The Evaluation of Remote Tower Visual Assistance System in Preparation of Two Design Concepts

Maik Friedrich[1(✉)], Stefan Pichelmann[2], Anne Papenfuß[1], and Jörn Jakobi[1]

[1] Institute of Flight Guidance, DLR German Aerospace Center, Braunschweig, Germany
Maik.Friedrich@dlr.de
[2] Institute of Psychology, University of Bern, Bern, Switzerland

Abstract. In the last couple of years, the interest in remote tower operation systems has increased, as a concept to provide cost-efficient air traffic control service to airports. Attempts are made to stretch the concept to enable multiple remote tower control. A couple of challenges have been identified that need to be mitigated to make the concept feasible. For instance, studies revealed the disadvantage of reduced quality of the video panorama and an increased dispersion of information gathering. Design concepts were derived from the top-down and bottom-up mechanisms of human information processing to support multiple remote tower. These concepts could potentially mitigate the negative effects of a reduced quality of the video panorama and also the dispersion of information gathering. This paper focusses on the reduced quality of the video panorama and uses these design concepts to implement three visual assistance systems. These systems were tested within a laboratory experiment with 40 participants for effectiveness, efficiency, satisfaction, situation awareness, and workload. The microworld FAirControl was used to simulate an air traffic control monitoring task at a single airport. First, the visual assistance systems were compared against the baseline to assess the improvement of each system. Second, in an explorative approach, the visual assistance systems were compared to identify the advantages and disadvantages against each other. The results show that visual assistance systems can mitigate the influence of the quality-reduced video panorama. The results also indicate that usability of the visual assistance systems is crucial. Participant selective information overlay had the best overall performance.

Keywords: Remote tower operations · Visual attention · Workplace design · Interaction design · Visual assistance systems

1 Introduction

The German Aerospace Center (DLR) project RapTOr (Remote Airport Tower Operation Research, from 2005 to 2007) focused on the feasibility of remote tower operation (RTO) [1]. Remote tower operation was pushed forward by a joint venture

D. Harris (Ed.): EPCE 2017, Part I, LNAI 10275, pp. 285–300, 2017.
DOI: 10.1007/978-3-319-58472-0_22

project (Remotely Operated Towers from 2006 to 2008) of the Swedish Civil Aviation Administration (LFV) and SAAB [2]. SAAB also coordinated the EU-Project ART (Advanced Remote Tower 2007–2009) [3] focusing on single remote tower control with possible extension to multiple airports. In the last couple years, the interest in remote tower operation (RTO) systems has taken a leap. This depends on an increased acceptance of the concept in the ATM community. RTO includes all technical equipment and procedures that are necessary to perform air traffic service (ATS) from a location where provision of ATS by direct visual observation is not possible. Instead a display shows the areas of responsibility controlled by the remote tower ATS Unit. The visual presentations can be transmitted anywhere as a video panorama. Following this concept, the idea of multiple RTO was proposed. This concept defines that one air traffic controller controls traffic at multiple airports in parallel. Multiple RTO for two small airports has already been successfully validated by LFV [4]. Nevertheless, the concept raises questions on how to best support monitoring tasks of the Air Traffic Control Officer (ATCO) in a multiple RTO setting in the future.

Monitoring is a main task of ATCOs in the tower. They use it to detect deviations between the preplanned and the real traffic situations. At the moment, two major influences on visual attention occur with the transition from conventional ‘single’ tower operation to multiple RTO. First, the quality from the out of window (OTW) view to the video panorama at the remote workplace is reduced compared to the quality of the human eye. Studies [5, 6] revealed that, since the RTO concept aims at reducing costs, the video panorama has some disadvantages in comparison to the OTW view, e.g. the resolution of the image. Second, visual attention of the ATCOs under multiple RTO conditions is spread. Möhlenbrink and Papenfuß [7] addressed the feasibility of controlling two airports at the same time, with a focus on visual attention. Visual attention in search tasks is mainly guided by top-down processes, namely experience and expectations [8]. In general, their results indicated that the human cognition is able to handle the applied concept for multiple RTO. Papenfuß and Friedrich [9] discussed two design concepts as support for the ATCOs to deal with the spread of visual attention. Due to the theoretical state of the concepts, Papenfuß and Friedrich [8] were not able to verify the technical feasibility. The work presented in this paper evaluates these design concepts by deriving, three visual assistance systems (VAS) and implementing these into a microworld. The conducted lab study evaluates VAS and their influence on the ATCOs performance in an air traffic control task, focusing on the first major influence. The VAS are evaluated with non-experts in a lab environment.

2 Design Concepts

Papenfuß and Friedrich [8] derived two remote design concepts from the results of their study and a workshop with an ATCO that helped interpreting the feasibility of their concepts. The basis for both concepts is that the ATCO developed in a traditional single airport environment demonstrated a good understanding of heuristics scheduling their tasks and attention. For example, a runway on an airport restricts that either one take-off or landing at the same time, and also automatically prioritizing the attention to this single event. This heuristic is challenged taking the influence of reduced quality and spread visual attention for multiple remote towers into account. Especially the reduced quality of the video panorama increases the work load for monitoring and search task. The following design concepts aim to mitigate the influences of RTO by providing additional support that is not available at the tower today.

