The Sixth Sense of an Air Traffic Controller

Eva Eggeling*, Volker Settgast*, Nelson Silva*,‡, Michael Poiger†, Theodor Zeh† and Dieter Fellner‡,§
*Fraunhofer Austria Research GmbH, Graz, Austria
†Frequentis AG, Vienna, Austria
‡Institute of ComputerGraphics and KnowledgeVisualization (CGV), Graz University of Technology, Austria
§GRIS, TU Darmstadt & Fraunhofer IGD, Darmstadt, Germany

Fig. 1. Hamburg Airport

***Abstract*—The project Sixth Sense postulates that the users body language is different at "good" and "bad" decisions. Therefore, in Sixth Sense we are looking for patterns or hidden data signs that allow us to detect moments of bad and good decisions that could be incorporated in an automated system in order to detect and eventually predict the next actions of a user. In our case the user is an Air Traffic Controller (ATCO). Specifically, we intend to analyse the correlation between the change in the ATCO's behaviour - expressed through his body language - and the quality of his/her decision. For that, an experiment was set up to collect, explore and analyse data about the user behaviour. The results of our work may be used for early warnings for upcoming"bad" situations or decision aids for ATCOs.**

I. Introduction

User errors are one of the most critical errors in a safety critical environment. In the last years Human Factors has concentrated on eliminating those - often with the result of reducing flexibility and productivity of the users. And mostly also by blocking innovation. Modern technologies like eye tracking, voice recognition and gesture control were rarely taken into account. Even less a combination of those. The project Sixth Sense started with the idea to improve the fault tolerance of user interfaces by using multiple interaction sensors and comparing the result. That rough project idea cumulated in the research question: can the quality of the decisions a user is making be detected by using the whole body language of a user for communicating with a machine and thus be improved?
In our case the user is an Air Traffic Controller (ATCO) and the environment an Air Traffic Control Tower. We analyse the correlation between changes in the user behaviour - expressed through her/his body language - and the quality of her/his decision.
The result of our work might lead to a Sixth Sense module which supports users of safety critical systems. This module built into the controllers working position will give an early warning if bad situations are about to occur and guide the users to safe decisions. Thus the Sixth Sense module will form the safety net for user errors.

But before thinking about predictions we need to collect data that describe the human body language, and detect these patterns by smart analysis. In a later stage, we intend to predict these patterns. For the pattern detection we first break down the long-term challenge into smaller problems and simpler research questions, that can be answered by data analysis.
In our paper, we present the process from the design of the experiment, the sensor selection and data collection to the identification of the most promising combinations of sensor data and their visualization in order to give hints for the detection of hidden patterns for good or bad decisions.

II. The Challenge

Thinking ahead, the performance of the Sixth Sense module should be tested by answering the following hypothesis:
The module is able to detect situations in which the operator tends to make bad decisions by analysing user-input and user-tracking data. Furthermore, the module is able to identify good and bad workflow patterns.

Before behaviour patterns can be predicted by an algorithm we need to find them. Therefore, we first try to answer the following question instead:

> What kind of pattern was detected in the data and might be useful for the development of a prediction module?

To find an answer to this key question, we have formulated a short list of 15 concrete research questions (see [1]) that can be handled and discussed by analysing the data. These research questions guide our analysis, visualization and exploration of data when searching for good predictors of the users behaviours. And finally, the answers to these questions will lead us towards the aim of our project.

III. Related Work

Data stream oriented applications are used across different fields and play an important role in sensor integration projects. Several strategies can be applied to analyse streaming information. One of these strategies is the use of Complex Event Processing Systems (CEP) [2]. CEP combines several existent events to generate a new composite or derive new events. These events contain new correlated information for studying the underlying processes. Furthermore, CEP offers the opportunity for an improved loose coupling of software components [3].

The recognition of patterns is an important part in analysing decision-making mechanisms. It can be achieved by applying advanced outlier detection and machine learning techniques [4]. Recent research clarifies why the human brain makes mistakes and how the decision-making mechanisms work in reality. The decision-making tasks are now linked with sensory evidence delivered in randomly timed pulses where noise is playing a key role as a source of variability and errors [5].

A variety of controller errors involves perception, memory, decision-making, communication and team resource management. The classification of errors is essential to record data for the detection of trends in incident occurrence. Identifying situations where systems can fail or identifying risky strategies taken by users, makes error analysis a key component in safety management [6].

The topic of anomaly (outliers) detection and time series visualization are also important aspects for the analysis of time based data originated from users behaviour and from sensors data streams [7], [8]. Moreover, the projection of multidimensional data to a lower-dimensional visual display is a common approach to identify and visualize patterns in data [9].

