On leveraging cockpit data to extend Human Factors understanding on the flight deck

Sigoillot, C.[1] and Ligier, V.[2]
[1]Dassault Aviation, Mérignac, France
[2]Dassault Aviation, Saint-Cloud, France
camille.sigoillot@dassault-aviation.com
valentin.ligier@dassault-aviation.com

INTRODUCTION

While being taken into account in the aircraft design process for a few decades, understanding Human Factors on flight decks remain an open and active research topic. Nowadays aircraft systems are extensively monitored and any failure mode is characterized and taken into account in the aircraft design to guarantee safety of the aircraft operation. Human Factors are of course also considered in the aircraft design process (e.g. CS-25 §1302) but the extent of knowledge and “human failure mode” characterization (i.e. degraded human states), especially live monitoring in operation, is far from reaching the level of other aircraft systems.

Recent accidents (Bureau d’Enquêtes et d’Analyses pour la sécurité de l’aviation civile, 2012 and The House Committee on Transportation & Infrastructure, 2020) have shown that Human Factors are still a contributor to aircraft accidents and regulations tends to be more and more prescriptive on this topic in a continuous improvement approach (United States Congress, 2020).

In this context, Dassault Aviation, as an aircraft manufacturer and more specifically a cockpit designer, aims at improving the knowledge of these factors with several end uses or development focus:

- Getting a better objective understanding of cockpit operational use by flight crew to improve future cockpit development;
- Introducing a crew state monitoring function: identification of physiological, cognitive states and interaction state (e.g. behavior);
- Collecting observable data on user evaluation of a new design to provide detached assessment.

This extended abstract discuss the opportunities offered by the analysis of collected data based on the end use and the source to extend Human Factors understanding on the flight deck. It relies on aircraft cockpits, training simulators and development simulators.

OBJECTIVES OF COCKPIT DATA ANALYSIS

Dassault Aviation as for the cockpit design has identified the following needs.

Gathering feedback on an existing deployed design

Aircrafts with modern avionic systems compatible with cockpit data collection have been in-service for about 20 years with a substantial fleet of several hundreds aircraft around the world. These figures allow to compute statistics thanks to the large variety of crew and operations.

This data could be analysed in order to provide support to Human Factors engineers when designing cockpit. When used along with contextual data (environment, active failure, etc.), insights derived from this data can be for example:

- Recurrent error detection (due to poor design choice, e.g. clockwise vs anticlockwise rotation knob)
- Reaction time to alerts (mean and standard deviation)
- Recurrent high workload tasks
- Phase of flight with loss of vigilance of the crew
- Pattern detection and correlation to operating procedures: workflow optimization

Knowledge of the context and type of operation is essential to really understand the meaning of the data, hence it will be difficult to fully automate the analysis.

Identifying crew states

This objective require the use of a crew monitoring system (e.g. camera, heart rate sensor, electro-encephalograph…). Cockpit interaction data alone is not sufficient to infer a multidimensional comprehensive crew state.

Knowledge of the cognitive or physiological state of a crew member can be used for example:

- In flight:
    - To detect degraded states of the crew that require a system or procedural counter-measure to increase general flight safety;
    - To adapt a Human-machine interface to different collaboration states.
- On ground:
    - To add an impartial source that detects crew states to help investigations on inflight incidents;
    - To gather more feedback and context on existing deployed design to support the objective described in the previous section.

Some states are hard to identify and ongoing research is to be achieved before it can be deployed and used in operation.

Evaluating a new design in development phase

During a cockpit development phase (either an evolution or a complete new design from scratch), new concepts or functions are considered and prototyped to improve existing cockpits considered as a baseline. However, today's evaluations of these concepts rely on subjective data expressed by test subjects who have experience on the baseline design. The idea is to leverage data that is collected from these prototypes to derive more objective evaluations of the new proposed concepts or functions.