2.1 Attention Control and Guidance

The first design concept is based on guiding the attention of ATCOs to areas of interest. Especially, the mental change between airports could be supported via attention guidance assistance to the ATCO. This could help to focus the diversion of visual attention and support the top-down processes, guiding visual attention during a search task. The attention control and guidance (AttConG) concept increases the salience on information that is eminent important for the ATCO. The AttConG foresees an analysis of the current situation to identify what is important. Therefore it could use input and data of interaction of the ATCO. For instance, the interaction between the ATCO with the electronic stripe system provides information about the aircraft and the task that is currently in focus, e.g. landing clearance for DHL 2345.

Papenfuß and Friedrich [9] proposed to increase salience on the geographical areas at the video panorama or aircraft that have increased task relevance. For instance, in the electronic flight strip display the according strip with the landing aircraft could be highlighted and in connection the aircraft itself could be highlighted in the video panorama.

A supplementary application is the support of the monitoring task. An eye tracking system could identify the aircraft that the ATCO is looking at and then additional information for this particular aircraft could be provided. It also could support the visual search by highlighting aircraft within the video panorama that are close to the eye direction of the ATCO. An advantage of the video panorama is that it can be augmented with additional information directly. Therefore the visualization of bounding boxes around an area or aircraft is simple.

2.2 Integration of Information into External View

"Head-up Only" is the second design concept to mitigate the negative influence of visual attention. Head-up Only distributes the same information to several attention areas to minimizing the spatial distribution of information. The implementation focus is thereby the integration of all information into the video panorama. As a main part of the concept the radar and stripe information are duplicated into the video panorama. Therefore, the important information from head-down displays, like aircraft type or clearances given to the flight, are accessible whilst looking at the video panorama. This is expected to increase the visual attention spend on the video panorama. By accessing the relevant information within the video panorama, the ATCO attention can focus on the situation at hand and his peripheral attention should support him in detecting unforeseen events even though his foveal attention is on other information.

For instance, the implementation of radar screen information could be realized by bounding boxes around all aircraft within the control zone. These bounding boxes could be extended by the stripe information (e.g. call sign, speed, height) connected to them. Following this example, the danger of information cluttering [10] has to be considered with growing amount of traffic.

2.3 Evaluation of Design Concepts

AttConG and Head-up Only are different approaches to support the ATCO especially working with an video panorama performing RTO. In the interest of transferring them to VAS, Papenfuß and Friedrich [9] proposed further research for the two concepts to sharpen their influence on monitoring behavior. Considering the different aspects within Sect. 2 for AttConG (support monitoring with eye tracking versus highlighting task relevant information) and Head-up Only (increase the information distribution within the video panorama), a stepwise approach should allow insight into the potential of each concept.

The implementation of different VAS covering only one aspect from either AttConG or Head-up Only is only effective if a set of criteria helps to determine the impact of each concept. Evaluation of the VAS should also focus on the individuals' performance during the ATC task. Huber [11] described how to evaluate individuals' results for such kinds of complex problems: the difference between the optimal final state of the target variables and the achieved final state of the target variables as a measure of "performance quality". The air traffic control (ATC) task's target variables are safety and efficiency. These two variables correspond to the terms of effectiveness (error rate) and efficiency (required time) as defined by the International Organization for Standardization [12]. But these two criterias only measure the final outcome. International Organization for Standardization additionally demands user satisfaction as an usability criteria which is defined as "freedom from discomfort, and positive attitude to the use of the product" [12].

Both design concepts aim at influencing the monitoring behavior of the ATCOs. Monitoring is strongly related to the concept of situation awareness [13, 14]. According to Endsley's model of situation awareness in dynamic decision making [15], situation awareness is part of a loop of continuous interaction between the environment and a person. Hence, achieving a high level of situation awareness is demanding on cognitive resources. If a VAS makes achieving situation awareness easier, more cognitive resources are available for the ATCO for other tasks. Even if the system does not help to ease the achievement of situation awareness it might help to increase the level of understanding of the situation.

Current research in ATC [16] emphasizes the importance of a reduction of ATCO workload. For example, Metzger and Parasuraman [17] found that increased automation reduced ATCO workload. Comparable to situation awareness, mental workload is also influenced by the current situation and available information and in turn influences ATCO performance. Therefore, workload was also selected as a criteria.

Based on the presented theoretical background of the ATC task, the following five criteria seem appropriate: two performance measures (effectiveness and efficiency), satisfaction, situation awareness, and workload. As the evaluated VAS focus on an easier and faster overview of the airport situation, the research hypotheses for each criteria assume a positive impact of the implemented VAS compared to a baseline without any VAS. As an explorative research question the criteria results for each VAS are also compared to each other to identify advantages or disadvantage between the different implementations.

3 Method

3.1 Participants

In total, 40 participants took part in the experiment. Their average age was 24.73 years (SD = 3.19 years) and 25 were male and they were students. The recruitment did not allow people with a color vision deficiency because of the usage of color by all VAS. 17 participants relied on optical aids. Every participant received a compensation of 20 €.

