A Human-Autonomy Teaming Approach for a Flight-Following Task

Summer L. Brandt[1], Joel Lachter[2], Ricky Russell[1], R. Jay Shively[2]

[1]San José State University, NASA Ames Research Center, Moffett Field, CA 94035
[2]NASA Ames Research Center, Moffett Field, CA 94035
{Summer.L.Brandt, Joel.Lachter, Ricky.Russell, Robert.J.Shively}@nasa.gov

Abstract. Human involvement with increasingly autonomous systems must adjust to allow for a more dynamic relationship involving cooperation and teamwork. As part of an ongoing project to develop a framework for human-autonomy teaming (HAT) in aviation, a study was conducted to evaluate proposed tenets of HAT. Participants performed a flight-following task at a ground station both with and without HAT features enabled. Overall, participants preferred the ground station with HAT features enabled over the station without the HAT features. Participants reported that the HAT displays and automation were preferred for keeping up with operationally important issues. Additionally, participants reported that the HAT displays and automation provided enough situation awareness to complete the task, reduced the necessary workload and were efficient. Overall, there was general agreement that HAT features supported teaming with the automation. These results will be used to refine and expand our proposed framework for human-autonomy teaming.

Keywords: Human-Autonomy Teaming · Ground Station · Recommender System · ACFP

1 Introduction

Managing aircraft is becoming more complex with increasingly sophisticated automation responsible for more flight tasks. With this increased complexity, it is becoming more difficult for operators to understand what the automation is doing and why. Human involvement with increasingly autonomous systems must adjust to allow for a more dynamic relationship involving cooperation and teamwork.

As part of an ongoing project to develop a framework for human-autonomy teaming (HAT) in aviation [1], a part-task study was conducted to demonstrate, evaluate and refine proposed tenets of HAT. The HAT features were derived from three tenets and were built into an automated recommender system on a ground station. These HAT tenets include:

- *Bi-Directional Communication*: For automation to act as a teammate, there needs to be bi-directional communication about mission goals and rationale. This requires a clear communication channel with a shared, understandable language [2].

- *Transparency*: Automated systems often do not facilitate understanding or tracking of a system [3]. Providing the automation's rationale for selecting particular actions helps the human understand what the automation is doing and why. Again, to be truly transparent, communication should use a shared language that matches the operators' mental model.
- *Operator Directed Interface*: While increased automation can help with manual performance and workload, recovering from automation failure is often worse [4,5]. A dynamic allocation of tasks based on operator direction and context allows a much more agile, flexible system and a greater opportunity to keep the operator in the loop.

This study focused primarily on interactions with one piece of automation, the Autonomous Constrained Flight Planner (ACFP). The ACFP is an automated recommender system designed to support rapid diversion decisions for commercial pilots in off-nominal situations [6]. The ACFP was designed to generate a list of diversion options in a ranked order. It compiles information from several sources such as ATIS broadcasts, METAR weather reports, GPS location and terrain, aircraft condition, and airport/runway characteristics. Evaluations are made for various factors (e.g., risk associated with the enroute, approach, and landing phases of flight, fuel usage, weather, terrain, distance, facilities). These evaluations are then aggregated to produce an overall score. The initial implementation of the ACFP provided little transparency regarding the evaluation and weighting of these factors [7]. Much effort has gone into enhancing this tool not only in capability but also in transparency [8,9,10]. For the purpose of this study, participants used the ACFP at a ground station to reroute aircraft in situations such as inclement weather, system failures and medical emergencies. Participants performed this task both with HAT features enabled and without and provided feedback.

2 Method

2.1 Participants

Four dispatchers (median dispatch experience was 11 years) and two pilots (both active duty with over 10,000 hours flown as a line pilot) participated in this simulation.

2.2 Simulation Environment

Our simulation ground station has been developed through a series of human-in-the-loop simulations to examine issues associated with collaboration between an onboard pilot and ground support from a dispatcher or ground pilot [11,12,13]. Each successive simulation advanced the ground station and evaluated a more fully evolved operational concept. Our current framework envisions a role for ground support in monitoring and assisting aircraft in an advanced flight following mode, which requires increasingly sophisticated automation and an opportunity to enhance collaboration between the operator and the automation.

The following sections describe the components of the ground station for multi-aircraft monitoring and support. The HAT features included for this simulation are called out separately.

