INDIVIDUALIZING FLIGHT SKILL TRAINING USING SIMULATOR DATA ANALYSIS AND BIOFEEDBACK

J.O. Entzinger[*] , T. Uemura[*] , S. Suzuki[*]
[*]The University of Tokyo, School of Engineering, Department of Aeronautics & Astronautics
7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-8656, Japan. Tel. +81-3-5841-6632, Fax. +81-3-5841-8560

Keywords: *Flight training, Human pilot, Manual flight skill, Flight Simulator, Psychophysiological data*

Abstract

We propose tools and methods to analyze and interpret both simulator and psychophysiological data in order to support pilot trainees and instructors and improve manual flight control training effectiveness.

Current pilot training uses a fixed curriculum, with all trainees proceeding at the same pace. There is little guarantee that they really develop the skills that are being taught, until the subjective evaluation by instructors. In the "Pilot's Individualized Learning using Objective Data" (PILOD) project we provide methods to individualize the training program based on the feedback of objective data to the trainees during their training.

In this paper we discuss experiments where we trained university students in a fixed base Boeing 747-400 simulator. From a large number of initial trainings we developed analysis methods and evaluation metrics, as well as training recommendations and specific practice exercises. We applied these in a follow-up experiment with 10 students who received 6 hours of training each. Eye, heart, and brain wave data were recorded, but at this moment only the analysis of heart data is sufficiently robust and meaningful for practical use.

The results show that the analysis of simulator and psychophysiological data can be a useful addition to subjective observation by an instructor in the evaluation of manual flying skills during flight training and the creation of personalized training programs. The results also showed the effectiveness of the training in general, and of the specific flare and longitudinal control training modules in particular.

1 Introduction

1.1 Background

Developments in the aviation market have led to 3 big challenges in pilot training:

Quantity: There is a pilot shortage that will only get worse due to the rapid globalization and increased wealth, especially in Asia [1, 2].

Quality: With the increased use of cockpit automation, pilots' manual control skills have degraded [3]. Manual aircraft control has become a factor in the majority of fatal accidents.

Cost: More efficient pilot training will be needed to assure safety in the increasingly competitive market where Low Cost Carriers (LCCs) and government-supported airlines from the Middle-East put pressure on airlines and pilots to cut costs.

On the other hand, the following developments create opportunities for pilot training:

Better understanding of human behavior: Interdisciplinary research between the fields of engineering, medicine, and psychology is blooming because both simulators and instruments to measure eye movements, heart rhythms, brain waves, etc. have become cheaper and easier to use.

Data-driven society: With the increased availability of data, and our increased ability to analyze such data, objective decision making and personalization have become easier than ever.

Educational innovation: The aviation industry and regulators world-wide have started pushing a modernization of training methods. The Advanced Qualification Program (AQP) [4] and Evidence Based Training (EBT) [5] can be seen as implementations of problem/project based learning (PBL), continuous assessment, and the systematic definition of learning objectives that are already mainstream in regular education at schools and universities around the world. The next step would be the 'individualized curriculum'.

1.2 The "Pilot's Individualized Learning using Objective Data" (PILOD) Project

In the PILOD project we develop tools to identify the needs of individual pilot trainees, and to support the individualization of training curricula. It builds builds on research in control, aviation psychology, and data analysis to move Evidence Based Training from the fleet level to the personal level (Fig.1). The aim is to improve the efficiency and effectiveness of pilot training in order to mitigate the challenges mentioned in the previous section.

Flying is not only about keeping the aircraft within safe operational limits, but also doing so in a robust and systematic way, with sufficient spare capacity to deal with unforeseen events. Therefore we propose to incorporate all three levels in the analysis — pilot (resource), control (process), and performance (output). Beside flight (simulator) data, we also use bio data to analyze the trainee's visual workload, mental effort, and stress. This could provide a valuable addition to the subjective information an instructor can obtain through observation. The analysis outcomes can support the instructor an recommend appropriate practice exercises on a challenging level (Fig.2). This means some trainees will get exercises to specifically improve their personal weak points, while others may get integrated tasks or move ahead slightly quicker (they may need additional practice on different points in the future).