IV. The Experiment

As a description of the human body language, communicating with the machine interface, we use sensor data. The following sensors were taken into account for reading the body language:

- Kinect for body movement
- Eye tracking for gaze detection
- Speech recognition
- Mouse cursor position
- Room temperature
- Heartbeat of the user.

In addition to that, we added ATC expert observations of the user behaviour and questionnaires (NASA-TLX) answered by the user to collect information about user preferences and working experiences.
We defined two exercises for the experimental phase of the project. In a preparatory exercise we evaluated the accuracy of the sensor system that we plan to use within the experiment. Technical details on the performance and the evaluation results of the sensors can be found in Appendix A of [1].

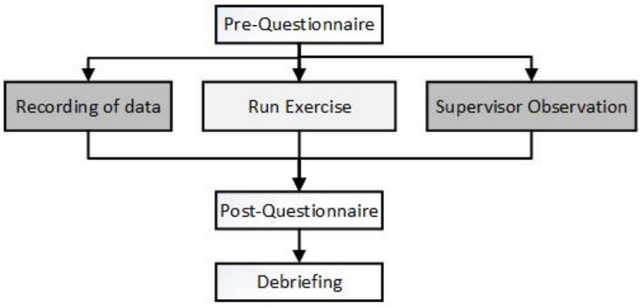

Fig. 2. Workflow of the designed Experiment to record all possible data of the test persons human body language.

To collect the actual body language data we designed In the second exercise, the actual body language data collection exercise, a participant performs a simulated 60 minutes ground controller shift at a simulated ground controller position. Figure 2 shows the workflow of the exercise. We recorded all the sensor and observation data during the experiment. After that, when replaying the video of the performance, ATC experts annotated in the time-line of the experiments when errors or suboptimal situations occurred, for example a blocking situation on the runway.

A. The Scenario

We chose the Hamburg Airport (Figure 1) scenario for our experiment since its layout has sufficient complexity to bring the test persons in a difficult situation while still simple enough to assess the quality of the decisions through experts.

The following constraints apply for the scenario:

- Simulation prepared for approx. 60 min.
- Arrivals are automatically simulated until touchdown, no change of route
- Departures are controlled until take off
- No runway change foreseen
- Taxiway routes can be selected by the operator
- 31 arrival flights, 27 departure flights.

The setup is based on a single simulated controller working position. No 3D view is available at the experiment, it concentrates on the ground traffic management. At the ground position we have seated the test person and the expert supervisor for taking notes and answering questions. A simulator position including the pilot simulation is located next to the ground position. The observer position with an observer taking additional notes (on the fly) is set up in a neighbouring room. The following modules have been used during the experiment:

- Traffic Simulator
- CWP with electronic flight strips and support information
- Active Message Queue[1] (AMQ) broker
- Eye-Tracker
- Mouse
- Microsoft Kinect

[1] http://activemq.apache.org/

The AMQ broker is the central distribution system for all data communication between all components. It allows an one-stop data exchange between different systems, sensors and modules by using customized XML messages. An SQL[2] data base is used for data logging allowing the course of events to be replayed as often as needed. The setup of the simulated working position is described in Figure 3.

Fig. 3. Working Position

B. The Exercises

The experiment consists of two exercises. Exercise 1 was performed to evaluate the sensor hardware and develop the data recording work flow as shown in Figure 3. Exercise 2 serves is data collection exercise. At the beginning of exercise 2 every participant received a map of the airport (Hamburg). They were informed that they could ask the ATC supervisor questions about the use of the simulator user interface to. When all questions were answered, the air traffic information was loaded into the simulator and the exercise started. Every 10 minutes, the supervisor asked the test person about the stress level and took notes of the current user performance from the expert's point of view. The experiment could run for 45 - 60 minutes maximum, depending on the current air traffic situation. Exercise 2 was performed eight times.

All participants work in the field of air traffic control but at different expert levels: as ATC, one En-route, two ground, one trained as a ground controller but works only in simulations experiments.

The following roles were participating in the experiment:

- Ground Controller: Participant
- Runway Controller: Manually Simulated
- Pseudo Pilots: Manually Simulated
- Supervisor
- Observers

[2]https://www.mysql.com/

V. The Data

An overview of the data collection is shown in the following table:

Topic	# Variables	# Events
Supervisor and Observer	3	693
StressLevel	3	49
FlightObject	42	2.846
Selections	6	1.773
Eye	10	211.739
GlobalMouse	4	187.652
Mouse	7	12.310
Kinect	23	34.912
Voice	12	7.320
Waspmote	4	10.978
Heart Rate	12	20.295
Eye AOIs	13	8.788
Mouse AOIs	13	9.418
Total:	**152**	**508.773**

Supervisor and Observer are events containing decisions recorded by the supervisor and the observers. The *StressLevel* is a rating from 1 to 5 (both combined with time stamp and user ID). To create the final observer list the recorded exercise was revised by a domain expert. The observer list consists of selected events which are rated positive, neutral and negative. A positive event occurs when the participant can successfully resolve a negative event. A negative event is for example: a blocking situation which delays other aeroplanes, too many arriving aeroplanes in queue (e.g. threshold bigger than 3), communication errors between pilot and ATCO.