Aircraft Control List. The center of the station hosts an Aircraft Control List (ACL), the primary tool for managing multiple aircraft and switching the focus between aircraft (see Fig. 1A). The ACL provides information crucial for situation awareness such as callsigns, departure and destination city pairs, estimated time of arrivals, flight plans, souls on aboard, and pilot details.

This version of the ground station was designed to monitor, with the help of automation, a large number of aircraft (up to 30). Automated alerts are provided in three priority levels and two burdening levels. Priority refers to whether the situation must be addressed immediately (high shown in red), can wait before being addressed (medium shown in amber), or is merely advisory (low shown in green), corresponding respectively to flight deck warning, caution, and advisory alerts. Burdening refers to whether the operator is expected to take an action or if another agent or onboard pilot is responsible. In the current simulation, alerts were issued for failure to adhere to a clearance, failure to stay on path, environmental threats (weather on flight path or at the destination, and airport closures), system issues, and failure of the pilot to acknowledge a flight deck alert.

Fig. 1. Simulation ground station. A: Aircraft Control List (ACL), augmented with timeline, alerting information and HAT features. B: Traffic Situation Display (TSD). C: Flight controls and displays for the selected aircraft in read-only mode. D: CONUS map and charts.

Creating an Operator Directed Interface with Plays. We have adopted the playbook approach to set system goals and manage roles and responsibilities between the operator and the automation [14]. Our first implementation provides 13 different plays the operator could call to address various scripted events. When the operator selects a play, the ACFP is triggered with preset weights and the corresponding checklist appears on the display identifying operator tasks in white and automation tasks in blue (see Fig. 2).

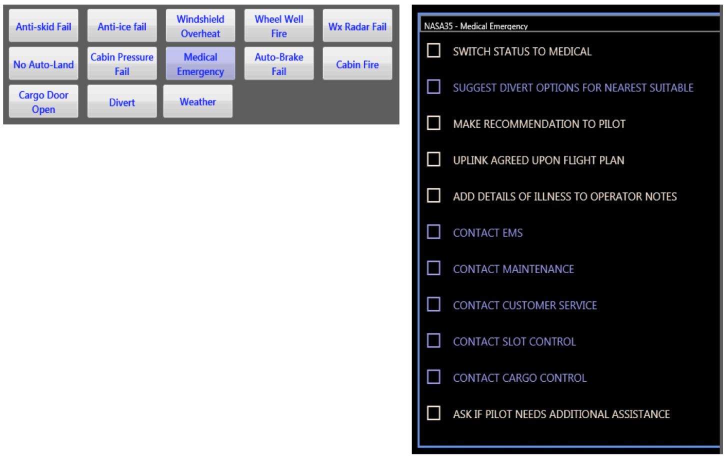

Fig. 2. Operators could call plays for various events and in the HAT condition were provided a checklist of roles and responsibilities. This is an example of a Medical Emergency play.

Building in Transparency. The ACFP takes into account more than just risk; it looks at fuel, distance, services available, etc. It also has the capability of weighting these factors differently based on the situation. We increased transparency by explicitly showing the factors and weights of the recommended divert airports when the ACFP is enabled (see Fig. 3). Additionally, we translated the scores for the ACFP factors to more meaningful numbers and descriptors for the operator (e.g., presenting nautical miles (nm) instead of a score). In the example above, a Medical Emergency play was called which resulted in the distance to medical facilities (Medical row) and time to destination (ETA row) given more weight than other factors. As a result, Cheyenne (KCYS) was the top recommendation showing a trauma care facility 3 nm from the airport. Although Denver (KDEN) was closer, the trauma care facility is further from the airport.

Building in Bi-Directional Communication. We preset weights for each play and presented the weight settings (top of Fig. 3). The operator is able to negotiate with the system by changing these weights to better fit the situation. The operator can adjust the weights and see how the divert recommendation is affected. Again, using the example in Fig. 3, if the operator decided that distance to the airport or estimated time of arrival were a higher priority than available medical facilities, the operator could adjust the ACFP weights and re-run the query.