3.2 Visual Assistance Systems

The microworld FAirControl [18] was selected as implementation platform for the developed VAS. Figure 1 shows, how FAirControl represents the areas of a real airport. Each arriving aircraft must be given six clearances (Landing, Taxi In, Taxi In Apron, Push Back, Taxi Out, and Take Off) before it departs. A participant gives clearances by clicking on one of the six buttons (Fig. 1 center below). Participants select an aircraft by selecting the flight strip. Aircraft shown in the left flight strip bay only request and execute Landing, Taxi In and Taxi In Apron clearances, whereas aircraft shown in the right flight strip bay only request and execute Push, Taxi Out and Take Off clearances. Each flight strip shows the time, the aircraft identifier (Callsign), the necessary clearances, and whether the clearance is still requested or was already deployed.

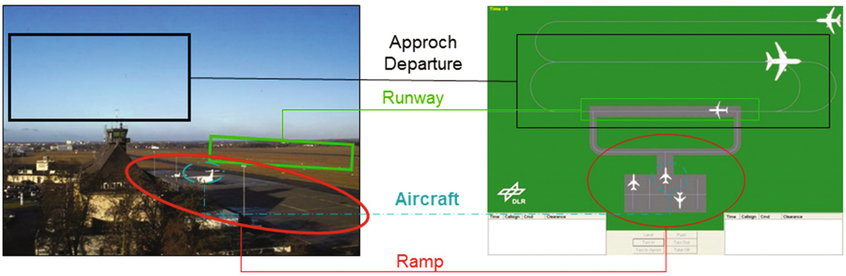

Fig. 1. Schematic representation of FAirControl (Color figure online)

FAirControl consists of three components: a human model, an airport process model, and an interaction model [10]. The human model was replaced by the participants, using a graphical user interface to interact with the airport. The airport process model controls the airport. Aircraft within FAirControl request and execute clearances based on their current status and position. The interaction model describes which information is presented to the ATCO and which clearances are given to the aircraft. Therefore, the VASs are implemented into the interaction model.

In total, three VAS were tested. Each of them aimed on improving salience with augmented information. All of them marked the position of aircraft on the screen by providing colored bounding boxes for the aircraft. Colors were chosen to provide high contrast to the surrounding area. All VAS used at least two colors: yellow and blue. A yellow frame around an aircraft symbolized a pending request of this aircraft, while a blue frame showed that no command was necessary. Throughout the experiment the same monitor was used for all participants to avoid effects due to different color representations.

The AttConG concept was implemented in two versions. The first version is called VAS-SEL and surrounds only aircraft with a bounding box that are selected via theirs flight strip. It is assumed that these aircraft are most relevant for the ATCO, e.g. aircraft that are currently requesting clearance from the participant.

The second version is called VAS-EYE and incorporated eye tracking technology. VAS-EYE works similar to VAS-SEL but uses eye tracking to select the aircraft with highest task relevance. An algorithm identified the aircraft that was currently closest to the focus of the participant's eye gaze position. A low saturated orange frame was shown around this aircraft. If the participant pressed the space bar, the aircraft was selected and a blue or yellow frame was shown around it, additionally selecting the flight strip at the same time. The implementation uses the Tobii EyeX eye tracker with a 60 Hz capturing rate.

The Head-up Only concept was implemented as VAS-ALL. It showed bounding boxes around all aircraft that were currently depicted on the screen. VAS-ALL enhanced the visibility of all aircraft, independent from their importance to the current situation.

3.3 Scenario

The scenario consisted of 27 aircraft (). At the beginning, one aircraft is always on the apron and two arrive through the approach area within 20 s. If two aircraft request landing at the same time, other aircraft arrival was delayed automatically. The remaining 24 aircraft enter the airport roughly every 20 s depending on participants' behavior. The scenario finishes once the last aircraft takes off. The reduced visual quality of the video panorama was simulated by a partially transparent layer. The contrast was reduced stronger (91%) in the approach and departure area and less for the ramp. Hence, aircraft and airport visibility was highly diminished to the degree that aircraft were still visible but harder to spot and identify. The partially transparent layer had no effect on the visibility of the bounding boxes of the VAS (Fig. 2).