The *FlightObject* contains events coming from the simulator, for example current status of aeroplanes. *Selection* events are created by the participant when selecting the digital flight strips.

For the mouse tracking we installed a mouse hook in the system to capture all mouse events (*GlobalMouse*). We also added a listener to the simulator (*Mouse*) collecting only specific feedback from the user interface, for example the ID of user interface (UI) elements.

Waspmote is an environmental sensor platform. We used it to capture temperature, humidity and lighting variations in the room.

After performing the experiment the collected data was further processed: The freely available eye tracking analysis software Ogama (Open Gaze And Mouse Analyzer) by Voßkühler (FU Berlin) [10] was used to calculate gaze and mouse fixations. The areas of interest (AOI) were defined within Ogama. The calculation of fixations then automatically takes the AOIs into account and connects the results.

The total number of 508.773 events was further analysed.

VI. Analysis

In the first stage after post-processing the measured data collection, we created simplified visualizations (using Tableau[3]) for initial discussion and exploration of the data, for example mouse events over time or stress level over time. As a result we agreed on a reference list of metrics, charts and initial findings.

Furthermore, we got an idea of more complex visualizations with combinations of several metrics and visualization types. Based on this we created a short list of the most promising metrics that could be combined in order to better represent the overall status of the users at each point in time (which is in our case, per minute).

Literature suggests that a combination of different measures assessing the same mental aspect like workload leads to more robust results than considering each measure on its own ([11], [12]). Therefore, we grouped this short list into four categories: task load, mental workload, attention and other metrics. These categories are known as factors that are related to operator performance. As these metrics may influence performance they can be regarded as independent variables whereas the performance measures serve as dependent variables.

At this stage we were able to formulate target oriented research questions, such as:

- How to improve the user interfaces usability?
- How to detect main causes that lead to mistakes (e.g., using air traffic information, eye tracker, mouse, heart rate data, body pose)?
- What are the unknown factors that contribute to higher stress levels or to the lack of situational awareness?
- Can air traffic information be combined with sensor information to improve the detection and classification?

Based on these, still very general, research questions we created a list of 15 concrete questions that guide our analysis, visualization and exploration of data when searching for good predictors of the user's behaviours. This includes exploration of

- the number of arrivals and departures per minute in relation to errors,
- increases in eye movements when the user is having periods of high workload that relates to the occurrence of negative observations,
- the relation between mouse pauses and increases in eye fixation times, the number of areas of interest visited per minute, lower heart rate variability, how the voice communications (number and speed of words spoken) is related to negative observations,
- the most preferred areas of interest by the users,
- how we might use the Kinect head pose and sound source angle variables to detect problematic time periods that might allow us to reduce the amount of data that needs to be analysed in real time.

[3]http://www.tableau.com

The answers to these questions (see [1]) for the complete list of research questions) will lead us towards the aim of the Sixth Sense project.

VII. Results

Combinations of multiple sensor recordings and different visualisations were used to detect striking user behaviour correlating to good or bad decisions. The data analysis uncovered several hints which can be used to further develop a suitable prediction algorithm. In the next sections the most promising hints are presented.

A. Mouse, Eye and Observations

We investigated the relation between the occurrence of bad decisions and the increase of eye and mouse events. According to the literature, pauses in the mouse movement are known to be linked with high workload periods when working with user interfaces [13]. The data indicates that there is a possible link between reductions in mouse movement and increases in the eye movements that coincide with the occurrence of negative errors (errors indicated by the experts). This indication could be useful for creating an algorithm that is able to detect or predict error periods. In Figure 4 we visualized the different fixation frequency changes (represented by lines) and what is happening in the observation data (red bar plots).

Another finding is, that the users were never moving the mouse at negative observations, they really stopped moving the mouse, probably to analyse the current situation.

B. Heart Rate and Workload

In addition to the heart beat per minute we measured the heart rate variability (HRV). The HRV indicates the fluctuation of the heart rate around an average heart rate. An average heart rate of 60 beats per minute does not mean that the interval between successive heartbeats would be exactly 1.0 second. Instead the interval may vary from 0.5 seconds up to 2.0 seconds. HRV is affected by aerobic fitness. The HRV of a well-conditioned heart is generally large at rest. During exercise, the HRV decreases as heart rate and exercise intensity increase. The HRV also decreases during periods of mental stress.

