The Sustainable Balancing Act: Operator's Situation Awareness and Forecasting Weather-Dependent Energy Resources

Remy Rey, Henrik Artman, and Andrii Matviienko
dept. of Human Centered Technology
KTH Royal Institute of Technology
Stockholm, Sweden
{rjfrey, artman, andriim}@kth.se

Mario Romero
dept. Media and Information Technology
Linköping University
Linköping, Sweden
mario.romero@liu.se

***Abstract*—Mass deployment of sustainable, weather-based energy sources will present new challenges to the power balance, as production forecasts are expected to fluctuate randomly in timing and intensity. In this study of a novel user interface, the influence of the magnitude of the forecast error is tested in a simulated power balance act. Participants, with no formal experience of control rooms, activate regulation reserves in three conditions and are evaluated on actual performance and self-reported measures through standardized questionnaires. It is shown that despite self-reporting similar levels of cognitive load and Situation Awareness (SA) in all conditions, higher magnitude of forecast error have a large impact on participant performance. This hints at the importance of providing the operator with the appropriate tools and enough information to manage changes in forecast and deviations from earlier planning. In particular, designers of visualizations should include a representation of current and obsolete forecasts in real-time operations timelines.**

***Index Terms*—visualization, grid balancing, user study, hci, situation awareness, cognitive load, forecast**

I. Introduction

In the power grid, electricity needs to be balanced: at any instant, what is produced must be consumed. Any substantial deviation, under- or over-production, which causes the frequency to depart from the norm of 50hz/60hz, can damage all connected assets and machines and, in extreme cases, lead to blackouts. A grid balancing operator has the role of activating reserves to increase or decrease production in real time to maintain the balance.

Endsley and Connors [1] state that high levels of Situation Awareness (SA) are required to maintain safe and reliable operation in the power grid. SA is commonly defined as a hierarchichal, three-tier framework: "[Level 1:] the perception of the elements in the environment within a volume of time and space, [Level 2:] the comprehension of their meaning, and [Level 3:] the projection of their status in the near future" [2]. Higher SA translates into a higher understanding of how a projected future will affect and be affected by ongoing decisions and actions. The authors [1] conclude, in the context of grid balancing, that most control rooms have poorly configured user interfaces. However, they do not give concrete suggestions for the design of such a system, and in particular the representation of information.

Furthermore, it is expected that a mass deployment of weather-dependent energy sources, while necessary for reaching net zero goals, will, conversely, render less accurate production forecasts. In the best cases, forecasts tools for the very short term (15 min horizon) still exhibit error rates of 3% to 5% [3] [4] which are substantial for power frequency control. For the balancing operator, this means paying attention to both the immediate situation of consumption and production and their projected values in the near future, and taking cost effective actions to correct the imbalance trend. The present study investigates how and if a dynamic and interactive visualization supports operators in achieving SA for handling weather-dependent renewable energy sources which "can be quite variable in both timing and intensity" [1].

The contribution in this paper is two-fold: First, an interface is designed following state-of-the-art information visualization guidelines to provide an overview of the power imbalance and regulating reserves. Second, the interface is used to study the impact of inaccurate forecasts brought about by weather-dependent energy sources on the SA of the operator. Inspired by Brehmer's [5] micro-world approach to dynamic decision making, a focused and simplified, yet structurally equivalent, game of the balancing task has been designed. Participants are presented with clear goals: maintaining a (1) safe and (2) affordable power grid and a forecast evolution mechanism that challenges their decision-making. This gamified version of an operator task exhibits delayed feedback and mental tracking of trends over time as some information remains hidden. If participants achieve high levels of SA they can predict the future grid state and plan margins accordingly. By facing different levels of forecast error, users of the system will show how behavior and SA in operations can be affected by the increasing penetration of renewable energy.

The research leading to these results was supported by the Graduate School in Energy Systems (FoES) funded by the Swedish Energy Agency (Grant no. P2022-00177).

979-8-3315-2503-3/25/$31.00

II. Methodology

A. Design of the interface

The user interface (Fig. 1) was elaborated iteratively by data visualization experts and followed the best practices of the field. The work is inspired by other contributors in the field working on maps [6] with contour [7], pseudo-maps like mosaics [8] or force-directed graphs [9]. Limited effort from the visualization community is dedicated to the state of imbalance and reserves.