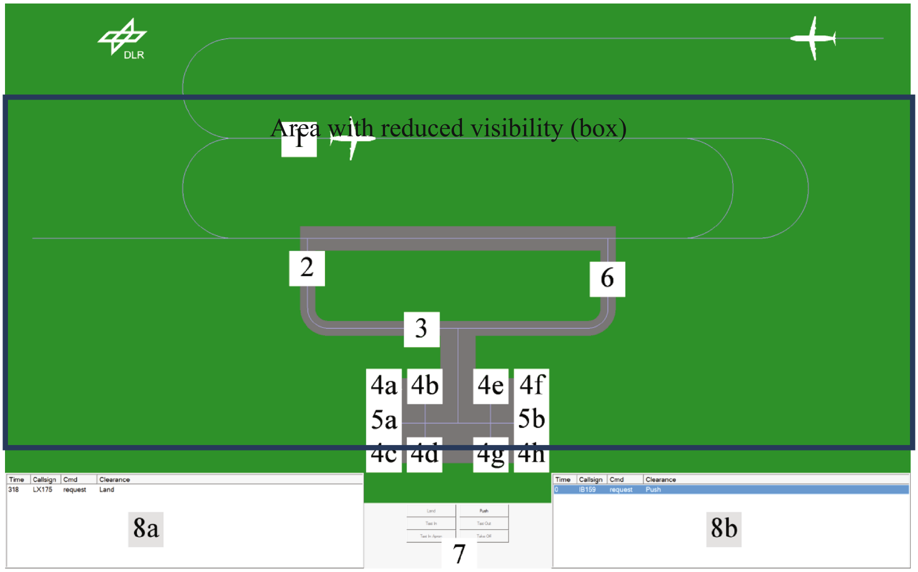

Fig. 2. User interface of FAirControl: request areas: Land (1), Taxi In (2), Taxi In Apron (3), Push (4a–h), Taxi Out (5a, b), Take Of (6). Command buttons (7) and tables containing the current aircraft (8a, b)

3.4 Design

A 1-factorial within subject design was applied to test all VAS and a trial without visual assistance (VAS-NONE) with each participant. Every trial used the same scenario. Participants were randomly assigned to one of four different sequences:

1. VAS-EYE — VAS-ALL — VAS-SEL — VAS-NONE
2. VAS-SEL — VAS-NONE — VAS-EYE — VAS-ALL
3. VAS-ALL — VAS-EYE — VAS-NONE — VAS-SEL
4. VAS-NONE — VAS-SEL — VAS-ALL — VAS-EYE

As mentioned above, the five criteria effectiveness, efficiency, satisfaction, situation awareness, and workload were used to determine the differences between the VAS: The criteria were measured either by internal calculation of FAirControl or by questionnaires. FAirControl recorded the required time to finish a trial as well as the amount and kind of mistakes. Mistakes were derived from actual ATC rules concerning the proximity between aircraft, e.g. at every time only one aircraft is allowed on the runway. The reciprocal of the mistakes quantity was used as a measure for effectiveness and the reciprocal of the required time as a measure for efficiency.

Questionnaires were applied to measure satisfaction, situation awareness, and workload for each trial. Satisfaction was measured using a single item "Please rate now how satisfied you have been with each of the assistance systems". User answered on a 7-point scale. Situation awareness was measured with the situational awareness rating technique SART [19]. The SART uses three scales to calculate a general score of situation awareness: demands on attentional resources (by the situation), supply of attentional resources (by the user), and understanding of situation. Participants answered on a 7-point scale. Workload was measured with the NASA - task load index (NASA-TLX, 20-point scale) [20].

3.5 Procedure

After arrival, participants got a brief introduction to FAirControl which included an explanation of the simulated airport and a description of how to manage the aircraft. Furthermore, they were instructed which mistakes to avoid (e.g., collision between two aircraft). The participants completed a 10 min training with one aircraft already at the apron and another two aircraft arriving shortly after at the approach area. Afterwards, participants completed the pairwise comparisons used to calculate the NASA-TLX weights [13] and answered a demographic questionnaire.

Following the design, each participant completed four runs depending on the sequence they were randomly assigned to. The task of the participants was to control the arriving and departing traffic using the six clearances. Each run took approximately 11 min to complete. After every run, participants answered the NASA-TLX and SART. Satisfaction was measured once a participant finished the final run. Furthermore, participants were asked to write down any technical problems which they experienced.

4 Results

4.1 Effectiveness

Effectiveness scores of all runs were pooled by the VASs. Table 1 shows the descriptive statistics after outliers were removed. To ease interpretation, these statistics show the reciprocal effectiveness (i.e., the total amount of mistakes per run). Hence, higher values represent less effectiveness. In the VAS-ALL and VAS-SEL conditions, participants made fewer mistakes. In contrast, participants in the VAS-EYE condition made more mistakes than participants without any VAS.

Table 1. Descriptive statistics of Effectiveness^{-1} (n = number, M = mean; SD = standard deviation)

Visual assistance system	n	M	SD
VAS-SEL	38	17.82	6.71
VAS-EYE	39	21.92	5.78
VAS-ALL	38	18.82	6.41
VAS-NONE	37	20.49	7.10

All participants with outliers in any VAS condition were removed from the sample and a repeated measures test was conducted to test for significances. The Friedman test was used instead of the repeated measures one-way ANOVA, because the Shapiro-Wilk test showed significant deviations from a normal distribution for all VAS conditions. The Friedman test showed a significant effect of VAS conditions on effectiveness (n = 34, $\chi^2(3) = 11.43$, $p = .010$). The post-hoc tests between different VAS conditions were conducted according to Dunn [21] and p-values were adjusted with the Bonferroni correction. Table 2 shows the results for the Dunn test. The VAS-EYE condition yielded significant lower effectiveness scores than the VAS-ALL and VAS-SEL conditions, corresponding to medium negative effects. The difference between VAS-ALL and VAS-SEL was close to zero. VAS-ALL and VAS-SEL yielded small positive effects compared to VAS-NONE but these effects were not significant. In contrast, VAS-EYE showed a small negative but not significant effect on participants' effectiveness scores when compared with VAS-NONE.