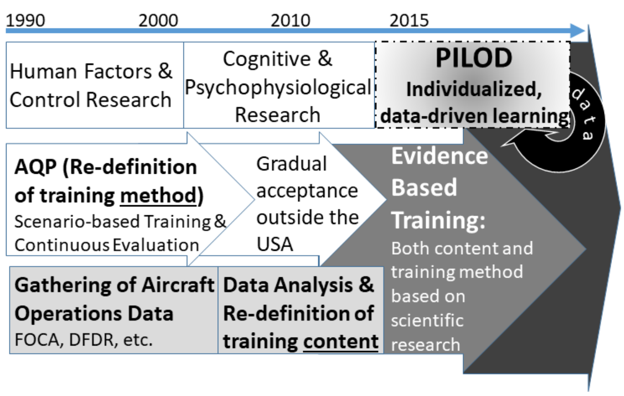


Fig. 1 Developments in pilot training.

1.3 Workflow

To develop a tool that can support continuous assessment and individualized training, we need to

1. find a practical means to acquire data,
   (a) cost-effective measurement devices
   (b) ease-of-use (by non-researchers)
2. identify *meaningful* features in the data,
3. automate the extraction and analysis of the selected features,
4. perform an evaluation or classification of the analysis results,

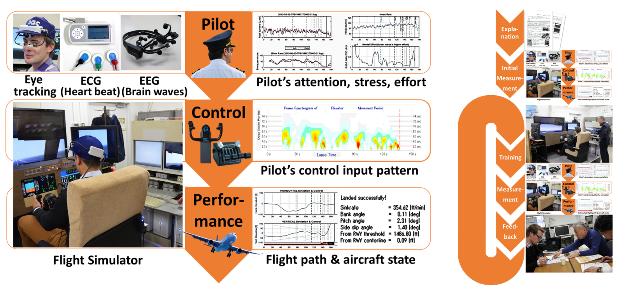


Fig. 2 Left: Integral evaluation of the trainee's psychophysiological state, his control actions, and the resulting flight. Right: PILOD starts with a briefing and initial data collection, followed by a cycle of personalized training, data recording, and feedback.

   (a) build a base or reference data
   (b) identify what is 'good' performance or 'desirable' behavior (i.e., 'metrics')
5. suggest specific actions based on the evaluations.
   (a) create a pool of exercises to train various skills on various levels
   (b) verify that the exercises have the desired effect on the trainee's skills

Obviously, the '*meaningful*' in item 2 strongly depends on what is thought to be desirable (item 4b) and how it relates to training possibilities (item 5). Identifying these points is the major challenge in this research. Another challenge is developing automated feature extraction methods that are sufficiently reliable and robust to cope with the relatively low quality data that can be expected under the restriction arising from item 1.

Figure 3 illustrates the general idea of the skill level classification and action recommendation.

2 Materials & Methods

2.1 Simulator

For the experiments we used the fixed-base Boeing 747-400 simulator at The University of Tokyo (Fig.4). Custom software is used to simulate the dynamics, which is based on publicly available models but has been extended and tuned with the help of professional pilots. The simulator states are logged at 20 Hz.

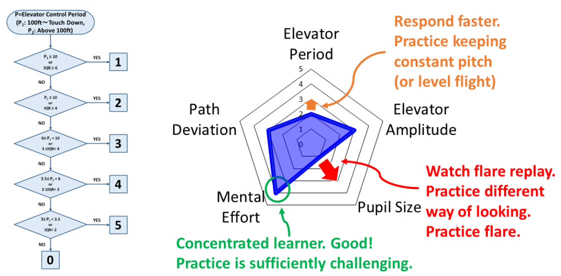


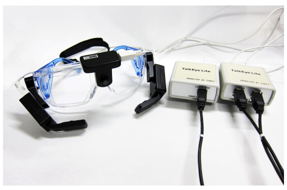
Fig. 3 Left: skill level classification example for elevator control style. Right: results can be used to individualize training to balance the trainee's skill set.

2.2 Other measurement devices

We recorded eye-data (gaze direction, saccade speeds, blinks, and pupil diameters) at 30 Hz using the Takei TalkEyeLite and electrocardiograms (ECG) at 250 Hz using the ParamaTech EP-301. In some experiments we additionally recorded brainwave data (electroencephalogram, EEG) using the eMotiv EPOC+ at 128 Hz. The devices are shown in Fig. 5.

All recordings were automatically synchronized using custom built, serial port based trigger system, activated by the beep of the ECG recorder (which does not have external trigger functionality built in).

