The participants seemed certain in their assessment during this task, and did not come up with more than one solution each. The users' approach for solving the task was to use the workflow graph to identify the flight with the longest visual distance between the actual take-off time (ATOT) and the CTOT. The users did not use the radar image or timeline controls for solving this task.

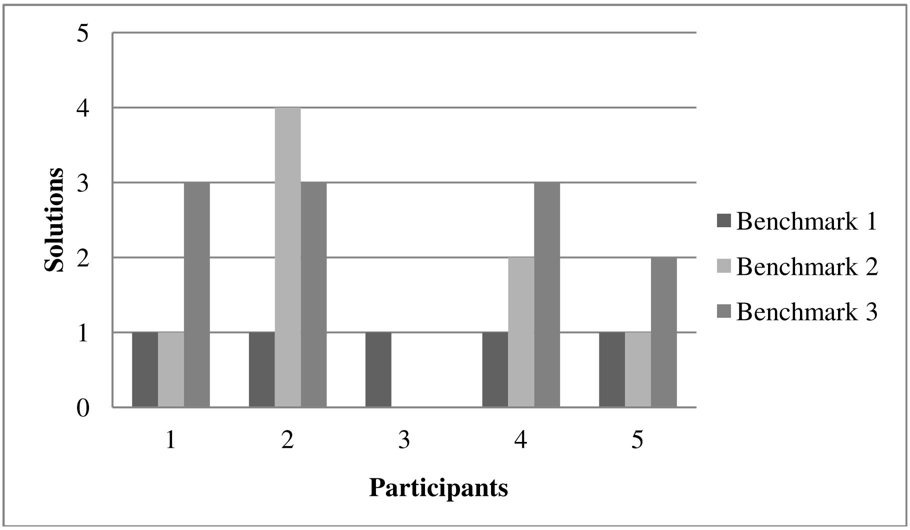

Fig. 4. Frequency of identified solutions per participant

In the second benchmark task, the participants were to find a possible cause for the delay identified in the first task. Excluding participant 3, whose dataset was omitted from the results, all participants managed to come up with a plausible solution for this task. Two of the participants identified more than one possible solution (see Fig. 4). The typical approach for solving the task was to fast forward the radar image to the point in time where the delayed flight was handled and then review the flight's progression from gate to runway. During this process, the participants often used the workflow graph in combination with the radar image to keep track of which controller was handling the flight at any given point in time. Although the participants managed to come up with plausible explanations, they emphasized that there were several possible causes of the delay, and that they were uncertain with respect to the validity of their answers. One participant postulated that he could not establish a solid explanation for the delay based solely on the information in the tool.

In the third benchmark task, the participants were to find an alternative solution for the situation causing the delay of the flight in the first task. Again excluding participant 3, all participants managed to find more than one plausible solution. The participants' approach for solving the task was to replay the flight's progression from gate to runway using the timeline controls, and to identify flights in the graph that were given departure clearance before the delayed flight, despite having a later CTOT. The participants also studied the route taken by the delayed flight, looking for points where it could sneak in before other flights. In similarity with the second task, the participants emphasized that there were several possible solutions, and that their suggestions were only speculations.

4.2 Interviews

The first part of the interview addressed the usefulness of the tool in terms of its capability to facilitate learning in ATC rooms. The response from the participants was generally positive, and all participants emphasized that the tool could be a useful aid for reviewing one's own work following a training simulation. In particular, it was postulated that the tool should be used immediately following a training simulation, and that the data in the tool had to be reviewed by the controller that had worked on the actual simulation from which the data were collected. The most useful aspect of the tool seemed to be the possibility of reviewing the actual situation by using the radar image in combination with the timeline. One participant emphasized that there was a risk that students could come to the wrong conclusions while using the tool on their own, unless they had a sufficient level of knowledge in the field:

"This [the visualization tool] makes it much more concrete than just thinking back [on the situation]. It's easier to understand [the situations] by using such a tool. It is possible to use the tool on your own, and you don't need an instructor telling you what you did wrong. This requires that you have quite a lot of knowledge [in the field] so that you don't create a source of error in yourself…"