Table 2. Results of the Dunn test for effectiveness differences between VAS conditions (n = 34; [a]values adjusted using Bonferroni correction; $^{*}p_{adj.} < .05$)

VAS conditions	z	p	$p^{a}_{adj.}$	d
VAS-NONE - VAS-SEL	1.55	.121	.727	0.38
VAS-NONE - VAS-ALL	1.22	.222	1.000	0.30
VAS-NONE - VAS-EYE	−1.46	.145	.872	−0.36
VAS-ALL - VAS-SEL	−0.33	.742	1.000	−0.08
VAS-ALL - VAS-EYE	2.68*	.007	.045	0.69
VAS-SEL - VAS-EYE	3.01*	.003	.016	0.78

4.2 Efficiency

As for effectiveness, efficiency scores of all runs were pooled by the VASs. Table 3 shows the descriptive statistics after outliers were removed. To ease interpretation, these statistics show the reciprocal efficiency (i.e., the required amount of time to finish the trial in minutes). Hence, higher values represent less efficiency. In the VAS-ALL, VAS-SEL, and VAS-NONE conditions, participants needed roughly eleven minutes on average. In contrast, participants in the VAS-EYE condition required twelve minutes on average.

Table 3. Descriptive statistics of Efficiency^{-1} in minutes (n = number, M = mean; SD = standard deviation)

Visual assistance system	n	M	SD
VAS-SEL	39	10.11	1.19
VAS-EYE	40	11.78	1.83
VAS-ALL	38	10.02	1.09
VAS-NONE	40	10.38	1.19

Again, all participants with outliers in any VAS condition were removed from the sample and the Friedman test was used because the Shapiro-Wilk test showed significant deviations from a normal distribution. The result was significant (n = 37, $\chi^2(3) = 32.44$, $p < .001$). The post-hoc test results are shown in Table 4. The VAS-EYE condition showed significantly lower efficiency scores than any other condition, corresponding to large effect sizes. All other comparisons were not significant and their effect sizes were close to zero.

Table 4. Results of the Dunn test for effectiveness differences between VAS conditions (n = 37; [a]values adjusted using Bonferroni correction; $^{*}p_{adj.} < .05$)

VAS conditions	*z*	*p*	$p^{a}_{adj.}$	*d*
VAS-NONE - VAS-SEL	0.36	.719	1.000	0.08
VAS-NONE - VAS-ALL	0.14	.893	1.000	0.03
VAS-NONE - VAS-EYE	−4.46*	< .001	< .001	−1.21
VAS-ALL - VAS-SEL	−0.23	.822	1.000	−0.05
VAS-ALL - VAS-EYE	4.59*	< .001	< .001	1.26
VAS-SEL - VAS-EYE	4.82*	< .001	< .001	1.35

4.3 Satisfaction

Again, scores of all runs were pooled by the VASs. Table 5. shows the descriptive statistics. Participants reported highest scores for the VAS-SEL condition. In contrast, participants showed lowest satisfaction when no VAS was available.

Table 5. Descriptive statistics of Satisfaction (n = number, M = mean; SD = standard deviation)

Visual assistance system	n	M	SD
VAS-SEL	40	5.83	1.28
VAS-EYE	40	3.83	2.04
VAS-ALL	40	4.53	1.55
VAS-NONE	40	3.25	1.93

The Friedman test was used for Satisfaction because the Shapiro-Wilk test showed significant deviations from a normal distribution. The result was significant (n = 40, $\chi^2(3) = 33.60$, $p < .001$). The post-hoc test results are shown in Table 6. Participants showed significantly higher satisfaction in the VAS-SEL condition than in any other condition. Two differences in participants' satisfaction corresponded two large positive effects: VAS-SEL versus VAS-NONE as well as VAS-SEL versus VAS-EYE. Participants' satisfaction with VAS-ALL was moderately smaller than their satisfaction with VAS-SEL and moderately higher than their satisfaction with VAS-NONE but only the difference to VAS-SEL was significant. Furthermore a small positive effect for VAS-EYE was found when compared with VAS-NONE, but the effect was not significant.

Table 6. Results of the Dunn test for satisfaction differences between VAS conditions (n = 37; [a]values adjusted using Bonferroni correction; $^{*}p_{adj.} < .05$)

VAS conditions	*z*	*p*	$p^{a}_{adj.}$	*d*
VAS-NONE - VAS-SEL	5.50*	< .001	< .001	1.56
VAS-NONE - VAS-ALL	2.43	.015	.092	0.56
VAS-NONE - VAS-EYE	1.60	.109	.655	0.36
VAS-ALL - VAS-SEL	−3.07*	.002	.013	−0.73
VAS-ALL - VAS-EYE	0.82	.411	1.000	0.18
VAS-SEL - VAS-EYE	3.90*	< .001	.001	0.97

4.4 Situation Awareness

The Situation awareness scores of all runs were pooled by the VASs. Table 7 shows the descriptive statistics. Participants reported highest satisfaction scores for the VAS-SEL condition.

