EFFECTS OF AIRPORT TOWER CONTROLLER DECISION-SUPPORT TOOL ON CONTROLLERS' HEAD-UP TIME

Miwa Hayashi
NASA Ames Research Center
Moffett Field, California, USA

Jose M. Cruz Lopez
Universidad Autónoma de Nuevo León
San Nicolás de Los Garza, Nuevo León, Mexico

Visually monitoring aircraft traffic outside the window is an important part of tower controllers' tasks. Introducing new controller decision-support tools in the tower may increase or decrease their head-up time. The present study investigates the effect of NASA's Spot and Runway Departure Advisor (SARDA) tool on controllers' head-up time. A small video camera was mounted on controllers' heads to measure their head-movement activities during simulated tower operations with or without the SARDA advisories. Simple, pixel-color-based classification algorithms were able to achieve reasonably high accuracy of head-up detection in the training video frames (91% on average). The results showed that the controllers were looking outside 33% of the time, on average. The Local Controller looked outside about 8.5% more when the SARDA advisories were provided. No other major impact of the SARDA advisories on the head-up time was found in the study.

Air Traffic Control Tower (hereafter, *tower*) controllers are responsible for the separation of aircraft and the efficiency of traffic flow on the airport surface and in its designated airspace. Visually monitoring the traffic outside the tower window is an important part of fulfilling this responsibility. NASA is developing a tower controller decision-support tool called Spot and Runway Departure Advisor (SARDA) (Jung, et al., 2010), which provides scheduling advisories to the tower controllers. The present paper examines the effects of this tool on controllers' head-up time.

SARDA computes the runway-use sequence optimized in terms of taxi delay, Traffic Management Initiative constraint conformance, and runway throughput, and presents advisory information to controllers on their Electronic Flight Strip (EFS) system. More specifically, SARDA provides the Local Controller (LC), who controls the active runway traffic, with runway-use sequence advisories, and the Ground Controller (GC), who controls all of the other traffic in the Airport Movement Area, with departure metering advisories from the ramp area. One hypothesis of SARDA's impact is that the tool alleviates controllers' planning work and allows them to complete head-down tasks more quickly. As a result, the controllers can look outside more. A counter argument is that the tool demands more head-down time to read the advisories, and consequently reduces the head-up time. Every time the controller visually monitors the traffic outside, it adds an additional layer of protection against runway incursion and other traffic conflicts. Thus, any effects on controllers' head-up time have direct implications for the safety of the airport traffic management operation. To maintain safety, the proposed tool must not decrease controllers' head-up time.

Some researchers have investigated tower controllers' head-up and/or head-down time. Bos, Schuver-van Blanken, and Huisman (2012) reported that the controllers in their study were head-down for 74% of the time when they were using paper flight progress strips and 79% when they were using the EFS system in their tower operation simulation. Bos, et al. (2012) speculated that the increase of the head-down time may have been caused by additional visual attention required for moving electronic strips, and the head-down time might eventually decrease after the controllers became accustomed to handling the EFS. Pinska and Bourgois (2005) video-recorded the controllers at the Arlanda Airport tower in Sweden, and concluded that the controllers looked outside the window for about 30-40% of the time. Hilburn (2004) used an observer and a video camera to record tower controllers' activities at two major European aerodromes, and found that the controllers spent 43-49% of the time looking outside the window. Pavet (2001) also filmed the tower controllers in the South Tower at the Paris Charles de Gaulle airport, and reported that the controllers looked outside the window for about 20% of time. Together, these past

studies suggest that, at busy airport towers, controllers generally spend about 20-50% of their time monitoring traffic outside the window.

The present study examines the effects of SARDA advisories on controllers' head-up time. Measurements were collected while the SARDA advisories were either provided to the controllers (*Advisory* runs) or not provided (*Baseline* runs). The EFS system was used in both types of runs. A small, lightweight wearable camera mounted on the side of the controller's head was used to record the controller's line of sight and later identify the controller's head-up time. The measurements were conducted during a one-week dress-rehearsal session of the main SARDA simulator evaluation session. The dress-rehearsal session was chosen for this data collection, rather than the actual main session, because of the unknown magnitude of the performance impact of mounting a camera on the controllers' heads. The dress-rehearsal session did not yield any performance results other than the head-up time measurement results reported in this paper. Readers interested in cross-referencing the current paper's results with the other performance results (e.g., workload ratings) are referred to the findings of the main evaluation session conducted in May 2012 with a similar simulation setup but with a different set of controller participants (Gupta, et al., 2013; Hayashi, et al., 2013). The results of the main session demonstrated that the use of SARDA significantly reduced the taxi delay, fuel consumption, and controller workload.