The aim of the design was to represent the state of power balance in the grid and the associated reserves, to support the goal of a safe and affordable transmission. The visualization includes the following concepts:

- **system imbalance** the difference between production and consumption, representing the Area Control Error
- **net imbalance** the system imbalance plus the sum of all activated reserves
- **automatic reserves** the procured reserves that are automatically activated by the system, representing the different types of automated balancing services
- **manual reserves** the procured reserves that needs activation from the participant, representing the manual frequency restoration reserve or equivalent

These concepts abstract the complexity of the frequency ancillary service markets while retaining the main characteristic of the imbalance task: combining manual and automatic response to deal with unexpected events. Considering the grid congestion is outside the scope of this study.

The manual reserves are sliced in five-minute slots that can be specified independently, with a fixed ramp time of three minutes. The automatic reserves provide a perfect and instant response to imbalances and compensate for manual reserve overshoots, if the user provides more than the grid requires.

On the overview panel (Fig. 2), the user sees the planned imbalance and reserve status for the next 30 minutes and what happened in the last 15 minutes. In a desirable state, the net imbalance of the system, represented by the green line, should be at zero at all times. The dashed lines (automatic balancing limits) shows the users how much margin capacity the automatic balancing has not activated yet and give a sense of the safety level. When the user hovers the graph, a tooltip shows the exact values for the corresponding time. The elements that a user needs to perceive to achieve the first level of SA are the following: (1) bumps in the green line (as shown in Fig 4), (2) the red and blue areas (Fig 2a), (3) the dashed lines (Fig 2a) when close to zero and (4) the changes in the white line between two updates. Those elements help the user achieve the second level of SA if they can connect them to the simulation model explained above. (1) and (2) represent violations of the goals and (3) and (4) indicates changes in the forecast.

Below the overview, a control panel enables the user to change the planned level of activation for manual reserves (Fig. 3). A single power level is set for every five-minute interval, and a line shows the actual energy output including

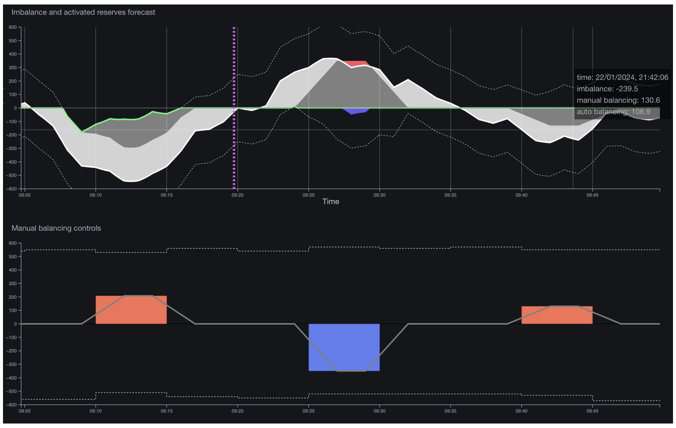
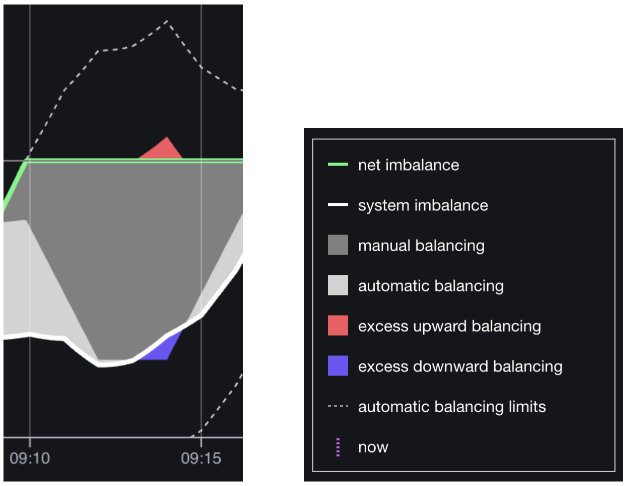


Fig. 1: The UI presents a 45 minute timeline of imbalances and reserves (top) and allows manual activation (below)

the ramp up and down. The exact value is shown in text next to the cursor when hovering.

The full interface (Fig. 1) shows a timeline of 10 time slots over the length of the screen. The UI was designed for a 32” QHD display (2560x1440 pixels) and a mouse. The UI is dynamic and updates in two-seconds intervals. This is close to standard SCADA refresh rates and gives the user an opportunity to perceive changes during the animation.