During the interview the participants were also asked if there was any additional information and/or functionality that should be available in the tool in order to enhance learning. Several participants suggested that the radar image should be supplemented with audio playback of the communication between ATCOs, and between ATCOs and pilots. This would give the users a better understanding of how the controllers were reasoning, and an opportunity for the users to assess the clarity of their own commands. Furthermore, the participants also suggested that the radar image should display other traffic in the airport (e.g. ground vehicles, fire trucks, luggage trains), as well as more detailed information about the flights. Another highlighted issue was that the general quality of the radar image needed to be improved so that users can make out fine details, such as airplane codes and statuses. The icons on the radar screen should be color coded according to the graph so that they are easier to locate in order to see which ATCO responsibility area they are in.

When asked to reflect upon how well they managed to solve the given tasks, several participants made it clear that the tool was completely new for them, and that this affected their ability to make efficient use of it. They would need more time to get to know the tool before they could use it in an efficient manner. Several participants also emphasized that they had not seen any system similar to the visualization tool in their line of work, and that the tool would mostly be useful for students. A quote exemplifying this is included below.

"Not of this type. We do runs in a simulator where we can freeze the situation, but we cannot rewind. I think it can be used, but I don't know how actively. When you have worked for a while you have accumulated routines so that you know what you have done wrong independently of this tool. The tool will be most useful for students."

5 Discussion

All the participants in our study managed to find possible solutions to the three given benchmark tasks although they had no earlier knowledge of the traffic patterns and layout of the airport in the given scenario. In general, these findings suggest that the tool is capable of generating insight that can enable ATCO students to identify bottlenecks and potential for improvement in their work. However, due to the limited size of the study, both in terms of the number of participants and the time they spent using the tool, the results should be treated cautiously. Further research is needed to investigate the effects of this tool. An accurate measure of the tool's impact on learning quality would require longitudinal, continuous, and comparable studies, as suggested by earlier studies [9-11]. More studies are needed to explore the effects of different visualizations and to fine-tune the tool.

Although the participants were positive towards using this tool in training, they also mentioned the possibility of misuse. To reduce the possibility that users would come to the wrong conclusions while using this type of tool, they would need to have a certain level of knowledge in the field of ATC. Other studies, such as [12], have indeed shown that students with domain knowledge gained more from certain visualizations than students lacking that knowledge. This might seem obvious, but it also highlights the point that users must know what they are looking for in order to understand and make sense of the information that is presented in visualizations [13]. With regards to this, it is clear that the tool should not be used as a replacement for normal classroom lectures, but rather as a supplemental learning aid.

Several RTS facilities have the ability to freeze and replay simulation scenarios [4]. The training tool proposed in this paper extends this functionality by providing a workflow graph that helps students assess where errors have been made, as well as a timeline for navigating back and forth in the simulated scenario. The tool is also different from a regular simulator by not being restricted to a classroom setting. Students could use the tool to review simulations individually, and an instructor could give training tasks that the students could solve on their own time. The individual solutions could then be presented and compared in group discussions, allowing the instructor to provide feedback in a less obtrusive manner. As a result, one could avoid the disruptions of the internal cognitive processes of the trainees which are often observed during traditional RTS training [5]. Further, as the tool does not require the presence of an instructor or access to RTS facilities, it could also save resources in terms of cost and time.

Another suggestion is to use the tool in a collaborative way by allowing students to compare and discuss the performance of their individual solutions. By introducing such game-like elements in a student environment, one could increase the students' interest in using the tool [14-15] and in turn increase the time they spend on learning.

6 Conclusion

In this paper we have proposed and evaluated a post-simulation assessment tool that aims to improve the training of ATCOs. The results from the study indicate that visual feedback consisting of interactive workflow graphs and radar replays can be a useful means to support self and peer assessment during ATCO training. By adopting such assessment techniques in the training process, one could expect significant enhancements in the learning quality [16-17], and also help mitigate the known challenge of obtrusive instructor feedback during RTS training. Such enhancements will be increasingly important during the forthcoming modernization of the ATM industry introduced by the SESAR and NextGen programs [18-19]. The findings may also be relevant for other domains where RTS is used for training purposes, such as the fields of emergency and crisis management, defense, and health care.