Method

Simulation

Facility. The simulation was conducted in the FutureFlight Central tower simulator facility at NASA Ames Research Center (Figure 1). The tower cab simulator provided a 360-degree computer-generated out-the-window view projected onto twelve 10-foot by 7-foot screens. The GC-position workstation consisted of a 24-inch touch screen monitor showing the EFS and a surface surveillance map monitor on the left side of the EFS monitor. The LC-position workstation included the same set of monitors as the GC workstation plus a third monitor on the right side of the EFS showing the airborne radar image.

Figure 1. FutureFlight Central Tower Simulator.

Traffic. The east-side traffic of the Dallas/Fort Worth International airport (DFW) in the south-flow configuration (runway 17R used for departures and 17C for arrivals) was simulated. Two levels of traffic, medium and heavy, were simulated, and two scenarios per each traffic level were generated for use. The medium-traffic scenarios consisted of 30 departures and 20 arrivals, and lasted for about 35 minutes. The heavy-traffic scenarios contained 40 departures and 27 arrivals, and lasted for about 45 minutes.

Experimental design. The measurements were collected in eight runs. Two retired tower controllers, who each had over 25 years of experience in DFW tower operations, participated in the simulation. The two participants were asked to perform the GC and LC tasks, alternating between these positions after each run. The four independent variables were Advisory (Advisory vs. Baseline runs), Position (GC vs. LC), Traffic Level (Medium vs. Heavy), and Participant (Controller 1 vs. 2). The runs were ordered to counterbalance potential learning or fatigue effects within and between each participant.

The dependent variables were the controllers' head-up behaviors, calculated from the video data using the methods described in the next section. Once their head-up times were computed, the head-up durations and frequencies were calculated and the effects of the SARDA advisories on these quantities, as well as the effects of the Position, Traffic Level, and Participant, were examined.

Video Data

Camera and video clips. The study used the Looxie 2 video camera (Looxie, Inc., Sunnyvale, CA) (Figure 2). This small, lightweight, wearable camera (3.25 inches long, 22 grams) was clipped onto the controller's one-ear headset,

Figure 2. Looxie 2 Video Camera.

positioned just above the ear not covered by the earphone, facing forward. Each controller wore one camera. The camera recorded a 480-pixel-by-320-pixel video clip at a rate of 15 frames per second. The recording format was MPEG-4. In this study, a 20-minute video-clip segment between the 5th minute and the 25th minute in each run, during which the traffic was sufficiently built up and the controllers were usually busy managing the traffic, was used for the analysis.

Classification. The Image Processing Toolbox of the MATLAB software package (release 2012b; MathWorks, Inc., Natick, MA) was used to extract frames from the recorded video and analyze the RGB triplet values of each pixel. After some trial and error, the following two simple classifiers shown in Table 1, or a combination of them, were found to perform the classification task sufficiently well:

Table 1. *Head-Up/Down Classifiers*

Classifier	Value	Classification
A	X_A = the sum of the RGB triplet values of all the pixels in the top half of the video frame	$X_A \geq \tau_A$: Head Up $X_A < \tau_A$: Head Down
B	X_B = the count of the black pixels in the top one-third of the video frame	$X_B \geq \tau_B$: Head Down $X_B < \tau_B$: Head Up

Classifier A utilizes the bright colors of the out-the-window image, and classifier B takes advantage of the dark color of the image inside the tower. A combination classifier AB first applies the classifier A to each frame, then the frames classified as Head Down are sent for further classification by classifier B. Which classifier works best differs for each video clip because of the camera angle. Classifier A works best if the camera was pointed slightly downward, and B works best if the camera was facing slightly upward. For some clips, combination classifier AB works the best.

The optimal threshold for each classifier (τ_A, τ_B) also varies among the clips. To obtain the best threshold for each classifier for a certain clip, and then the best classifier for the clip, the following training process was performed for each clip:

1. Extract video frames at a 3-second interval. A 20-minute clip generates 400 such frames. These serve as a training set.
2. Manually label each of the training frames as *Up*, *Down*, *Ambiguous* (i.e., it is hard to judge Up or Down), or *Other* (i.e., the controller looked at a different place). See Figure 3 for examples.
3. Run each of the classifiers, A and B, with a varying threshold on the training frames labeled Up or Down. Find the best threshold (τ_A, τ_B) that resulted in the maximum match rate between the classification result and the labels placed in Step 2. Run the combination classifier, AB, with the best thresholds for A and B.
4. Select the classifier among the three that resulted in the maximum match rate.