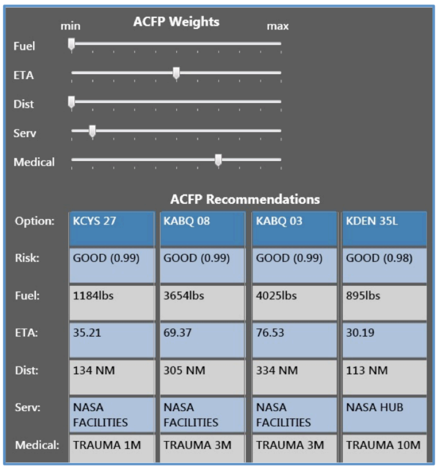

Fig. 3. In the HAT condition, operators were provided ACFP factors (on bottom) and weights (on top) to increase transparency and bi-directional communication. This is an example of a Medical Emergency play.

Traffic Situation Display. The Traffic Situation Display (TSD) is a 3D map display of company aircraft (see Fig. 1B). Information such as flight plans, trajectories and data tags are selectable. Color-coding from the ACL is maintained allowing the operator to, at a glance, identify the priority level of each aircraft. NextRad weather and turbulence boxes are graphically displayed.

Building in Transparency. The ACFP was augmented to display ATIS at the destination airport as well as indicate which of a number of risk factors are present in any potential divert location (see Fig. 4) [9,10]. Operators could also request such ratings for airports that are not recommended by the tool. The recommended route to a given airport is displayed in grey on the TSD.

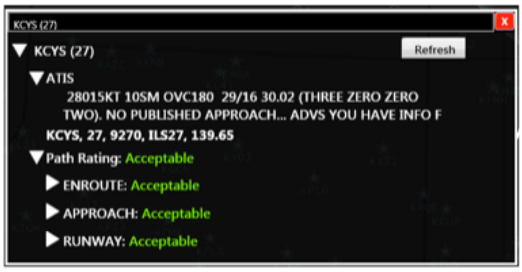

Fig. 4. When using the ACFP, regardless of condition, a transparency window appeared on the TSD for the recommended airport. An ATIS report, runway information, path rating and reasoning statements were also included.

Additional Displays. The left-side display contains aircraft flight controls and instrumentation for the selected aircraft (see Fig. 1C). A graphical display of the Flight Management System (FMS) through a GUI Control Display Unit (CDU) was carried over from previous builds of the ground station in part because dispatchers previously reported that seeing the flight controls improved their situation awareness. The controls are view-only and do not allow for any manipulation of the aircraft. Instrumentation and information displays include a Primary Flight Display (PFD) and some Engine Indication and Crew Alerting System (EICAS) functionality. The right-side display contains a CONUS map with an overlay of company aircraft and weather (see Fig. 1D). Airport charts are provided below the map.

Voice Interaction. Operators are able to perform some functions by voice, such as selecting specific aircraft and invoking the ACFP. Alerts and certain system changes (e.g., aircraft landing) are also announced vocally. All stations (simulated ground station and flight decks) are equipped with push-to-talk communication.

Airspace. Participants managed the ground station while confederates supported the simulation by piloting and initiating various off-nominal situations. Aircraft approximated 737-800s and were located in the western half of the U.S. The Multi Aircraft Control Systems [15] was used to simulate the airspace and aircraft.

2.3 Experimental Design

The experimental design consisted of a single fixed factor, HAT, and a random factor, Subject. There were two levels of HAT: HAT (ground station with HAT features enabled) and No HAT (ground station without HAT features). Participants performed the flight-following task once in each condition with the order of trials counterbalanced across participants. We collected behavioral and subjective data. Subjective data results are presented below; behavioral data are reported separately [16].

One participant was tested per day. Each participant received approximately 3.5 hours of training before running two 50-minute experimental scenarios. Questionnaires were administered post-scenario and post-simulation. A debrief session was conducted post-simulation to gather additional feedback.

Participants were provided with a concept of operation where automation and ground personnel provide “another set of eyes” monitoring aircraft. The role of ground in this flight following task was to support aircraft in high workload and off-nominal situations. Our primary interest was in participant feedback of the HAT features.

Scenarios were developed to test HAT features in making diversion decisions under different weather conditions and emergency landing situations. Each scenario required participant ground operators to make approximately six diversions using the ACFP. Confederates supported the simulation by piloting and initiating various scripted off-nominal situations for the ground station operator. Example reasons for diversions include deteriorating weather conditions at the destination airport, a mechanical emergency such as an aft cargo door open, or a critically ill passenger needing immediate medical care.

