Assessment of Human Pilot Mental Workload in Curved Approaches*

By Jorg Onno ENTZINGER[1], Marijn NIJENHUIS[2], Tsuneharu UEMURA[1], and Shinji SUZUKI[1]

[1] *School of Engineering, The University of Tokyo, Tokyo, Japan*
[2] *Faculty of Engineering Technology, The University of Twente, Enschede, The Netherlands*

Which cognitive challenges do human pilots face during the execution of curved (RNP-AR) approaches? We hypothesize that the mental model required for a curved approach will be more complex than for a straight one. To investigate this, we compare risk and mental effort through physiological factors, control-input, and performance. Our current experiments focus on straight landing approaches under different visibility conditions and we compare the long and short final to establish which methods can be used to analyze mental effort and safety. Both student and professional pilots took part in our fixed base B747-400 simulator experiments. The control-input, performance, and ecg analyses appear to be particularly useful, whereas blink and especially pupil diameter data obtained from an eye camera is more difficult to use in analysis. To safely implement RNP-AR, we need to further investigate the necessities of cognitive skill training.

Key Words: RNP-AR Curved Approaches, Flight Safety, Workload, Cognitive Models, Human Factors

1. Introduction

Our aim is to understand which cognitive challenges human pilots face during the execution of curved approaches flown under Required Navigation Performance Authorization Required (RNP-AR) procedures.

From a literature review and interviews with researchers and pilots who actually fly RNP-AR approaches on a regular basis, we know that various safety issues arise during actual operations. Although path design and cockpit automation for RNP-AR approaches have received much attention and are quite well established, descriptions of operational human factors issues in this complex environment are mostly anecdotal. This is in part due to the high reliance on automation, and due to the fact that RNP-AR approaches are currently mostly carried out under "ideal" conditions (not during peak-times at airports, in good weather conditions, etc.). For RNP-AR to be successful and to guarantee safety in the future, we will, however, have to investigate rare-event cases and particular necessities in (cognitive) pilot training.

Even though RNP-AR operations are generally carried out relying heavily on cockpit automation, the human pilot still has the final responsibility, and should at any time be able to intervene. He therefore needs sharp skills to verify the proper operation of the on-board automation, and be actively involved so that he can smoothly take over the control from the autopilot if needed. We believe this will inherently be more difficult in curved approaches than in straight approaches, and therefore we defined the two main objectives of this research:

1. To understand differences in pilots' mental models and cognitive processes between curved and straight approaches
2. To find out how best to support the pilot (through training or interfaces) in his supervision of automation and in decision making.

In this paper, we report on the first, explorative phase of this 3-year research project. We focus on the development of a set of tools and techniques to measure and analyze mental effort and risk to compare various approach types.

2. Risk Analysis

Particularly in highly automated cockpits, the human pilot's main task is to verify whether the autopilot controls the aircraft appropriately, and to initiate a go-around when this question cannot be answered positively. The need for a go-around is (luckily) a rare-case event, with the drawback that pilots have relatively little experience of dealing which such cases. In addition, pilots may be susceptible to plan continuation error (PCE), and stick to the originally planned landing operation, rather than intervening and rescheduling when necessary. This may be due to automation complacency, but Causse et. al[1] add that a contributing factor to PCE may be that *"an airline that emphasizes productivity (e.g. on time arrivals or saving fuel) may unconsciously set up conflicts with safety. Pilots may be willing to take a risk with safety (a possible loss) to arrive on time (a sure benefit)."* With the increased competition in the aviation sector, this forms a growing risk.

In this research, we will limit our scope of risk to

*Peer-reviewed by the international program committee of APISAT-2013

those risks related to "human error". We will not investigate the probability that automation errors or failures occur, but treat them as a given fact, and only include the risk that the human pilot may not notice the error or failure, or may not be able to deal with it appropriately.