Figure 4 shows an example of the selection of the best threshold (Step 3). In this example of classifier A, a threshold for the sum of the RGB triplet values, $\tau_A = 1.8 \times 10^7$, yields the highest match rate—93%. A graph like Figure 4 also provides some information about the sensitivity of the match rates to the selection of a particular threshold. If the peak of the curve is relatively flat, the selection affects the match rate little. If it is sharp, the match rate is more sensitive to the selection. In the example in Figure 4, the curve exhibits a relatively flat peak. Any
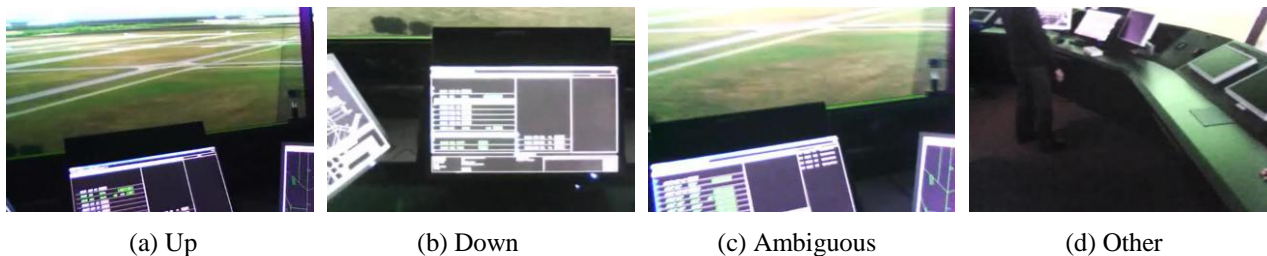

| (a) Up | (b) Down | (c) Ambiguous | (d) Other |

Figure 3. Examples of Manually Labeling the Training Frames.

threshold between 1.6×10^7 and 2.0×10^7 would result in a match rate over 90%, which would be reasonably close to the highest rate of 93%. After performing the above four steps, all video clips resulted in a match rate of 91% on average (range: 83-98%), which was sufficiently accurate for the goals of this study.

Once the classifier and the threshold for each clip were selected, the head-up versus head-down classification could be applied automatically to the remaining frames in the clip. In this study, frames at an interval of 0.25 seconds were processed, rather than all 15 frames per each second, assuming that this sampling rate would be fast enough for identifying human head movements. The frames labeled Other in Step 2 and any frames within ±1.5 seconds of these frames were labeled Other in the classification to prevent them from mixing into either the Up or Down category.
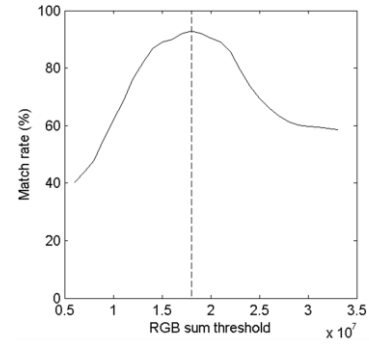

Figure 4. An Example of the Selection of the Best Classifier Threshold.

Results

Head-Up Behavior Statistics

Table 2 lists the statistics of the controllers' head-up behaviors. *Std Dv* is the standard deviation, and *N* is the number of observations used to compute the mean and the standard deviation.

- The % Head-Up Time column shows the statistics of the percentage of the time in the entire length of the 20-minute segment that the controllers were looking outside in the applicable video-clip group.
- The Head-Up Duration (sec) column shows the statistics of the duration of out-the-window viewing observed in the applicable video clips. (The distributions of these values are heavily skewed to the right.)
- The Head-Up Frequency Per Minute column shows the statistics of how frequently the controllers looked outside within each of the 20 one-minute segments in the individual video clip.