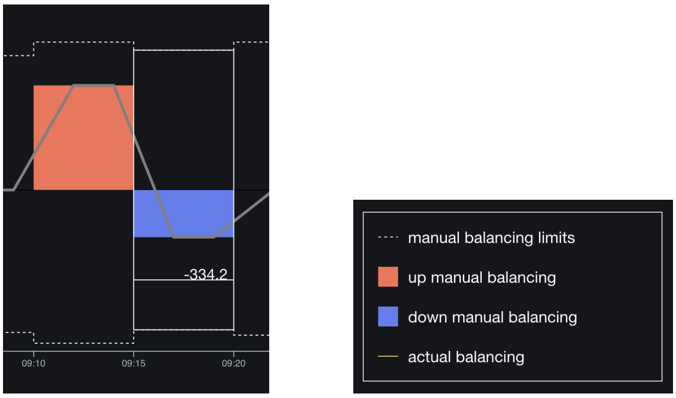
(a) A time slot presenting the imbalance and status of the reserves.

(b) Legend of the imbalance graph continuously viewable by the user.

Fig. 2: User interface showing an imbalance overview.

B. Experimental Protocol

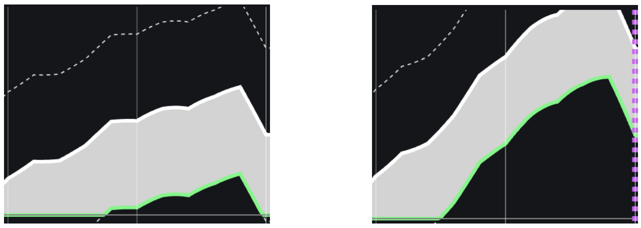
Eighteen (11F / 7M) participants, aged 19 to 41, were recruited through the broadcasting of emails within and outside the university. The experiment lasted one hour and the participants were compensated with 150 SEK. Six participants were electrical engineering students recruited after a course on control and communications in the power grid (EE). The rest had little to no knowledge of power system dynamics.



(a) Manual activation control for two timeslots.

(b) The legend for the control panel.

Fig. 3: The control panel where one changes planned activations of reserves. In (a) the user can click within the dotted lines to change the power level. (b) is displayed on the side.



(a) The event as it first appears on the forecast.

(b) The event after it happened and crossed the now line.

Fig. 4: In a high condition, the participant sees an imbalance of 200MW 30 minutes in the future (a), the imbalance increases linearly over time and reaches the 'now' line at 450MW (b), as the -250MW forecast error vanishes.

The experimenter read a script that introduced the need for reserves, the role of the balancing operator, and the user interface (UI). The participant then completed a test scenario with one balancing event to familiarize themselves with the controls. The testing consisted of 3 scenarios, each independently combined with a forecast error of 50MW (Low condition), 150MW (Mid condition) or 250MW (High condition). Imbalance time series come from historic data downloaded on the ENTSOE transparency portal, onto which are added artificial events and forecast error. In this within-subject design, conditions and scenarios were balanced independently using a balanced Latin square [10], and combined in 6 different sequences to control for learning effects and differences between scenarios.

Each scenario represented one hour of operation and was sped up six times to fit in 10 real-time minutes (all time references below will be expressed in the sped up simulation time). One scenario presents a sequence of six balancing events each lasting 15 minutes and evenly spaced over the time visible to the user (60 minutes of simulation time and 30 minutes of forecast). The participant is only evaluated on the actionable events whose forecast had time to evolve, therefore the first and sixth events are only providing context to the others and are excluded from the analysis. Because the viewable forecast window is 30 minutes into the future, the participant always sees two events at a time. This challenges the distribution of attention and raise the bar to achieve first level SA (perception) and track changes in both events.

The participant had access to enough reserves to cater to every balancing event, including the forecast error. When an event starts showing on the screen, the forecast error is at 100% of the condition value, then it linearly decreases to 0% until the event happens (an example is shown in Fig 4).

After each scenario the participant completes a NASA task load index (NASA-TLX) [11] questionnaire on a tablet and a SA Rating Technique (SART) [12] questionnaire on paper. The experimenter also records answers to a set of follow up questions for the SART questionnaire on SA and the participant's strategy. Any unprompted comments, for example on the user interface design, during or after each scenario is also recorded.