Acknowledgments. 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. The authors are particularly grateful to the controllers that participated in the study, for their valuable input and great effort. We also thank Prof. Carl-Herbert Rokitansky and Thomas Gräupl from the University of Salzburg for providing us with realistic simulation data.

References

1. Battiste, H., Choi, W., Mirchi, T., Sanchez, K., Vu, K.-P.L., Chiappe, D., Strybel, T.Z.: The effects of early training with automation tools on the air traffic management strategies of student ATCOs. In: Yamamoto, S. (ed.) HIMI/HCII 2013, Part II. LNCS, vol. 8017, pp. 13–21. Springer, Heidelberg (2013)
2. Higham, T.M., Vu, K.-P.L., Miles, J., Strybel, T.Z., Battiste, V.: Training air traffic controller trust in automation within a nextgen environment. In: Yamamoto, S. (ed.) HIMI/HCII 2013, Part II. LNCS, vol. 8017, pp. 76–84. Springer, Heidelberg (2013)
3. Fisher, S.G., Kulick, I.: Air traffic controller training: A new model. In: Smolensky, M.W., Stein, E.S. (eds.) Human Factors in Air Traffic Control, 1st edn., pp. 273–298. Academic Press, United States of America (1998)
4. Hitchcock, L.: Air traffic control simulation: Capabilities. In: Smolensky, M.W., Stein, E.S. (eds.) Human Factors in Air Traffic Control, 1st edn., pp. 327–340. Academic Press, United States of America (1998)
5. Voller, L., Fowler, A.: Human Factors Longitudinal Study to Support the Improvement of Air Traffic Controller Training. In: Kirwan, B., Rodgers, M., Schäfer, D. (eds.) Human Factors Impacts in Air Traffic Management. Ashgate Publishing Limited, England (2005)
6. North, C.: Toward measuring visualization insight. IEEE Comput. Graph. Appl. 26(3), 6–9 (2006)
7. Weiss, R.S.: Learning from Strangers, 1st paperback edn. The Free Press (1995)

8. Lewis, C.: Using the "thinking-aloud" method in cognitive interface design. IBM TJ Watson Research Center (1982)
9. Saraiya, P., North, C., Lam, V., Duca, K.A.: An insight-based longitudinal study of visual analytics. IEEE Transactions on Visualization and Computer Graphics 12(6), 1511–1522 (2006)
10. Shneiderman, B., Plaisant, C.: Strategies for evaluating information visualization tools: multi-dimensional in-depth long-term case studies. In: Proceedings of the 2006 AVI Workshop on Beyond Time and Errors: Novel Evaluation Methods for Information Visualization, pp. 1–7. ACM (2006)
11. Isenberg, P., Zuk, T., Collins, C., Carpendale, S.: Grounded evaluation of information visualizations. In: Proceedings of the 2008 Workshop on Beyond Time and Errors: Novel Evaluation Methods for Information Visualization, p. 6. ACM (2008)
12. Sarayia, P., North, C., Duca, K.: An insight-based methodology for evaluating bioinformatics visualizations. IEEE Transactions on Visualization and Computer Graphics 11, 443–456 (2005)
13. Wills, G.: Visualizing time. Springer (2012)
14. Groh, F.: Gamification: State of the Art Definition and Utilization. In: Proceedings of the 4th Seminar on Research Trends in Media Informatics (RCMI 2012), pp. 39–46 (2012)
15. Zichermann, G., Cunningham, C.: Gamification by Design: Implementing Game Mechanics in Web and Mobile Apps. O'Reilly, Sebastopol (2011)
16. Race, P.: A briefing on self, peer & group assessment, vol. 9. Learning and Teaching Support Network, York (2001)
17. Boud, D., Falchikov, N.: Rethinking Assessment in Higher Education Learning for the Longer Term. Routledge (2007)
18. Sesar Joint Undertaking, `http://www.sesarju.eu/about/background`
19. NextGen, `http://www.faa.gov/news/fact_sheets/news_story.cfm?newsid=8145`