Fig. 4 The fixed-base flight simulator at The University of Tokyo



(a) The Takei TalkEyeLite eye-mark recorder

(b) Eye-tracking screenshot

(c) The ParamaTech EP-301 portable ECG recorder

(d) The eMotiv EPOC+ brainwave recorder

Fig. 5 Measurement devices used in the experiments

2.3 Initial Training Experiments

The analysis methods and evaluation criteria were developed from our experience with and data from the basic flight (simulator) training of ca. 40 university students, some of who also completed an intensive follow-up training. The gathered data were automatically analyzed after each training session and used to provide the trainees feedback on their performance and psychophysiological state, as well as to give them advice such as what skill to focus on in the next session. The interpretation was done by the authors, and typical patterns in the graphed data were explained to the participants so that over the training course they developed the ability to self-evaluate using the analysis result graphs.

During these experimental training sessions we tried out various explanation and feedback methods and training practice exercises. Together with data from a few private pilots as well as a few active and retired airline pilots, this data was used to develop the analysis methods, to create a baseline of typical progress, and to inventory common training difficulties.

2.4 PILOD-based Training Experiments

The main experiments discussed in this paper consisted of the training of 10 male university students aged between 21 and 23. All participants were right-handed and had normal or corrected to normal vision. None of the subjects had any experience flying an actual plane, nor more than 1 hour experience in the simulator used in the experiment. One subject had about 30 hours hang-glider experience.

All subjects received a short training syllabus and 30 minute individual explanation about the aircraft operation (controls, cockpit displays, out-the-window visual cues/aids, etc.) and 6 one hour training sessions (on 6 different days). In the 1st, 4th, and 6th session we recorded data for analysis and the remaining time was spent on practice, the other sessions were practice only. About half of the practice time followed a predetermined curriculum shared by all subjects, the other half was individualized based on the analysis results.

Before the first data recording, subjects got a little time to familiarize with the simulator, watched a landing replay, and flew a straight-and-level flight for 3 minutes and a landing from 1800 ft (550 m) altitude. After that we recorded simulator data, eye data, ECG, and EEG for 2 landing replays (including 'relax' time after touchdown), 2 landings in good visual conditions, and 2 landings in bad visual conditions (1600 m visibility). We also recorded simulator data only for a straight-and-level flight. In all scenarios there was moderate turbulence.

Individual practice exercises were assigned based on the analysis results from the first day measurements and could be one or two of the following:

- High Altitude[1] Longitudinal Control
- High Altitude Lateral Control
- Scanning (i.e., systematically checking each cockpit display/indicator)
- Aiming (i.e., focusing on the movement of the landing zone markings on the runway to decide one's low altitude control)
- Flare (i.e., the pitch-up maneuver a few seconds before touchdown to reduce sink rate and land on the main gear)
- Integral landing practice (i.e., no specific practice)

The experiment protocol was approved in advance by the ethics committee of The University of Tokyo's School of Engineering. Each subject provided written informed consent before participating.

3 Results

3.1 Initial Training Experiments

The initial training experiments have resulted in a large set of data analysis tools, knowledge about

[1] 'High Altitude' in this paper refers to altitudes above ca. 300∼500 ft (100∼150 m), where the cockpit instruments are the pilot's main source of information, as opposed to lower altitudes where the outside visual scene with runway geometry provides more salient cues.

the interpretation of various data, and special practice exercises. Several of the analysis methods have been reported earlier [6, 7]). We added a number of new features to our simulator in order to record additional parameters and facilitate new training exercises, for example freezing motion in a specific direction so that the trainee can focus on one subtask at a time, and a warning buzzer when a flight parameter deviates too much from its desired value to improve scanning. A number of new software tools were developed, such as a simple secondary task tool [8].

In the experiments we presented the trainees with various graphs of their psychophysiological state, control style, and performance and showed them what to aim for. We noticed that this visualization of the flight analysis can help to explain problems or desired behavior. It also help to convince the trainee, as it provides some evidence with the instructor's subjective comments.

We carried out a meta-analysis of the data gathered in the initial training experiments, trying to develop a 5-level skill level classification scheme from beginner to veteran for a large number of metrics. Large progress was made in the automatic extraction of numeric indices that reflect features we had thus far subjectively identified from graphs when providing feedback to trainees. However, a few challenges remain on the way to an automated evaluation and recommendation system, and some expert interpretation, contextual awareness, and cross-verification with the original graphs and (subjective) cockpit observations will probably always remain necessary to some extent.