Table 7. Descriptive statistics of situation awareness (n = number, M = mean; SD = standard deviation)

Visual assistance system	n	M	SD
VAS-SEL	40	6.59	1.52
VAS-EYE	40	6.03	1.88
VAS-ALL	40	6.56	1.50
VAS-NONE	40	6.03	1.53

The results of the Shapiro-Wilk test showed no significant deviations from a normal distribution and therefore Mauchly's sphericity test indicated that the assumption of sphericity was not violated ($\chi^2(5) = 5.68$, p = .338). Hence, the repeated measures one-way ANOVA was conducted without corrections. It revealed a significant effect of VAS conditions on situation awareness ($F(3, 117) = 3.08$, $p = .030$). The post-hoc dependent t-tests for paired samples were conducted to analyze which of these four conditions differed significantly from other conditions. Again, p-values were adjusted using the Bonferroni correction. Results are shown in Table 8. Participants in the

Table 8. Results of the dependent t-Tests for paired samples for situation awareness scores comparing different VAS conditions (n = 40; [a]values adjusted using Bonferroni correction; $^{*}p_{adj.} < .05$)

VAS conditions	*df*	*t*	*p*	$p^{a}_{adj.}$	*d*
VAS-NONE - VAS-SEL	39	2.82*	.008	.045	0.45
VAS-NONE - VAS-ALL	39	2.12	.040	.242	0.34
VAS-NONE - VAS-EYE	39	0.41	.685	1.000	0.06
VAS-ALL - VAS-SEL	39	−0.12	.906	1.000	−0.02
VAS-ALL - VAS-EYE	39	1.76	.086	.514	0.28
VAS-SEL - VAS-EYE	39	1.86	.071	.425	0.29

VAS-NONE condition had significant lower situation awareness than participants in the VAS-SEL condition. VAS-ALL and VAS-SEL yielded small positive effects when compared against VAS-EYE and VAS-NONE but the corresponding t-tests were not significant.

4.5 Workload

Workload scores of all runs were pooled by the VASs. Table 9 shows the descriptive statistics. Average workload was lowest in the VAS-SEL condition, but absolute numbers for workload are rather medium (scale from 1 to 20).

Table 9. Descriptive statistics of workload (n = number, M = mean; SD = standard deviation)

Visual assistance system	n	M	SD
VAS-SEL	40	8.21	3.72
VAS-EYE	40	9.35	3.41
VAS-ALL	40	9.01	3.71
VAS-NONE	40	9.26	3.64

The results of the Shapiro-Wilk test showed no significant deviations from a normal distribution and therefore Mauchly's sphericity test indicated that the assumption of sphericity was not violated ($\chi^2(5) = 5.00$, $p = .416$). Hence, the repeated measures one-way ANOVA was conducted without corrections. It revealed no significant effect of VAS conditions on workload ($F(3, 117) = 2.41$, $p = .071$).

4.6 Relationships of the Criteria's

Calculation of the average effect sizes of all criteria's would be misleading due to their different importance (e.g., effectiveness is more important than satisfaction for ATC). The design of the study allows a comparison between the different VAS separately for each criterion. A major aspect of the comparison within the design is comparing each VAS against VAS-NONE, because it represents the baseline. Figure 3 presents a summary for each criterion. It shows the standardized VAS scores with Effectiveness, Efficiency and Workload scores inverted, to ease the interpretation. Therefore, values closes to the outer edge represent a better result.

Except for efficiency, VAS-SEL yielded small to high positive effects when compared with VAS-NONE. Two of these effects (satisfaction and situation awareness) were significant. For VAS-EYE, only one small positive effect was found for satisfaction when compared to VAS-NONE. It was not significant. VAS-ALL showed a medium positive effect for satisfaction and small positive effects for efficiency and situation awareness in comparison with VAS-NONE but none of them were significant.

The results comparing the different VAS against each other showed for VAL-SEL small to high positive effects for all criteria's in comparison with VAS-EYE and three of them were significant (effectiveness, efficiency, and satisfaction). Between

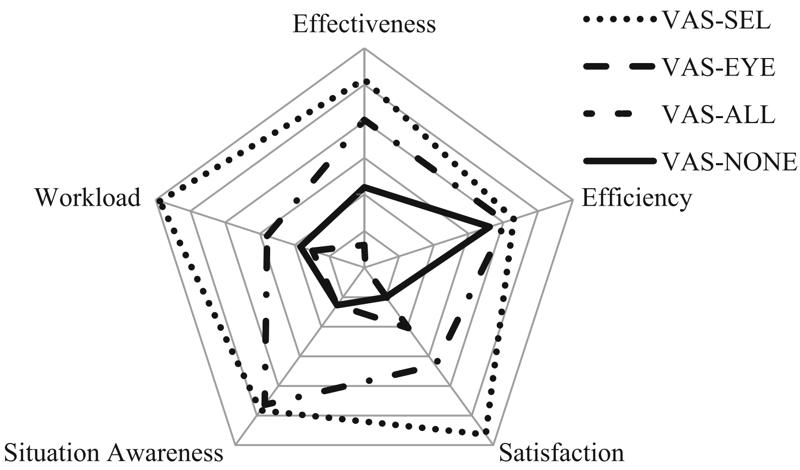

Fig. 3. Standardized VAS scores for each criterion with invented effectiveness, efficiency, and workload scores