A new design can be evaluated considering several factors, e.g. its interface, utility and performance, as a single function and as a part of a complex avionic system. The crew monitoring system mentioned above can be used for instance to detect an evolution of workload after adding a new function in a cockpit.

On the short term, this objective enables to gather impartial data to be used for the evaluation of a new design by and with crews. On the long term, such a feedback can also help create a baseline of knowledge on a system's performances in order to help qualification and certification.

SEVERAL MEANS OF DATA COLLECTION

Aircraft cockpits

Real aircraft cockpits provide the most representative data since flight crew is actually operating the aircraft in real-time with real external conditions. This data is extremely valuable for analytics.

Modern cockpit systems and especially avionic systems make it possible to easily collect interaction data (e.g. from physical controls on panels to soft keys in avionic displays) thanks to a modular architecture with data exchange across systems. However, integration of a dedicated system to existing aircraft design to collect additional observables is difficult since most of the aircraft are already deployed to customer home bases or in mission. Standalone solution are an option but require an additional step of post-recording synchronization and a different data pipeline.

Training simulators

Training simulators provide the best trade-off between cost/access and representativeness. The data quality depends a lot on the simulator quality. The more representative the simulator, the more representative the data. Quality of the simulated external environment, quality of the sounds, the movements of the simulator can help improve the data representativeness.

However, even with the best quality simulator, there remain biases from training. Training simulators are often used to help crews familiarize with a plane or to help dealing with rare but critical situations. Thus, training data will be representative of such flights with repetitive exercises or scarce failures and seldom of real operational ones.

Finally, training data will include a learning bias that decreases data representativeness. Nevertheless, carefully handled, this data can help identify issues in the cockpit design that training would cover. Also, repetitive training on a same situation with controlled scope and with a large crew panel provides a good opportunity to compute statistics on a controlled environment.

Development phase simulator

Development simulators provide the most accessible data. It is also the least representative of operational data. Indeed, development simulators are part task simulators that must be easily updated, adjusted and executed; they also often lack an immersive environment making the full involvement of operator difficult to maintain.

Thus, development simulators can hardly provide representative operational data.

Nevertheless, such simulators will enable the evaluation of the integration of a new function in the whole cockpit. It is then also possible to characterize, for example, the new mental load of a crew with the new function on a whole flight.

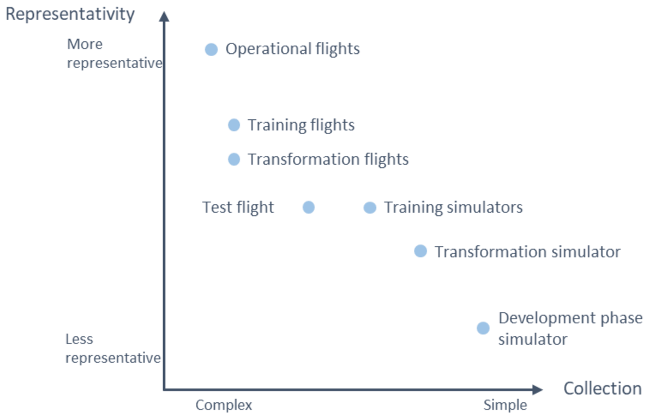

Figure 1: Data representativeness versus collection effort

The core idea of the work is a reading grid for the valorization of cockpit data based on the two axis: different means of collection are used to benefit some of the objectives depending on the strengths and weaknesses of the means for the targeted interests (see Figure 1).

COCKPIT DATA CHALLENGES

Dassault Aviation undertook some studies around cockpit data collection and analysis (on ground and in flight). The following challenges have been emphasized during these projects.

The problem of data valuation and contextualization

Besides basic statistical analysis, the potential of cockpit data for learning purposes has been considered. A task inference model has been derived with the idea that in the end it could be used to provide new adaptive functions to the crew.