3.2 PILOD-based Training Experiments

To evaluate the trainee's control input we make use of spectrogram analysis. Figure 6 shows an example. The top image shows a typical beginner's pattern, with almost no control, except for some slow (long period) control at low altitude, short before touchdown. With increasing experience, we see more, stronger, and faster control. In the bottom figure we see the control is concentrating at the shorter periods, as the pilot notices deviations quicker and his response becomes more adequate (less over-control).

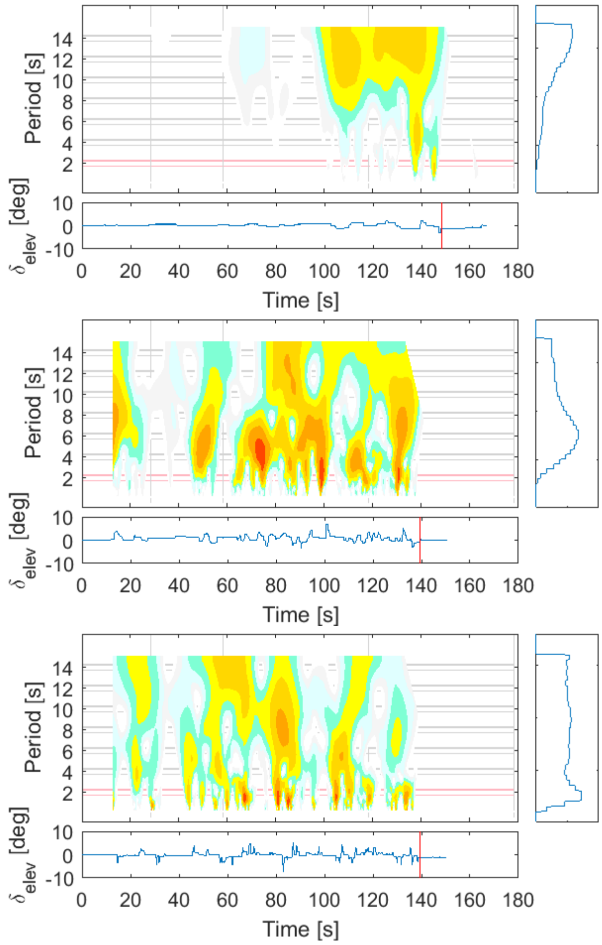


Fig. 6 Analysis of elevator control in a 'good visibility' flight the 1st, 4th and 6th training session (top to bottom). Each image shows a spectrogram analysis of the control input (top-left), the distribution of control input periods (right) and the elevator deflection command (bottom). The vertical red lines indicate the moment of touchdown.

Figure 7 shows how the heart rate of one of the subject changed during the bad visibility landings. It is interesting to mention that this subject showed almost no progress on any of the control or performance metrics, and some even got worse. Looking at the heart rate patterns however, we see some clear changes. On the first day there is no difference between the 'Not Flying' (watching replay / relaxing) case and the Flight cases: the subject did not know what to do and did not feel engaged. On day 4 we see a rapid

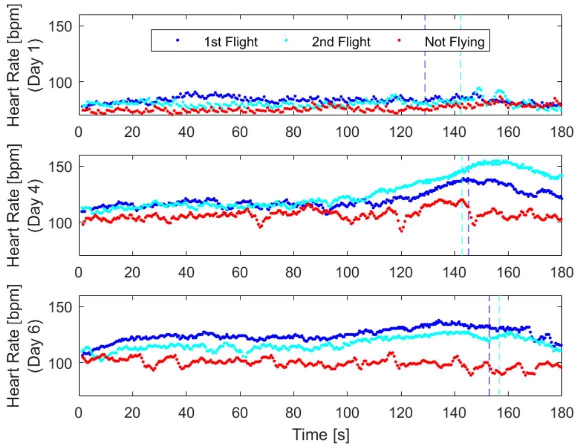


Fig. 7 Heart rates in the 'bad visibility' landings and a landing replay of the 1st, 4th and 6th training session (top to bottom). The dashed vertical lines indicate the moment of touchdown.

increase of the heart rate from ca. 20-30 s before touchdown, when the runway becomes visible through the fog: the subject didn't feel he could make a difference in instrument flight, but knows what needs to be done when seeing the runway. On day 6, we see more engagement earlier on, and a less distinct stress peak just before the landing: he starts to understand the cockpit instruments. The subject probably made progress in understanding, without (yet) being able to turn that into effective actions. After the experiments the subject requested some more training, and his control and performance improved steadily. He appeared to be a late-bloomer, something that would have been impossible to see from cockpit observations only.