At the start of a scenario, the operators had up to 30 aircraft to flight-follow after takeoff until landing. Shortly into the scenario, the ground station began alerting the operator to various situations. An event started either with an alert on the ground station, elevating the aircraft priority and queuing the operator to contact the pilot, or with a radio call from the pilot. In either case, if it was determined that the aircraft needed to divert to a new destination, the operator would invoke the ACFP by selecting the appropriate play. In the HAT condition, once the play was selected, a checklist of procedures appeared with the automation responsible for a certain set of identified tasks. In the No HAT condition, operators had a paper checklist available for procedure items. In both conditions, the ACFP provided multiple recommendations in rank order and the transparency window was displayed on the TSD for the selected airport (see Fig. 4). The ACFP would select the highest rated airport based on the event and related factors, though the operator could explore additional airports and view both the suggested route and transparency window. In the HAT condition, the ACFP factors and weights were displayed on the ACL providing additional transparency and allowing for manipulation of factor weights. When there was consensus on the new airport, the operator would datalink the route to the pilot who would, according to the concept of operations, contact air traffic control for approval. Operators could then determine to what extent they needed to follow that aircraft and either leave the aircraft priority elevated (amber or red) or reduce the aircraft priority (green).

3 Results

3.1 Post-Scenario Comparisons

After each scenario, participants completed a post-scenario questionnaire rating agreement with (1 = strongly disagree, 5 = neither agree nor disagree, 9 = strongly agree) and confidence in (1 = no confidence, 5 = confident, 9 = extremely confident) several aspects of the ACFP. Paired-samples t-tests were run between the HAT and No HAT conditions. Given the small sample size and range of responses, few significant differences were found.

Little differences were found in participant ratings of reliance on the ACFP. In both the HAT (M = 5.67, SD = 1.37) and No HAT (M = 5.17, SD = 2.40) conditions, pilots neither agreed nor disagreed that they would rely on the ACFP recommendation without hesitation, p = .41. However, if faced with a very hard and time constrained task in the future, participants agreed they would rely on the ACFP in both the HAT (M = 8.00, SD = 1.55) and No HAT (M = 7.33, SD = 1.63) conditions, p = .24.

Although mean scores were not significantly different, four of the six participants agreed (ratings 7-9) that overall, the diversion decisions recommended by the ACFP were acceptable in the HAT condition (M = 6.67, SD = 2.16), compared to two of the six participants (ratings 8-9) in the No HAT condition (M = 5.33, SD = 2.58), p = .12. There was a significant difference in confidence where participants reported greater confidence that the diversions they chose were appropriate in the HAT condition (M = 7.83, SD = 1.47, all participants rating 6 and above) compared to the No HAT condition (M = 6.33, SD = 2.07, four participants rating 6 and above), $t(5) = 4.39$, p = .01. Five of the six participants reported confidence that the diversions recommended by the ACFP were appropriate in the HAT condition (M = 6.67, SD = 2.25) compared to three of the six participants in the No HAT condition (M = 5.33, SD = 2.42; p = .17).

3.2 Simulation Ratings

In addition to the post-scenario questionnaires, a final, post-simulation, questionnaire was administered after both trials were completed.

Display Preference. Participants were asked to rate their preferred displays and automation on a 1 = No HAT to 9 = HAT scale. Participants unanimously preferred the HAT displays. Specifically, HAT displays were preferred with regard to:

- keeping up with operationally important issues (*M* = 8.67, *SD* = 0.52);
- ensuring the necessary situation awareness for the task (*M* = 8.67, *SD* = 0.52);
- integrating information from a variety of sources (*M* = 8.67, *SD* = 0.52);
- reducing workload necessary for the task (*M* = 8.33, SD = 0.82); and
- efficiency (*M* = 8.33, *SD* = 0.82).

Participants were in agreement that overall they preferred interacting with the automation in the HAT condition (*M* = 8.50, *SD* = 0.55).

ACFP Recommendations. In additional to ratings of agreement, participants rated several items in terms of usefulness (1 = not useful, 5 = somewhat useful, 9 = very useful). Post-simulation, the ACFP was rated a useful tool (*M* = 7.33, *SD* = 1.37). Four of the six participants agreed that the recommendations were generally consistent with what s/he would have recommended (*M* = 5.83, *SD* = 2.56). The ACFP seems to be particularly helpful during emergency situations, as explained by one participant, *"Everything is easy and accessible in emergency situations. No need to consult many other programs to get various info."*

HAT Features. Items specific to the HAT tenets were rated both post-scenario and post-simulation.