We could observe in the HRV data (for user4, user6 and user8) that: If we cross check the heart rate variability with the negative observations in the observations list, we can see that every time before an increase of severe negative observations, there is a steep descent (lower heart rate variability) on the inter-beat interval values.

According to the literature HRV can be a good indicator of high stress (e.g. [14], [15] and [16]).

We suggest to study the angle of the line plot (steep descent or steep climbing) because it could be used as a good predictor for moments of high stress and for the detection of intervals where negative observations are more prone to occur. When combined with the monitoring of the negative inter-beat interval value, this can be a good hint for negative situations.

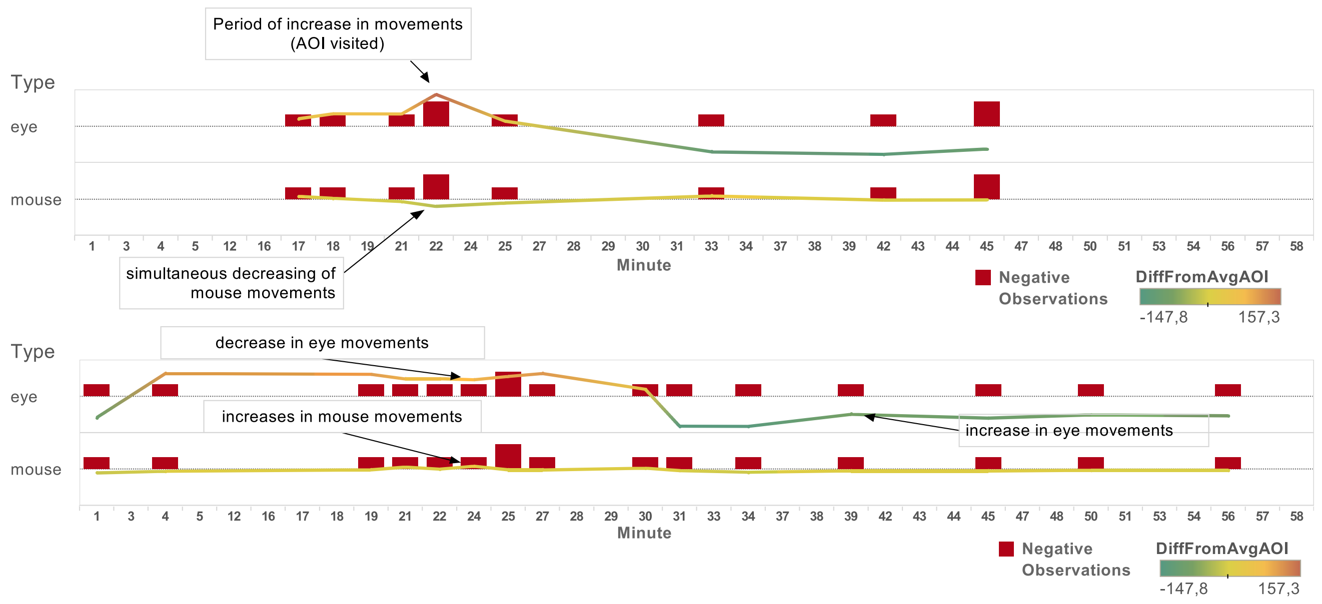

Fig. 4. Reductions in mouse movement and increases in eye movements correlate with the occurrence of negative observations (top: user6, bottom: user8).

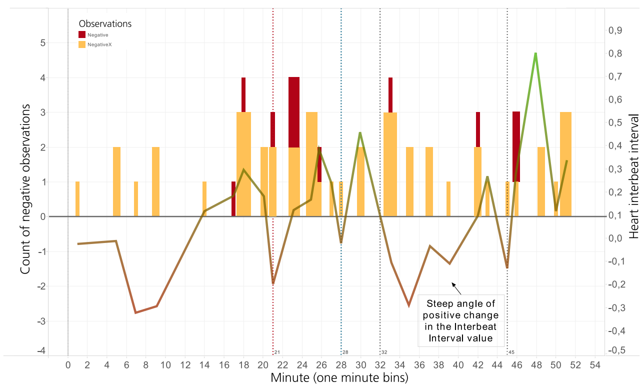

Fig. 5. The HRV together with the reduction in mouse activity, number of visual UI objects to be managed and eye tracking AOI frequency and duration provide good clues for moments of stress and high workload.

In Figure 5 we see dotted vertical reference lines when the HRV decreases (negative changes). Usually, this event is followed by an increase in negative observations.

It would be worth to analyse how quick this change (the slope of the inter-beat interval values) is still associated with the occurrence of negative observations when the user appears to be relaxed and not in a high stress situation.

C. Voice and Observations

There are fundamental differences between human speech and the more traditional forms of computer input. With the voice recognition system the number of words spoken by the ATCO was extracted.