Individual differences and the difficulty of robustly de-blinking data made it difficult to extract practical metrics from the eye tracking data. For several subjects the eye tracker could not consistently detect the pupil for unclear reasons.

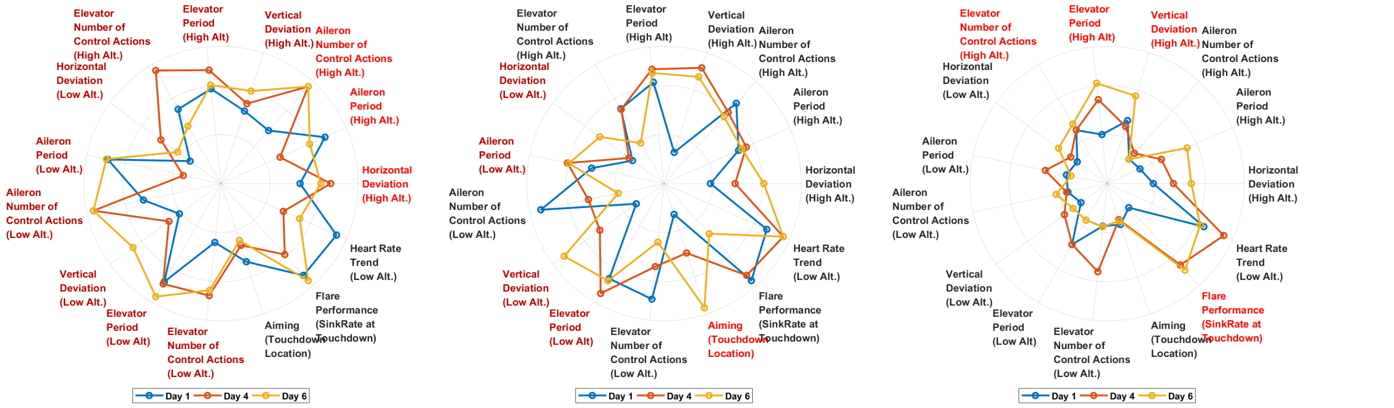
We are still in the process of developing EEG based metrics. A spectral band power analysis [9] and Approximate Entropy [10] did not provide useful results. It is still unclear if this is due to the low quality of the EEG data or that these metrics are not sensitive enough for our application (with a small number of trials per subject, limited measurement time, and a multidimensional and time-varying task).

The progress of 3 of the subjects is shown in Figure 8. The subject of Figure 8(a) received specific 'High Altitude Lateral Control' training, so we expect his horizontal deviation, aileron control period, and number of aileron control actions to improve. It is interesting to note that the aileron control period may seem to get worse, but this is due to the fact that on Day 1 the subject made only very few control actions, which happened to be short period, resulting in a good evaluation on this metric. In short, if the number of control actions is too low, their dominant period becomes meaningless. The subject also received scanning training, which can be expected to result in an overall improvement of all variables except aiming, flare, and heart rate. The heart rate trend getting worse at first and then recovering somewhat indicates a pattern similar to that of the subject shown in Figure 7.

Figure 9 shows a comparison between groups of subjects who received a particular specific training and those who didn't. Subjects who received the integral landing practice (i.e., no specific practice) and in the case of Fig. 9(b) also those who received scanning practice were omitted from the analysis, since they would fit equally well in either group. It can be seen that the groups receiving special practice clearly improved, while the groups that didn't receive such practice didn't improve. For the other specific trainings the results were less distinct. This may be due to the small sample size, lesser sensitivity of the metrics, ineffectiveness of the training exercises, higher difficulty of the task (e.g., lateral control takes more time to master than longitudinal control, so a few hours may not suffice), or a combination of these.