Figure 2: Task inference based on learning on cockpit data (small diamonds with black border are predictions while larger borderless are ground truth samples)

The lessons learned from this project are: 1) labellization is a very time-consuming task but mandatory for supervised-learning applications; 2) context is very important for the human analyst to fully understand the situation going on in the flight deck at any given time. Providing this context in an easy, human-understandable format is challenging (video and data cross-referencing).

The problem of cockpit constraints

Cockpits are a collection of embedded systems, they usually are small and busy with knobs and information. Adding systems in such spaces to collect physiological crew data is a challenge. On one side, the sensors must be miniaturized to avoid interfering with the crews' piloting. And on the other hand, it must be very precise to detect information of interest, e.g. which piece of information a crew member is watching.

Gathering ocular data in a fighter's cockpit also proved complicated because of the crew's wearing of a helmet and snout. The helmet prevents the crew from wearing glasses, and the snout prevents a system to recognize and position a face and then eyes in space with usual face recognition algorithms. The combination of these elements makes the gathering of ocular data difficult before the addition of eye trackers inside helmet mounted displays.

The problem of data quality and certification

The use of learning technologies on the data led to the question of the certification of such applications. Authorities are pro-active and are publishing initial guidelines (regularly updated and enriched) to provide domain-wide guidance material for industrials to rely on (European Union Aviation Safety Agency, 2021).

The proposed guidelines introduce new design activities. Noteworthy is the whole learning assurance framework whose purpose is to compensate for the incomplete coverage of learning applications by DO-178. The goal is to guarantee appropriate system operation in the pre-defined operational design domain. The cornerstone of the process is the data pipeline that shall ensure integrity and traceability all along life-cycle: collection, labellization, set repartitions, use for model training, validation and testing.

The problem of personal data ethic

Some cockpit data is considered as personal data since for instance a camera will record pilot's face. As such, the European Union General Data Protection Regulation shall be complied with to make sure the collected data is used accordingly to its intended use (cockpit design improvement).

Initial guidelines mentioned in previous sub-section are quite prescriptive on the 'in operation' data recording capability of machine learning systems. It is justified by explainability requirements (for monitoring and post-accident investigation) but to ensure widespread deployment of such systems it is important to address legitimate crew privacy concerns by providing guarantees. Development of machine learning explainability methods not relying on full input recording is still to be addressed.

FUTURE WORK

Projects described above are still in their early phase. If successful and deemed appropriate, the next step will be the development of specific analysis tools, their integration in the design process: leveraging for design purposes, leveraging to support certification activities and operational leveraging for the improvement of flight safety.

One critical milestone to move onto the next step and scale up is the development of an integrated data flow that merges heterogeneous data coming from a variety of sources and entities: training simulators, development phase simulator, test and operation aircraft. The unity offered by such framework will allow efficient processing and cross-referencing to support analyst task.

REFERENCES

United States Congress. (2020). *Aircraft Safety and Certification Reform Act of 2020*. https://www.congress.gov/bill/116th-congress/senate-bill/3969/text#toc-ida1fb8f529afe4aea9f6edacd72646584.

Bureau d'Enquêtes et d'Analyses pour la sécurité de l'aviation civile. (2012). *Rapport final de l'accident survenu le 1er juin 2009 à l'Airbus A330-203 immatriculé F-GZCP exploité par Air France vol AF 447 Rio de Janeiro – Paris*. https://www.bea.aero/docspa/2009/f-cp090601/pdf/f-cp090601.pdf.

The House Committee on Transportation & Infrastructure. (2020). *Final committee report: The design, development & certification of the Boeing 737 MAX*. https://transportation.house.gov/imo/media/doc/2020.09.15%20FINAL%20737%20MAX%20Report%20for%20Public%20Release.pdf.

European Union Aviation Safety Agency. (2021). *EASA Concept Paper: First usable guidance for Level 1 machine learning applications – Issue 01.* https://www.easa.europa.eu/downloads/134357/en