It was not possible to check for mismatches between eye-tracking (of call signs) and the spoken call signs. This was due to the fact that the simulator UI in the radar area did not provide information about the call sign when the user is looking to a specific aeroplane. In addition the eye-tracker sensor did not allow us to track the small points representing aeroplanes in the radar with sufficient accuracy. We would need additional and improved selection strategies for that. For this reason, it was not possible to cross check if the call signs viewed by the user matched the call signs spoken by the user, although the voice recognition system could recognize the spoken call signs with a very high degree of accuracy.

In the future we could improve the simulator to provide feedback information when a user is looking to an aeroplane in the radar area.

However, we could observe a relation between the increase in the number of words used by the ATC and the occurrence of negative observations.

This seems to follow always the same pattern: there is a clear decrease in the number of words used followed by a significantly increase in the number of words spoken by the ATCs.

Especially by looking at the negative observation description this seems to correlate with the worst situations annotated by the experts (e.g., putting on hold several times the same aeroplane, resolving the crossing of runways or having too many aeroplanes to be resolved in the taxi or departure strip bays).

D. Head Position and Observations

With the Kinect we collected data about *head coordinate state, head coordinates, head pose coordinates, head rotation state (left, right, up, down) microphone beam angle and sound source angle, user in range* and *user tracked/not tracked*. We have found the *head coordinate state* and the user in range variables very promising for implementing a future error prediction system.

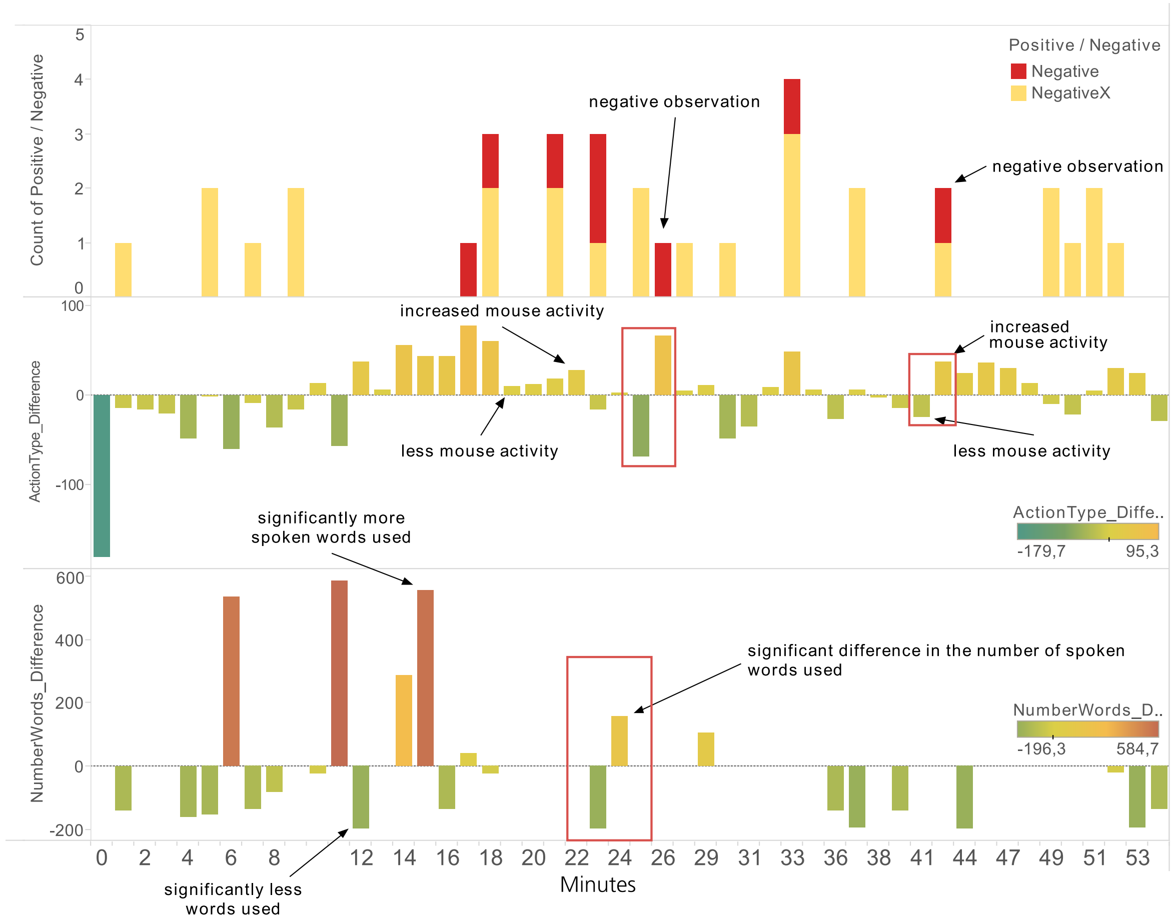

Fig. 6. Direct relation between increase in the number of words used by the ATCO and occurrence of negative observations.