4 Discussion

A remaining issue is that it is often very difficult to define what is 'good' or 'appropriate' control. In particular when we also consider the wide variety of pilot tasks and objectives, including ill defined concepts such as passenger comfort and his own workload, it is not always clear what deserves high priority and what performance level

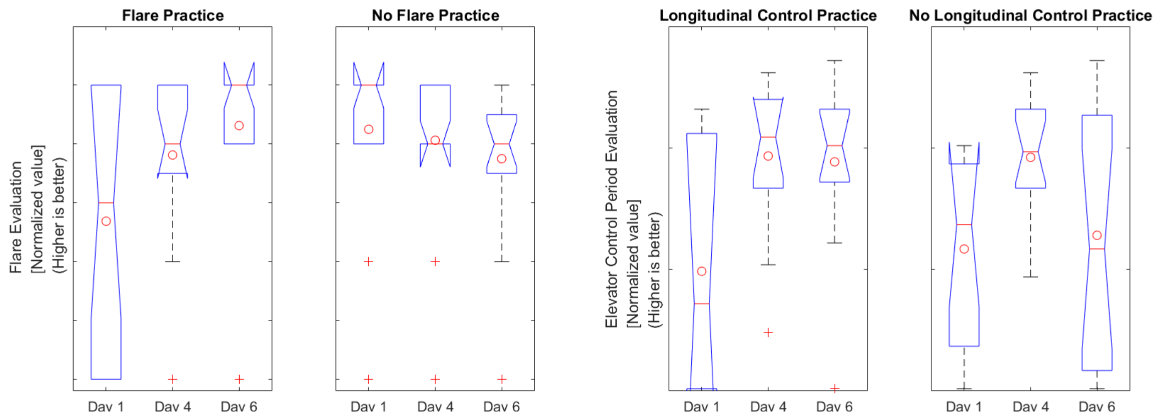


(a) Subject received special 'High Altitude Lateral Control' and 'Scanning' training

(b) Subject received special 'Aiming' training

(c) Subject received special 'High Altitude Longitudinal Control' and 'Flare' training

Fig. 8 Progress of 3 subjects on various evaluation metrics. Metrics expected to improve due to specific training are highlighted red; secondary effects in dark red.



(a) (b)

Fig. 9 Effect of special training. Boxes span from the first to the third quartile, + indicates an outlier value, red lines indicate the medians (excluding outliers), red circles indicate the means (including outliers).

is acceptable in the given situation. Unlike typical automatic feedback control systems which aim at driving the error signal to zero, a pilot will generally be indifferent to the error value as long as it is within some (soft) limits, and those limits may change over time with the flight phase or the pilot's workload.

Another issue is that sometimes veteran pilots and beginners show very similar behavior, while those in an intermediate phase act differently. This makes it impossible to automate evaluation without contextual awareness (e.g., number of training hours so far). A simple example would be the elevator control effort. A beginner who doesn't know what to do may rely on the plane's natural stability and have small path deviations without controlling anything at all. An intermediate trainee who tries to correct small deviations may over-control and stay busy correcting his/her self-caused deviations. The veteran notices deviations when they are still small and can correct them quickly and without much effort. Both in control effort and in performance, the beginner and veteran are very similar. This shows the risk of reducing complex data to a few simple indices.

The analysis of bio data poses some challenges due to the large interpersonal differences, and even differences between days for the same person. Normalization and standardization with reference to the data recorded while watching the replay and relaxing can help only to some extent. Some subjects will easily get excited or stressed and show an increasing heart rate, while others will always stay 'cool'. This may however be less of an issue in a real training environment, as the group of trainees has been homogenized by aptitude tests and medical examinations during the selection process.

Finally, it is hard to obtain reliable reference data. Although we accumulated data from many student trainees, it is hard to find advanced and veteran pilots who want to and are allowed to cooperate, particularly when it comes to recording bio data. An additional problem is that age difference between veterans and trainees may affect bio data (e.g., lower and more constant heart rates, slower eye movements). At the same time, veterans may have a accumulated rich practical experience, but are not necessarily better than early-career pilots [11] (cf. a veteran athlete becomes a trainer, but the early-career athlete wins the Olympic).

It should be noted that there are some subtle but important difference between evaluating pilots and evaluating pilot trainees. For example, whereas pilots need to be trained to a level where they have sufficient spare capacity to deal with unforeseen events, trainees need to be sufficiently challenged (without being overloaded) in order to learn efficiently. This has important implications for the interpretation of metrics using for example heart rate and heart rate variability. Another difference is that experienced pilots may have developed their own ways of doing things. These alternative methods may be equally valid (or even better suited to the individual), but may not be evaluated as such by an automated system, and maybe not even by a peer with even different preferences. For trainees however, it is important to acquire a basic set of skills to build on, so it will be appropriate to define a set standard for them.