The participant was given instructions to follow a set prioritization of goals: (1) maintain the net balance of the system at zero at all times, (2) limit overuse of the balancing services (3) maintain reasonable margins for the automatic balancing reserves and (4) plan ahead as changes cannot be ordered for the nearest 10 minutes. This is in line with the power grid operator's goal of keeping the system safe and affordable, and ensures the participant needs to adjust their plan with the evolution of the forecast. To measure performance, all the participants' decisions (clicks) were recorded along with the net imbalance and activated reserves level for each simulation timestep.

III. Results

A. Self-reported metrics

Participants self-reported in two questionnaires: NASA-TLX and SART. NASA-TLX was in general well understood, while the questions in SART were ambiguous for some participants and required explanations from the examiner. Fig. 5 displays all the measure from the TLX. For most dimensions there is no clear trend among the three conditions. A closer look at the self reported performance score shows a progression from Low to High (Fig. 6).

Only the performance dimension of the NASA-TLX showed a significant difference across conditions, as indicated by the Friedman test ($\chi^2 = 11.44$, $p < 0.01$). No other dimensions from the SART or TLX questionnaires reached significance (Table I). Post hoc analysis using the Nemenyi test revealed a significant difference between the High and Low conditions ($p < 0.01$).

B. User Performance: imbalance and reserve score

During the experiment, the participant could modify the manual reserve level from -500MW to +500MW by clicking on the control panel for any future timeslot. There was no

TABLE I: Friedman test results for TLX self-reported measures

TLX Dimension	χ^2	p-value
Mental	5.4	0.067
Temporal	2.8	0.247
Physical	2.0	0.368
Performance	11.4	< 0.01*
Effort	4.9	0.088
Frustration	2.7	0.262

*Statistically significant

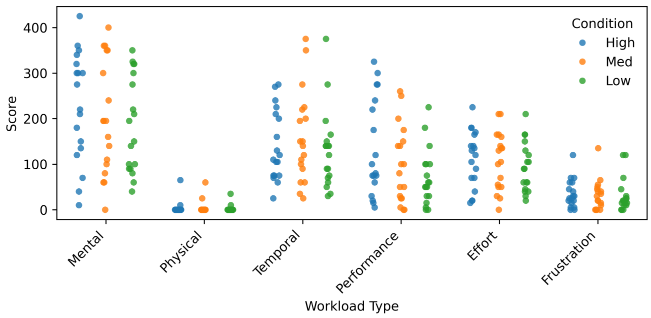


Fig. 5: Self-reported values for the six TLX dimensions

significant difference in the number of clicks, nor explicit patterns differences between the conditions.

The Friedman test revealed a significant difference between the net imbalance achieved in different conditions, $\chi^2 = 20.3, p < 0.0001$. Fig. 7 shows a very different distribution of score in High condition. Post hoc analysis using the Nemenyi test confirms significant differences between the High and Mid conditions ($p = 0.033$) and between the High and Low conditions ($p < 0.0001$). Most participants ($N = 13$) achieved no imbalance during the Low condition scenario (perfect score). No difference was found between EE and non-EE participants. The actual performance measured by imbalance contrasts the self-reported performance score as the High condition is now clearly distinct from the Mid and Low conditions.

The Friedman test showed difference for both the activated manual reserve score ($\chi^2 = 13.778, p < 0.001$, Fig 9) and the

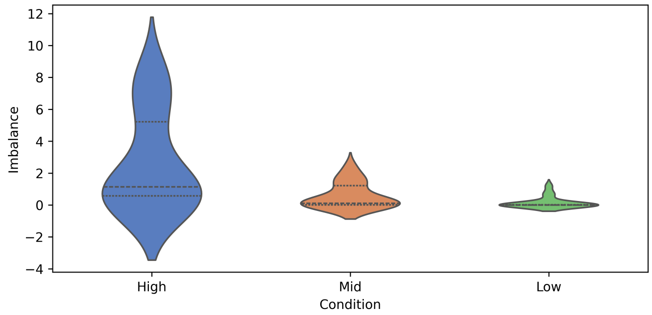


Fig. 7: Imbalance score per condition. Lower is better. (Violin plot with quartiles)