As we can observe by considering only the variable *head coordinate state* = 0, 2, 4, 5, 6, 7, or 9 (we removed states: 1, 3 and 8) and the variable *sound source angle* (between -26.6 and 36), we can include in the same time interval at least 96% of all negative observations reported by the experts. We envision the usage of filtering mechanisms to reduce the amount of date that needs to be processed using the Kinect data.

E. Sequential Pattern Analysis by Event Tracing

A first step towards the detection and prediction of interaction sequences that lead to errors was done by analysing traces of possible events like eye and mouse fixations by parametrizing a Variable Length Markov Model (VLMM) ([17]). VLMMs provide a simple but efficient mechanism to capture behavioural dependencies and constraints. States in a VLMM can easily be interpreted, since each state is labelled by a corresponding subsequence within the data. A state chart can be calculated from the VLMM and the most probable state sequences can be determined. The occurrence of a state can be associated with a timespan within the data. A state can have different attributes, for example complexity or the entropy of the probability distribution of next events. These measures can additionally be used to look for patterns associated with the observations. For this analysis we applied a tool of Fraunhofer FKIE called "Event Analyzer". The underlying probability theory can also be found in [17].

Figure 7 shows the user interface. On the right there is the state chart, where a state in our case is an eye or mouse fixation area on the simulator display which is divided into *radar*, the flight strip area (*taxiin*, *taxiout*, etc.) and the *offscreen* area (locate here below the display). The simulator display is on the left including the user whose event data is displayed in the state chart. In the centre is a tree view of the state sequences showing the state that follows most probably as next. The nodes of the tree view can be expanded for exploration of the sub sequences. At the bottom a time line histogram is displayed showing the distribution of the occurrence of a selected state. With the tabs above the simulator display different visualizations can be selected, in this case the UI visualization is chosen. In summary, Figure 7 displays the probability distribution of the selected state augmented on the screenshot of the Sixth Sense simulator user interface.

The general recipe for analysing the data with the Event Analyzer tool is the following:

- select a user, then a dataset on the left side, for example "eye" (areas of fixation)

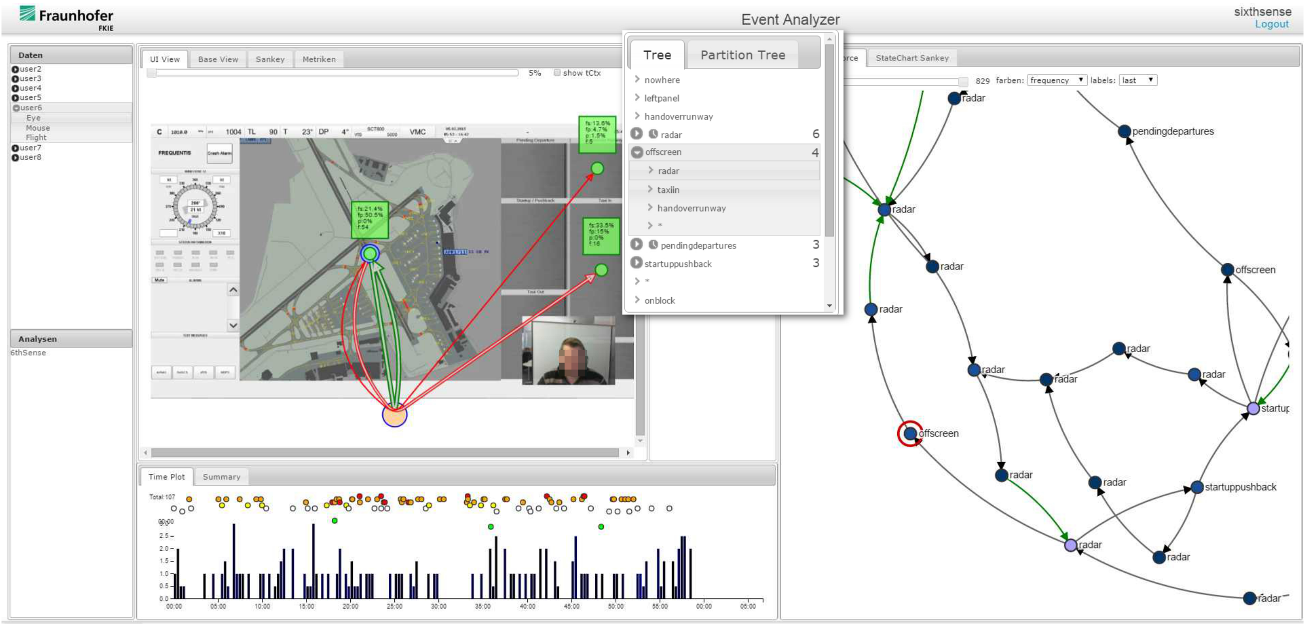

Fig. 7. Screenshot of the user interface with displayed transition probabilities of state sequences, here areas of eye and mouse fixations