5 Summary and Future Works

We proposed a method for the automated analysis of a variety of data collected during (simulated) practice flights, which can support the instructor in deciding which skills to focus on for each individual trainee. Such a system could even help a trainee to self-evaluate and perfect his/her skill to a specific level before requesting new instruction, so that the limited contact-time can be used more effectively. In addition, some issues may be hard to detect through 'subjective' observation by an instructor (e.g., effort, stress, workload, information use, situational awareness, control style/strategy), but may be possible to evaluate to some extent using measured 'objective' data. We showed examples of heart rate analysis and spectrogram analysis of elevator control that illustrate this.

We applied the PILOD method in an experiment where we trained 10 subjects very systematically. The metrics based on flight technical performance measures, control style, and ECG were shown to be sufficiently sensitive and robust to use for evaluation. Metrics based on eye tracker data and brain waves will need to be investigated further. The results also showed the effectiveness of the training in general, and of the specific flare and longitudinal control training modules in particular.

Additional effort will be needed to improve the robustness of several bio-feedback metrics, and to strengthen the base of reference data from advanced trainees and veterans. We also plan to validate the effectiveness of a number of special practice exercises in independent experiments, that is, without the complete flight basic training.

Acknowledgements

This research is financially supported by JSPS KAKENHI Grant Number JP16K21002. The authors like to thank Mr. Naoki Sekiguchi and Mr. Takuto Fukazawa for their contributions to the development of data analysis methods and all subjects who participated in the experiments for their time and comments.

References

[1] Boeing Commercial Airplanes Market Analysis. *Current Market Outlook 2016-2035*. Seattle, WA, USA, 2016.

[2] Airbus. *Global Market Forecast Growing Horizons 2017/2036*. Airbus, Blagnac Cedex, France, 2017.

[3] A. Haslbeck, E. Schubert, L. Onnasch, G. Hüttig, H. Bubb, and K. Bengler. "Manual Flying Skills Under the Influence of Performance Shaping Factors". *Work*, vol. 41, no. Suppl. 1, pp. 178–183, 2012.

[4] D. Farrow. "The Advanced Qualification Program: FAA Lessons Learned from 20 Years of AQP". In *ICAO NGAP Symposium*, Montreal, Canada, 2010.

[5] *Evidence-Based Training Implementation Guide*. International Air Transport Association (IATA), Montreal, Quebec, Canada, 1st edition, July 2013.

[6] J.O. Entzinger, T. Uemura, and S. Suzuki. "Mental Effort and Safety in Curved Approaches". In *29th Congress of the International Council of the Aeronautical Sciences (ICAS 2014)*, St.Petersburg, Russia, 2014.

[7] J.O. Entzinger, S. Kamo, T. Uemura, and S. Suzuki. "Objective Evaluation of Hand-Flying Skills and Styles in the Final Approach". In *The 2015 Asia-Pacific International Symposium on Aerospace Technology (APISAT 2015)*, Cairns, Australia, 25-27 Nov. 2015.

[8] J.O. Entzinger, T. Uemura, and S. Suzuki. "On the Use of Secondary Tasks in Addition to Other Objective Measures of Pilot Workload". In *The 2016 Asia-Pacific International Symposium on Aerospace Technology (APISAT 2016)*, Toyama, Japan, 25-27 Oct. 2016.

[9] G. Borghini, P. Aricò, I. Graziani, S. Salinari, Y. Sun, F. Taya, A. Bezerianos, N.V. Thakor, and F. Babiloni. "quantitative Assessment of the Training Improvement in a Motor-Cognitive Task by Using Eeg, Ecg and Eog Signals". *Brain Topography*, vol. 29, no. 1, pp. 149–161, 2016. ISSN 15736792. doi: 10.1007/s10548-015-0425-7.

[10] X. Li, Y. Jiang, J. Hong, Y. Dong, and L. Yao. "Estimation of Cognitive Workload by Approximate Entropy of Eeg". *Journal of Mechanics in Medicine and Biology*, vol. 16, no. 06, 2016. doi: 10.1142/S0219519416500779.

[11] M.A. Todd and M.J.W. Thomas. "flight Hours and Flight Crew Performance in Commercial Aviation". *Aviation, Space, and Environmental Medicine*, vol. 83, no. 8, pp. 776–782, aug 2012. ISSN 00956562. doi: 10.3357/ASEM.3271.2012.

6 Contact Author Email Address

Correspondence about this paper should be directed to the first author at jorg@entzinger.nl.

Copyright Statement