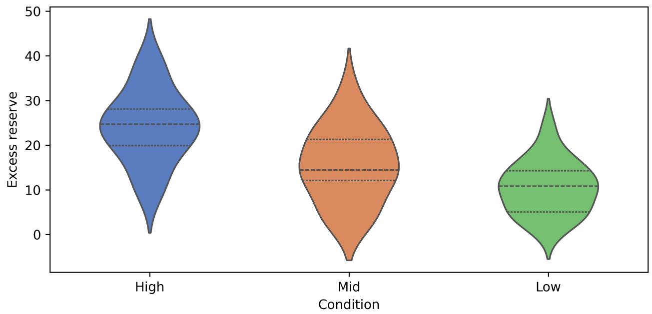


Fig. 8: Excess reserve score per condition. Lower is better. (Violin plot with quartiles)

excess reserve score ($\chi^2 = 14.778, p < 0.001$, Fig. 8). Post hoc analysis with the Nemenyi test only shows a significant difference between High and Low conditions ($p < 0.001$ for both scores).

C. *Qualitative feedback*

1) User experience: A majority of participants ($N = 10$) reported explicitly that the UI was ”nice” and ”intuitive”, or that they could solve the problem visually (in contrast to looking at values). For some participants the abstraction was too high and what was happening behind the scenes was unclear. They could solve the task without reasoning in terms

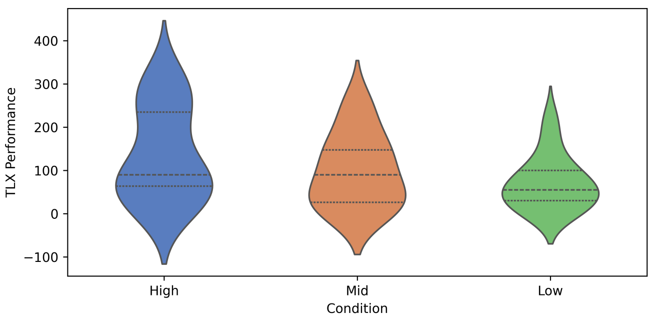


Fig. 6: Self reported performance (TLX). Lower is better. (Violin plot with quartiles)

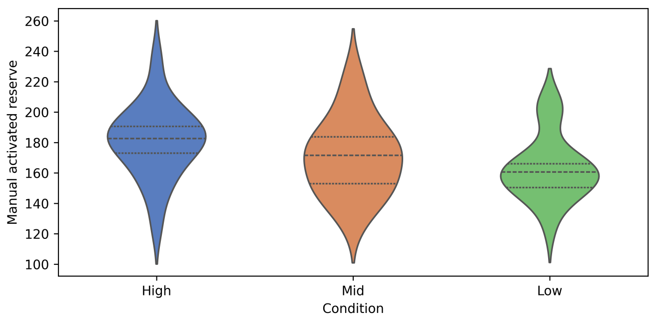


Fig. 9: Manual reserve score per condition. Lower is better. (Violin plot with quartiles)

of charging and discharging batteries or turbines spinning faster and slower. Only one participant who had previous experience from the control room made explicit use of exact numbers.

2) Strategies: Participants reported using two strategies. The first is to decide on a set point as soon as the event is visible and then only adjusting right before it slips out of the actionable timeframe. In a second variation, some users described scanning continuously all the actionable points.

Overall lack of feedback of the action was reported as an issue, even as it was provided in the UI. A few participants also raised the absence of alarms, to highlight a spent out margin. This hints at a challenge in achieving Level 1 SA or Level 2 SA, and therefore preventing the ability to anticipate imbalances (Level 3 SA).

IV. Discussion

The participants were facing an evolving forecast model and had to adjust their plan as the reality of the events became more precise. The High condition presented a challenge as performance decreased sharply (Fig. 7) but was not perceived as different from the Mid condition (Fig. 6). This clearly indicates that in the absence of proper tooling to understand how updated forecasts are departing from an older estimation, operators are at risk of underestimating the instability of the situation and suffer from a lack of margins. In this study, it was expected that participants would eventually grasp the pattern of change in the forecast as it was binary (amplification or dampening of the balancing event), but there is no clear indication they succeeded. The task would have been rendered trivial if users had access to the trend of the forecast value at the apex of the events, as it is a very predictable linear pattern.