Supporting Bi-Directional Communication. Participants agreed that the ACFP weights improved the automation's ability to handle unusual situations (*M* = 7.83, *SD* = 1.60) and were useful in making divert decisions (*M* = 8.33, *SD* = 0.82). Participants liked having the weights (*M* = 8.33, *SD* = 1.21) and one participant commented that, *"[the display] gave me the ability to see why, gave me control to change weights in variable(s)."*

Building in Transparency. Participants agreed that the ACFP table was helpful in making divert decisions (*M* = 7.67, *SD* = 1.51) and they liked having the table (*M* = 8.33, *SD* = 1.03). One participant commented that, "*This [table] is wonderful… You would not find a dispatcher who would just be comfortable with making a decision without knowing why*."

Creating an Operator Directed Interface with Plays. Participants liked having the electronic checklist for each play (*M* = 8.67, *SD* = 0.52) and did not prefer the paper to the electronic checklist (*M* = 2.67, *SD* = 1.97). One participant claimed that, *"The electronic list was easier because it was right there on the screen and it eliminated a* *couple of the steps."* Another participant was hesitant to rely solely on the electronic checklist and explained that s/he, *"found it necessary to have both on hand."*

Overall. Participants all agreed that they would like to have a tool like the ACFP with HAT features to use with real flights (M = 8.00, SD = 0.89).

Additional displays. Participants reported consulting ATIS information while making a diversion decision in both the No HAT and HAT conditions, and half of the participants consulted airport charts while making a diversion decision (the other half did not consult the charts because they were already familiar with the diversion airport).

Voice Input. Participants did not prefer using voice input over mouse input (M = 4.17, SD = 3.13), with one participant commenting that the preference for mouse was *"due to reliability."* Participants slightly agreed that voice input would be more useful if it worked better (M = 6.33, SD = 2.07).

Voice Annunciations. Participants found the voice annunciations for alerts useful (M = 7.17, SD = 2.56) with one participant commenting that it, *"gave another dimension to alerting."* However, another participant cautioned that, "*the voice annunciations became somewhat mundane when announcing the new routes when planning for a diversion."* Participants found the voice annunciations for arrivals somewhat useful (M = 5.50, SD = 2.66), with one participant commenting that they, *"helped with flight following."* Overall, participants rated voice annunciations for aircraft for the ACFP as useful (M = 7.00, SD = 1.67).

4 Discussion and Conclusion

HAT is a promising solution for increasingly complex systems. Guidelines for improving HAT are growing, e.g., [4], and this project is our early effort at designing for human-autonomy teaming in an actual system. Our approach emphasizes three tenets: bi-directional communication between the human and automation, transparency for automation, and an operator directed interface.

We acknowledge this was an initial study with a small sample size, as our primary purpose was to demonstrate HAT features and gather feedback for further refinement. Overall, participant feedback was positive, supporting our implementation of HAT features. Participants liked having a recommender system with factors and weights and expressed interest in having similar automation for real flights. They valued the integration of the displays, commenting that this level of integration is not currently available. Participants found the electronic checklist useful as were a number of the alerts.

Suggestions for improvement were also provided, which we are working toward. In regards to the ACFP, participants expressed reluctance to allow the ACFP to make diversion decisions for them at this point. As one person put it, it still *"needs some TLC to be trusted."* We are continuing our work on how transparency affects trust [9,10] and planning to incorporate the results into the ACFP. Furthermore, in our current implementation, plays included simple checklists. We are working towards making these more flexible with varying levels of automation and branch points. While the concept of voice input and annunciations was received well, our implementation is in its infancy. Voice input was unreliable, working well for some participants but not all, and voice annunciations lacked etiquette, speaking over the operator and pilot. Improvements are being made to increase the vocabulary and grammar and better manage the output to not interfere with operator tasks.

As a next step, we have ported some of these ground station tools to a tablet for use on a simulated flight deck and will again evaluate the tools with and without HAT features. Our goal is to develop a framework for HAT, consisting of tenets and guidelines for implementing them. We eventually hope to create software libraries that make this implementation easier.