- at the tree area find states with longer sequences, or, even better, go directly to the state chart area (right side) to select a node (AOI). The state chart shows the eye sequences and the intensity, the red color of the nodes will show the area of interest with a higher frequency
- select one of the interesting red nodes
- in the center area at the simulator display we see the representation of these sequences originated/ended at the selected AOI, red arrows represent a decreasing and green arrows an increasing probability
- look at the time plot at the bottom and check for the red dots, when errors are reported by the supervisors and at what time the selected AOI (state) occurred, check the length of the eye/mouse sequences

The selected state in Figure 7 is *offscreen*. We are interested in the area where the user tends to look at after focusing *offscreen*. We see that the user looks first to *radar* (second step on the tree) but also with an increasing probability (larger green arrow). On the bottom part we can study the time plot and see when errors occurred, especially at what time the importance of this state played a bigger role in the sequence by looking at the bigger bar plot.

As an outcome, the event trace analysis showed that the most frequent state sequence is very distinct for each user and occurs nearly two times more often than the second one in the ranking list. The detected mouse state sequences are significantly shorter than the eye state sequences. Many state sequences only contain one single event. However, the reason for this is that much less mouse events were present, since the mouse has not been moved as much as the eye.

Another possibility to study the data is to focus on the most complex state sequences. All complex state sequences contain many successive *radar* fixations, preceded by another fixation, mostly *taxiin*. This indicates that the probability to return to an AOI is increased after successive fixations on *radar*. This reflects that in the cognitive workflow the operator has to complete a task associated to the preceding AOI, for example *taxiin*, by collecting the necessary information on *radar* and returns back to the AOI where the information has to be placed.

Although, due to the limited number of test persons no obvious correlation of states with negative observations can be read from the histograms, we find this approach a very promising one that is worth to be continued.

VIII. Conclusion

In order to collect meaningful data for the Sixth Sense project, we prepared and performed an experiment with ATCOs to gain as much data as possible in the available time frame.

We designed and implemented a software framework to collect user behaviour data in real time. We integrated different systems, sensors and sources of information. We started by unifying all the air traffic data sources with all sensor technologies. We integrated the supervisor, observer and ATCO stress level reports inside the software framework to analyse and treat all the aspects related with ”thinking aloud” and observational protocols in an automated manner. We included questionnaires to extract valuable information about different preferences or working experiences. We used this information to evaluate the difficulty of the experiment, usability of the system, workload, situational awareness, and performance. Including questionnaires into the data analysis is definitely a promising option, but, again, due to limited resources and data quantity we could not exploit the maximum potential.

We collected about 600.000 events. The handling of the complexity (many events, different datasets, multiple variables, time series, and behavioural sensed data) required multiple strategies for pre-processing, analysis, discussion sessions, exploration and visualization. Guided by these findings, 15 research questions were established and addressed during data analysis.
The main results of the data analysis and visualizations are presented in the Sixth Sense report [1]. We found some very promising metrics and as a consequence promising hints in relations between different data streams. One is the link between reductions in mouse movement and increases in the eye movements, correlated with the occurrence of negative observations.
The combination of HRV together with the reduction in mouse activity, the number of visual UI objects to be managed and the eye tracking AOI frequency and duration provides good clues for anticipating moments of stress and high workload.
Furthermore, there are direct relations between an increase in the number of words used by the ATCO and the occurrence of negative observations. And finally, we found a correlation between the users head position and negative observations that indicates promising model creation for predictions.

The presented results show how important the incorporation of behavioural analysis is for the design of automated systems that are able to analyse, detect and predict in safety critical situations. Our results can also be applied to improve existent systems and UIs. We identified behavioural causes that play an important role in the report of higher stress levels, high workload or even to the lack of situational awareness. These behavioural causes are for example the number of visual objects to be handled (arrivals and departures per minute), number of areas to be monitored, delays and problems in communication with the pilot, time accumulation and also emotional factors.

Our test setup and process proofed right. The selected analytical tools and visualizations are feasible although there are numerous other possibilities which are worth to be explored and integrated in the future analysis. Due to the nature of this kind of explorative research projects with restricted resources no statistical relevance in the found patterns is recognisable. The number of test persons was too low.

However, the concrete patterns which have been found allow deriving early indications for good or bad decisions. There are good indications for positive results with sufficient statistical power, when more test data and more time is available for sensor permutation analysis.