In the control room, the forecast of events will not behave so civilly, and a weather pattern like a cold front can slide temporally and appear earlier than expected. The current design is limited, but we can reason that in a realistic setting, where the operators needs to juggle with different tasks and contexts, perceiving and understanding uncertainties in forecast brought by weather-dependent power plants will be challenging. Hock et McGuiness [6] also concluded that analyzing near-future situations requires specific visualizations and that changes in the datasets over time should be made clear to the operators, if they are to achieve Level 3 SA. Antonanzas et al. [3] also consider stochastic forecasting tools that, unlike the presented design, provide distributions instead of single values. They however do not propose any solution to visualize stochastic forecast of system imbalance or the state of planned reserves. Visualization of stochastic results constitute a good avenue for further research.

The second limitation of this study is that it considers balancing as a centralized task with a single degree of freedom. In reality, balancing services are distributed over the area and the activations will impact the power flow throughout the whole network. Future implementation of micro-worlds containing a power grid should address the spatial distribution as an additional constraint for operations. Moreover, that would set an interesting stage to understand dynamic decision making and collaborative work distributed between two operators: one monitoring the balancing between production and consumption, and the second limiting overloads on lines and stations.

Endsley and Connors [1] warn against adopting trendy new visualizations. The current study could be an example of that. At the same time different forms of mediating visualizations must be evaluated before put into practice. Furthermore, as it is hard to involve actual balancing operators in research for future control rooms design, experimental studies may suffice as is showcased here.

V. Conclusion

This user study had the participant face a balancing act on a novel interface. While they reported no increase in cognitive load for higher levels of forecast error, their performance showed a sharper decrease. It is concluded that perception of trends in forecast updates is hard and that constitutes a real concern for control rooms. Tool builders for the control room must account for this gap in their design as we are deploying weather-dependent energy sources.

References

[1] M. R. Endsley and E. S. Connors, "Enhancing situation awareness in power systems: Overcoming uncertainty and variability with renewable resources," in *Renewable Energy Integration*. Elsevier, 2017, pp. 395–404.

[2] M. R. Endsley, "Toward a theory of situation awareness in dynamic systems," *Human factors*, vol. 37, no. 1, pp. 32–64, 1995.

[3] J. Antonanzas, N. Osorio, R. Escobar, R. Urraca, F. J. Martinez-de Pison, and F. Antonanzas-Torres, "Review of photovoltaic power forecasting," *Solar energy*, vol. 136, pp. 78–111, 2016.

[4] X. G. Agoua, R. Girard, and G. Kariniotakis, "Photovoltaic power forecasting: Assessment of the impact of multiple sources of spatio-temporal data on forecast accuracy," *Energies*, vol. 14, no. 5, p. 1432, 2021.

[5] B. Brehmer*, "Micro-worlds and the circular relation between people and their environment," *Theoretical Issues in Ergonomics Science*, vol. 6, no. 1, pp. 73–93, 2005.

[6] K. P. Hock and D. McGuiness, "Future state visualization in power grid," in *2018 IEEE International Conference on Environment and Electrical Engineering and 2018 IEEE Industrial and Commercial Power Systems Europe (EEEIC/I&CPS Europe)*. IEEE, 2018, pp. 1–6.

[7] K. M. Gegner, T. J. Overbye, K. S. Shetye, and J. D. Weber, "Visualization of power system wide-area, time varying information," in *2016 IEEE Power and Energy Conference at Illinois (PECI)*. IEEE, 2016, pp. 1–4.

[8] A. B. Birchfield and T. J. Overbye, "Mosaic packing to visualize large-scale electric grid data," *IEEE Open Access Journal of Power and Energy*, vol. 7, pp. 212–221, 2020.

[9] P. C. Wong, K. Schneider, P. Mackey, H. Foote, G. Chin Jr, R. Guttromson, and J. Thomas, "A novel visualization technique for electric power grid analytics," *IEEE Transactions on Visualization and Computer Graphics*, vol. 15, no. 3, pp. 410–423, 2009.

[10] A. L. Edwards, "Balanced latin-square designs in psychological research," *The American journal of psychology*, vol. 64, no. 4, pp. 598–603, 1951.

[11] S. G. Hart and L. E. Staveland, "Development of nasa-tlx (task load index): Results of empirical and theoretical research," in *Advances in psychology*. Elsevier, 1988, vol. 52, pp. 139–183.

[12] R. Taylor, *Situational Awareness Rating Technique (Sart): The Development of a Tool for Aircrew Systems Design*, 07 2017, pp. 111–128.