Table 2. *Head-Up Behavior Statistics*

Video-Clip Group	% Head-Up Time			Head-Up Duration (sec)			Head-Up Frequency Per Minute		
	Mean	Std Dv	N	Mean	Std Dv	N	Mean	Std Dv	N
Advisory run	34.3%	9.5%	8	2.6	3.3	1254	7.8	2.4	160
Baseline run	31.5%	11.5%	8	2.7	3.3	1130	7.1	2.7	160
Medium traffic	35.1%	11.9%	8	2.8	3.5	1187	7.4	2.6	160
Heavy traffic	30.6%	8.5%	8	2.5	3.1	1197	7.5	2.6	160
Ground	38.6%	6.3%	8	3.3	3.9	1116	7.0	2.3	160
Local	27.1%	10.5%	8	2.1	2.5	1268	7.9	2.7	160
Controller 1	28.7%	9.6%	8	2.4	3.0	1166	7.3	2.7	160
Controller 2	37.1%	9.7%	8	2.9	3.5	1218	7.6	2.4	160
All video clips	32.9%	10.3%	16	2.7	3.3	2384	7.5	2.6	320

Analysis of Variance

The head-up frequencies per one-minute segment, whose means are listed in Table 2, were subjected to a mixed-model analysis of variance (ANOVA). The main effects included were Advisory, Position, Traffic Level, and Participant. In addition, the three two-way interaction effects involving Advisory effect (i.e., Advisory × Position, Advisory × Traffic-Level, and Advisory × Participant) were included in the model. Participant effect was treated as a random effect, and all the other effects were treated as fixed effects (Lindman, 1974). The results showed that

Participant effect was statistically significant ($F_{1,304} = 19.4$, $p < 0.001$). No other effect was found statistically significant.

Analogously, the same ANOVA model was applied to the total head-up duration per one-minute segment. The analysis found a statistically significant Participant main effect and an Advisory × Position interaction effect ($F_{1,304} = 14.3$, $p < 0.001$; $F_{1,1} = 3218$, $p = 0.01$; respectively) and a marginally significant Advisory × Participant interaction effect ($F_{1,304} = 3.7$, $p = 0.057$). Figures 5 and 6 plot the means to visualize the directions and magnitudes of these effects. (Error bars are omitted because the statistical significances were already inferred by the ANOVA.) Figure 5 shows that the total head-up duration per minute increased in the LC position when the SARDA advisories were provided (13.7 sec in Baseline, 18.8 sec in Advisory), but not in the GC position (24.1 sec in Baseline, 22.3 sec in Advisory). Figure 6 indicates that, for Controller 1, the total head-up duration per minute increased when the SARDA advisories were presented (19.6 sec in Baseline, 22.9 sec in Advisory), but not for Controller 2 (18.1 sec in Baseline, 18.2 sec in Advisory).
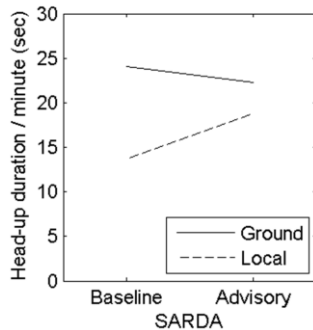

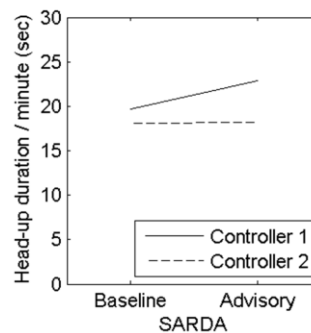*Figure 5.* Means of Total Head-Up Duration Per Minute by Advisory × Position.

Figure 6. Means of Total Head-Up Duration Per Minute by Advisory × Participant.

Discussion

Table 2 indicates that the controllers were looking outside the window for 32.9% of the time on average. The ±2σ-range was 12% to 53% (σ = 10.3%), which roughly overlaps with the previous studies' findings of 20-50% of head-up time. The % Head-Up Time was about 3% higher in the Advisory runs than the Baseline runs, but the difference may be too small to draw a reliable conclusion. Indeed, the ANOVA results did not find any statistical significance in Advisory main effect on either the head-up frequency or duration.

The ANOVA did find that the head-up durations were significantly increased in the Local position when the SARDA advisories were provided, but not in the Ground position (Figure 5). Table 2 suggests that, the % Head-Up Time is lower in the Local position than in the Ground position in general. It also suggests the LC makes more frequent and shorter glances at the out-the-window view than the GC does. These results together suggest that both GC and LC look outside just as frequently with or without the SARDA advisories, but that SARDA advisories allow the LC to monitor outside for a longer time (about 5 sec longer, or 8.5% more, per minute), probably because the advisories allow the LC's head-down tasks to be completed more quickly.