Acknowledgment

This work is co-financed by EUROCONTROL acting on behalf of the SESAR Joint Undertaking (the SJU) and the European Union as part of Work Package E in the SESAR Programme. Opinions expressed in this work reflect the author's views only and EUROCONTROL and/or the SJU shall not be considered liable for them or for any use that may be made of the information contained herein.

This project is supported by the government of the Austrian federal state of Styria within the research project "CUBA - Contextual User Behavior Analysis".

We wish to thank Carsten Winkelholz and Jessica Schwarz (Fraunhofer FKIE) for their invaluable contributions and expertise in the fields of air traffic control, evaluation methodologies, psychology, data analysis.

References

[1] N. Silva, E. Eggeling, V. Settgast, M. Poiger, F. Grill, T. Zeh, C. Winkelholz, and J. Schwarz, "Verification report: Experimental test report," *Sixth Sense, Deliverable 4.2*, pp. 1–98, 2015.

[2] S. Lehmann, R. Dörner, U. Schwanecke, N. Haubner, and J. Luderschmidt, "Util: Complex, post-wimp human computer interaction with complex event processing methods," in *10. Workshop "Virtuelle und Erweiterte Realität" der GI-Fachgruppe VR/AR, Aachen*, 2013, pp. 109–120.

[3] G. Cugola and A. Margara, "Processing flows of information: From data stream to complex event processing," *ACM Computing Surveys*, vol. 44.3, p. 15:115:62, 2012.

[4] C. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006, vol. 4. [Online]. Available: http://www.library.wisc.edu/selectedtocs/bg0137.pdf

[5] B. Brunton, M. Botvinick, and C. Brody, "Rats and humans can optimally accumulate evidence for decision-making," *Vol. 340. Science AAAS, New York (2013)*, vol. 340, no. 6128, pp. 95–98, Apr. 2013. [Online]. Available: http://dx.doi.org/10.1126/science.1233912

[6] S. Shorrock and B. Kirwan, "Development and application of a human error identification tool for air traffic control," *Applied Ergonomics (2002)*, vol. 33, no. 4, pp. 319 – 336, 2002. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0003687002000108

[7] M. Steiger, J. Bernard, S. Mittelstadt, H. Lucke-Tieke, D. Keim, T. May, and J. Kohlhammer, "Visual analysis of time-series similarities for anomaly detection in sensor networks," *Computer Graphics Forum (2014)*, vol. 33, no. 3, pp. 401–410, 2014.

[8] F. Stoffel, F. Fischer, and D. Keim, "Finding anomalies in time-series using visual correlation for interactive root cause analysis," *VizSec. ACM (2013)*, pp. 65–72, 2013. [Online]. Available: http://doi.acm.org/10.1145/2517957.2517966

[9] D. Perez, L. Zhang, M. Schaefer, T. Schreck, D. Keim, and I. Diaz, "Interactive feature space extension for multidimensional data projection," *Vol. 150. Neurocomputing. Elsevier (2015)*, vol. 150, Part B, no. 0, pp. 611 – 626, 2015. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0925231214012879

[10] A. Voßkühler, V. Nordmeier, L. Kuchinke, and A. M. Jacobs, "Ogama (open gaze and mouse analyzer): open-source software designed to analyze eye and mouse movements in slideshow study designs," *Behaviour Research Methods*, pp. 1150–1162, 2008.

[11] A. Haarmann, W. Boucsein, and F. Schaefer, "Combining electrodermal responses and cardiovascular measures for probing adaptive automation during simulated flight," *Applied ergonomics*, vol. 40, 2009.

[12] K. F. V. Orden, T. P. Jung, and S. Makeig, "Combined eye activity measures accurately estimate changes in sustained visual task performance," *Biological Psychology*, vol. 52, p. 221240, 2000.

[13] Q. Guo and E. Agichtein, "Exploring mouse movements for inferring query intent," in *In Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval, SIGIR '08, pages 707 - 708, New York, NY, USA, 2008. ACM.*, 2008.

[14] A. H. Roscoe, "Assessing pilot workload. why measure heart rate, hrv and respiration?" *Biological Psychology*, vol. 34, pp. 259–287, 1992.

[15] G. Costa, "Evaluation of workload in air traffic controllers," *Ergonomics*, vol. 36, pp. 1111–1120, 1993.

[16] B. Mulder, H. Rusthoven, M. Kuperus, M. de Rivecourt, and D. de Waard, "Short-term heart rate measures as indices of momentary changes in invested mental effort," *Human Factors Issues*, 2007.

[17] C. Winkelholz and C. M. Schlick, "Statistical variable length marov chains for the parameterization of stochastic usermodels from sparse data," in *IEEE interation Conference on Systems, Man, and Cybernetics*, 2004.