VAS-SEL and VAS-ALL, VAS-SEL yielded a medium positive significant effect for satisfaction. Furthermore, VAS-EYE yielded one large negative significant effect (efficiency) and one small negative insignificant effect (effectiveness). Additionally, VAS-ALL yielded small to large positive effects for three criteria's (situation awareness, effectiveness, and efficiency) in comparison to VAS-EYE. Only the results for effectiveness and efficiency were significant.

5 Discussion and Conclusion

The work presented in this paper had the goal of evaluating different VAS that were derived of two different design concepts. The design concepts were defined to mitigate the problem of decreased visual quality in comparison of OTW and VR. Therefore the influence and possible support of each VAS on the ATC monitoring task was analyzed in detail. The five criteria (effectiveness, efficiency, satisfaction, situation awareness, & workload) showed only partial superiority of the VAS conditions (VAS-ALL, VAS-SEL, &VAS-EYE) against VAS-NONE.

The hypotheses of positive impact on the criteria were only significant for satisfaction and situation awareness for VAS-SEL. For effectiveness, efficiency, and workload none of the VAS were significantly better then VAS-NONE. On contrast, for efficiency VAS-EYE showed a significant negative effect compared to VAS-NONE. Nevertheless, the results for the different criteria help to rank the VAS against each other. For effectiveness VAS-EYE yielded lower results in performance than the other two VASs. The same results were found for efficiency. For satisfaction VAS-SEL showed significant medium positive effect when compared to VAS-ALL and even large positive effects when compared to VAS-EYE. Situation awareness and workload showed no significant difference between the VAS.

The VAS concentrated on an aspect of the two design concepts AttConG and "Head-Up Only". VAS-SEL and VAS-EYE use top-down mechanism thus correspond to the AttConG design concept. These concepts make use of the task relevance (VAS-SEL) and expectations (VAS-EYE) of the ATCO to superimpose additional information into the video panorama. VAS-SEL uses mouse-input via the flight strips, VAS-EYE eye gaze position on the video panorama.

VAS-SEL was significantly better than VAS-EYE concerning effectiveness, efficiency and satisfaction. The following two reasons could account for that. First, VAS-SEL used the pre-processed information of flight strips. Selection of aircraft was accurate and allowed participants to determine their exact position in the video panorama. Whilst working in VAS-EYE, participants were required to have expectations about aircraft position in the video panorama. This indicates that additional training could increase the usage of VAS-EYE. VAS-SEL also depends on additional data sources and does not cover situations where unexpected aircraft arrive. Second, the insufficient usability and accuracy of the eye-tracking system led to the unacceptable efficiency of VAS-EYE when compared to VAS-NONE. When a system or function is cumbersome to use it can account for longer handling times. The ATC task is time critically, and therefore longer handling times might also correspond with less effectiveness. On the other hand, the satisfaction with VAS-EYE was better than VAS-NONE. It is of interest to repeat the study with a more accurate eye tracking system.

VAS-ALL was derived from the "Head-Up Only" concept. Information was directly superimposed into the video panorama. Thus, positions of aircraft under control of the participants were visible. So, for instance assessment of runway status (empty or occupied) was simple. VAS-ALL did not cause information clutter, but this might be a limitation of the microworld environment.

By examination of the methodology for this paper the following three major limitations have to be considered. First, the Participants of this study were students and no ATCOs. For this implementation of the different VAS and considering the rather low level ATC task, we decided that a bigger sample size for the statistical interpretation would be more suited than a small sample size. Hence, for this specific task the performance of ATCOs and students might not differ significantly. Second, FAirControl's interface provided only visual monitoring with aircraft moving much faster than in reality. Nevertheless, FAirControl is believed an appropriate environment to test different VAS in an early stage of development. The increased speed of aircraft in FAirControl compensates for the lower workload as other duties of the ATCO are not implemented. Third, 45% of all participants mentioned problems with the eye tracker even though the calibration went without any problems and prior tests showed only negligible deviations between the traced point and the point where participants were asked to look at. People found it hard to select small and closely positioned aircraft. This might have resulted in giving commands to unintended aircraft. This could also explain the quality of the VAS-EYE results.

Therefore, the recommendation is drawn that due to the results pattern of the VAS-SEL treatment aircraft with active request should be highlighted on a realistic remote tower visual presentation.

Summarizing the conclusions for design concepts, an integrated VAS might be best to suit requirements of the real remote tower working position. The amount of information for visual assistance should also be varied to assess effects like information clutter. In this study, aircraft position was superimposed, an information that is highly relevant and mandatory for ATCOs to build up their mental traffic picture. VAS-SEL has the most positive effects for all criteria and is therefore the most promising VAS to pursuit. VAS-ALL yielded the second best results and VAS-EYE showed the lowest results. It is of interest to understand in how far VAS-SEL also mitigates diversion of attention and the detection of unforeseen events.