Aside from the above Advisory × Position effect, overall, the SARDA advisories had little impact on the controllers' head-up time, at least in the simulation setup used in this study. Surprisingly, the effect of Traffic Level was not found significant, either. The Heavy traffic level used in the simulation may not have been high enough, relative to the Medium traffic level used, to cause any detectable differences in the controllers' performance. The effects of the controllers' personal biases, which appeared as the significant Participant main effect and Advisory × Participant interaction effect, were found most prominent in the head-up behavior data. That means that individual controller may have very different head-up behaviors, which may also be affected differently by SARDA.

Lastly, the study employed a low-cost, commercial, off-the-shelf video camera and simple pixel-color thresholds to identify the controllers' head-up time. This method is not the most accurate way to measure controllers'

scanning activities, but it allowed for fast setup and attained reasonably high classification accuracy (91% match rates in the training set) well suited to the goals of this study. A weakness of this method is that some frames were ambiguous as to whether the controller was looking up or down. In this study, to decrease the ambiguous frames as much as possible, the controllers were asked to be standing instead of sitting during the runs. This increased the required vertical movement of the head between looking up and down, and facilitated the classification of head-up vs. head-down to a certain extent. Still, about 10% of the training frames fell into the Ambiguous category. This is a limitation of not only the current study's method, but also any head-movement-based classification method. Researchers who are interested in higher accuracy are advised to use more sophisticated tracking systems, such as eye trackers. The other shortcoming of the current approach was the inability to detect sideways head motions. When the controllers looked sideways to look outside through a different window, the classification might have resulted erroneously in Head Down because the lower part of the window frames in the distance rises in a 3D perspective, making the intensity level of the frame similar to those of the Head Down frames (Figure 4d illustrates a similar situation). Also, the current method could not distinguish if the controller was looking at the EFS or the map display. Advanced object recognition techniques could be employed to overcome these shortcomings.

Conclusion

The study found that the controllers looked outside for about 33% of the time during the simulated tower operations, a result consistent with previous studies' findings. The results showed that the SARDA advisories helped the Local Controller to look outside for about 5 seconds or 8.5% more per each minute. No other major impact of the SARDA advisories on the head-up time was found in the study. Large Participant effects observed in the head-up behavior measurements suggest that, to obtain more generalizable results, it is essential for future research to include a wider range of controller participants.

References

Bos, T. J. J., Schuver-van Blanken, M., & Huisman, H. (2012). *Towards a paperless air traffic control tower* (Report No. NLR-TP-2011-192), Amsterdam, the Netherlands: National Aerospace Laboratory NLR.

Gupta, G., Malik, W., Tobias, L, Jung, Y., Hoang, T., & Hayashi. M. (2013, submitted). Performance evaluation of individual aircraft-based advisory concept for surface management. *Tenth USA/Europe Seminar on Air Traffic Management Research and Development*, Chicago, IL.

Hayashi, M., Hoang, T., Jung, Y. C., Gupta, G., Malik, W., & Dulchinos, V. L. (2013, submitted). Usability evaluation of the Spot and Runway Departure Advisor (SARDA) concept in a Dallas/Fort Worth airport tower simulation. *Tenth USA/Europe Seminar on Air Traffic Management Research and Development*, Chicago, IL.

Hilburn, B. (2004). *Head down time in aerodrome operations: a scope study*, Technical Report, Den Haag, The Netherlands: Center for Human Performance Research.

Jung, Y. C., Hoang, T., Montoya, J., Gupta, G., Malik, W., & Tobias, L. (2010). A concept and implementation of optimized operations of airport surface traffic. *Tenth AIAA Aviation Technology, Integration, and Operations (ATIO) Conference.* Fort Worth, TX.

Lindman, H. R. (1974). *Analysis of variance in complex experimental designs*. San Francisco, CA: W. H. Freeman and Company.

Pavet, D. (2001). Use of paper strips by tower air traffic controllers and promises offered by new design techniques on user interface. *Fourth USA/Europe Air Traffic Management Research and Development Seminar*, Santa Fe, NM.

Pinska, E., & Bourgois, M. (2005). Behavioural analysis of tower controller activity, *EUROCONTROL Experimental Centre, Activity Report 2005*, Brétigny-sur-Orge, France: EUROCONTROL.