Acknowledgments. We would like to acknowledge NASA's Safe and Autonomous System Operations Project, which funded this research.

References

1. Shively, R.J., Lachter, J., Brandt, S.L., Matessa, M., Battiste, V., Johnson, W.: Why Human-Autonomy Teaming? In: Proceedings of the 8th International Conference on Applied Human Factors and Ergonomics, this edition (2017)
2. Goldman, C.V., Degani, A.A.: Team-Oriented Framework for Human-Automation Interaction: Implication for the Design of an Advanced Cruise Control System. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, pp. 2354--2358. SAGE Publications (2012)
3. Lyons, J.B.: Being Transparent about Transparency: A Model for Human Robot Interaction. In: AIAA Spring Symposium Series (2013)
4. Endsley, M.R.: From Here to Autonomy: Lessons Learned from Human-Automation Research. Hum. Factors. 59, 5--27 (2017)
5. Onnasch, L., Wickens, C.D., Li, H., Manzey, D.: Human Performance Consequences of Stages and Levels of Automation: An Integrated Meta-Analysis. Hum. Factors. 56, 476--488 (2013)
6. Meuleau, N., Plaunt, C., Smith, D.E., Smith T.B.: An Emergency Landing Planner for Damaged Aircraft. In: Proceedings of the Twenty-First Innovative Application of Artificial Intelligence Conference, pp. 114--121 (2009)
7. Dao, A-Q., Koltai, K., Cals, S.D., Brandt, S.L., Lachter, J., Matessa, M., Smith, D., Battiste, V., Johnson, W.W.: Evaluation of a Recommender System for Single-Pilot Operations. Procedia Manufacturing. 3, 3070--3077 (2015)
8. Lyons, J.B., Koltai, K.S., Ho, N.T., Johnson, W.W., Smith, D.E., Shively, R.J.: Engineering Trust in Complex Automated Systems. Ergonomics in Design. 24, 13--17 (2016)
9. Lyons, J.B., Saddler, G.G., Koltai, K., Battiste, H., Ho, N.T., Hoffmann, L.C., Smith, D., Johnson, W., Shively, R.: Shaping Trust through Transparent Design: Theoretical and Experimental Guidelines. Advances in Human Factors in Robotics and Unmanned System. 499, 127--136 (2017)
10. Sadler, G., Battiste, H., Ho, N., Hoffmann, L., Johnson, W., Shively, R., Lyons, J., Smith, D.: Effects of Transparency on Pilot Trust and Agreement in the Autonomous Constrained Flight Planner. In: Digital Avionics Systems Conference (DASC) IEEE/AIAA 35th, pp. 1--9. IEEE (2016)

11. Lachter, J., Brandt, S.L., Battiste, V., Ligda, S.V., Matessa, M., Johnson, W.W.: Toward Single Pilot Operations: Developing a Ground Station. In: Proceedings of the International Conference on Human-Computer Interaction in Aerospace, Santa Clara, CA (2014)
12. Brandt, S.L., Lachter, J., Battiste, V., Johnson, W.W.: Pilot Situation Awareness and its Implications for Single Pilot Operations: Analysis of a Human-in-the-Loop Study. Procedia Manufacturing. 3, 3017--3024 (2015)
13. Lachter, J., Brandt, S.L., Battiste, V., Matessa, M., Johnson, W.W.: Enhancing Ground Support: Lessons from Work on Reduced Crew Operations. Cognition Technology & Work (in press)
14. Miller, C.A., Parasuraman, R.: Designing for Flexible Interaction Between Humans and Automation: Delegation Interfaces for Supervisory Control. Hum. Factors, 49, 57--75 (2007)
15. Prevot, T.: Exploring the Many Perspectives of Distributed Air Traffic Management: The Multi Aircraft Control System: MACS. In: International Conference on Human-Computer Interaction in Aeronautics, HCI-Aero, pp. 23--25 (2002)
16. Strybel, T., Keeler, J., Mattoon, N., Alvarez, A., Barakezyan, V., Barraza, E., Park, J., Vu, K.-P., Battiste, V.: Measuring the Effectiveness of Human Automation Teaming in Reduced Crew Operations. In: Proceedings of the 8th International Conference on Applied Human Factors and Ergonomics, this edition (2017)