The research presented in this paper brings use one step closer to an integrated VAS that could support the single and multiple remote tower operation and mitigate the restrictions that come with this kind of technology. Further research should concentrate on the evaluation of VAS-SEL, VAS-ALL, and their combination in a real RTO scenario. Also to extend future research, FAirControl also implements the Integration Guideline for Dynamic Areas of Interest (IGDAI) [22] that allows a detail view on dynamics moving Areas of interest and therefore helps to identify how the visual assistance influence the eye movement distribution.

References

1. Schmidt, M., et al.: Remote airport tower operation with augmented vision video panorama HMI. In: Proceedings of 2nd International Conference Research in Air Transportation ICRAT, Belgrade, pp. 221–230 (2006)
2. Saab Security: Remotely Operated Tower - The Future of Being Present (2008)
3. van Schaik, F.J., Lindqvist, G., Rössingh, H.J.M.: Assessment of visual cues by tower controllers. In: 11th IFAC/IFIP/IFORS/IEA, Valenciennes, France, p. 19 (2010)
4. Mullan, C., et al.: OSED for Remote Provision of Air Traffic Services to Aerodromes, Including Functional Specification for Single & Multiple Aerodromes, Brussel (2012)
5. Ellis, S.R., Fürstenau, N., Mittendorf, M.: Frame rate effects on visual discrimination of landing aircraft deceleration: implications for virtual tower design and speed perception. Proc. Hum. Factors Ergon. Soc. Ann. Meet. **55**, 71–75 (2011)
6. Friedrich, M., Möhlenbrink, C.: Which data provide the best insight? A field trial for validating a remote tower operation concept. In: Tenth USA/Europe Air Traffic Management Research and Development Seminar, p. 10 (2013)
7. Möhlenbrink, C., Papenfuß, A.: ATC-monitoring when one controller operates two airports: research for remote tower centres. Proc. Hum. Factors Ergon. Soc. Ann. Meet. **55**, 76–80 (2011)
8. Wickens, C.D.: Using interference models to predict performance in a multiple-task UAV environment-2 UAVs. In: Security (2003)
9. Papenfuß, A., Friedrich, M.: Head up only – a design concept to enable multiple remote tower operations. In: 35th Digital Avionics Systems Conference (2016)
10. Knabl, P., Többen, H.: Symbology development for a 3D conformal synthetic vision helmet-mounted display for helicopter operations in degraded visual environment. In: Engineering Psychology and Cognitive Ergonomics: Understanding Human Cognition (Part I), Las Vegas, USA. Springer, Heidelberg (2013)

11. Huber, O.: Complex problem solving as multistage decision making. In: Frensch, P., Funke, J. (eds.) Complex Problem Solving: The European Perspective, pp. 151–173. Lawrence Erlbaum Associates, Hillsdale (1995)
12. International Organization for Standardization, Ergonomic requirements for office work with visual display terminals (VDTs). In: Part 11: Guidance on Usability (1998)
13. Endsley, M.R.: Measurement of situation awareness in dynamic systems. Hum. Factors: J. Hum. Factors Ergon. Soc. **37**, 65–84 (1995)
14. Wickens, C.D., Mavor, A.S., McGee, J.P.: Flight to the Future. Human Factors in Air Traffic Control. National Academy Press, Washington, D.C. (1997)
15. Endsley, M.: Towards a theory of situation awareness in dynamic systems. Hum. Factors **37** (1), 32–64 (1995)
16. Burke, K.A., Wing, D.J., Lewis, T.: Pilot subjective assessments during an investigation of separation function allocation using a human-in-the-loop simulation. In: 2013 Aviation Technology, Integration, and Operations Conference (2013)
17. Metzger, U., Parasuraman, R.: Automation in future air traffic management: effects of decision aid reliability on controller performance and mental workload. Hum. Factors: J. Hum. Factors Ergon. Soc. **47**(1), 35–49 (2005)
18. Möhlenbrink, C., Oberheid, H., Werther, B.: A model based approach to cognitive work analysis and work process design in air traffic control. In: Annual Meeting of the Human Factors and Ergonomics Society Europe Chapter, Braunschweig (2008)
19. Taylor, R.M.: Situational awareness rating technique (SART): the development of a tool for aircrew systems design. In: Situational Awareness: Limitations and Enhancement in the Aviation Environment, pp. 6/1–6/14. Advisory Group for Aerospace Research and Development, Neuilly-Sur-Seine, France (1990)
20. Hart, S.G., Staveland, L.E.: Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. Hum. Ment. Workload **1**, 139–183 (1988)
21. Dunn, O.J.: Multiple comparisons using rank sums. Technometrics **6**(3), 245–252 (1964)
22. Friedrich, M., Rußwinkel, N., Möhlenbrink, C.: A guideline for integrating dynamic areas of interests in existing set-up for capturing eye movement: looking at moving aircraft. Behav. Res. Methods (2016)