2.1 Measuring Risk and Mental Effort

"The most objective measure of danger [...] is time until the aircraft is destroyed if control action is not taken.[2)] *"* Such a `time-to-crash (TTC)' is maybe a bit crude, but highly intuitive performance measure. It is closely related to the concept of `stabilized approach' approach, which states that at a certain minimum height the flight parameters (in particular aircraft track, flight path angle, and airspeed) do not exceed established criteria, published in the airline's standard operating procedures (SOPs).

The pilot's situational awareness, capacity, and actions during the remaining TTC largely determine the recoverability, and should therefore be included in the risk analysis as well. In figure 1 we analyse the effects of a "difficult" operation compared to a standard operation,

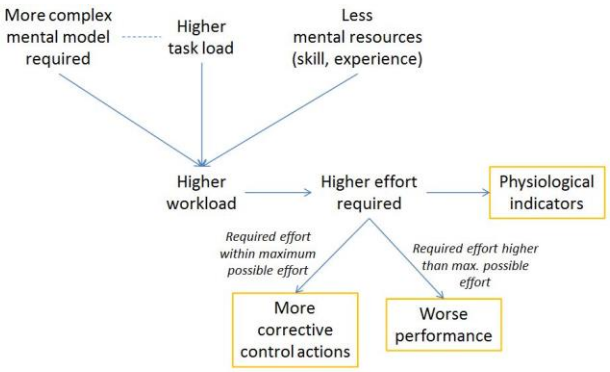

Fig. 1. Hypothesis tree.

In the closed control loop of pilot and aircraft, there are 3 places where we can gather data for the analysis of operational risk. The first is the pilot himself. We can ask him for subjective ratings about his awareness or workload, or measure related physiological reactions (e.g., heart rhythms, pupil dilatation). The second is an analysis of the pilot's control actions. The particular actions taken or the pilot's control style may be related to his mental state (awareness, effort). The third would be the final task performance. If the targets are not met, this would indicate that the workload was too high.

2.2 Time-to-Crash Analysis

The time-to-crash (TTC) can be determined at any point during a flight by running a simulation with the current aircraft state as initial parameters and removing any control inputs. The safest flight would be the one where the area under the TTC graph is highest. However, since this would mean no landing, we define ΔTTC, the TTC index with respect to the ideal stabilized approach.

Within the TTC we can identify two different deviations from the ideal TTC. One is a prolonged deviation (often as a result from an inadequate thrust setting), indicating that the pilot might not notice the problem or may not give priority to solving it. The other is a short deviation, because quickly corrected by the pilot, but with possibly larger consequences in a high-workload situation.

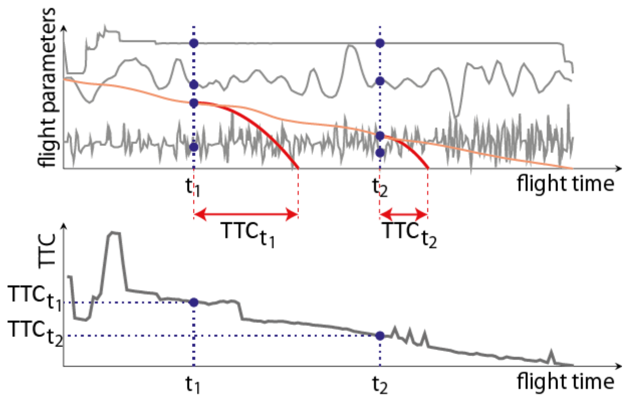

Fig. 2. Creating a time-to-crash graph (bottom) by calculating the new altitude profile (red) for the case that the flight would be resumed without any control inputs.

For these 2 types of risk, we define the following TTC indices:

$$f_{\text{Prolonged}}(\Delta TTC) = \frac{1}{t_d}\int_0^{t_d} \Delta TTC(t)dt \tag{1}$$

$$f_{\text{Temporal}}(\Delta TTC) = \frac{1}{n}\sum_{k=0}^{n}|D(k)| \tag{2}$$

$$f(\Delta TTC) = C_1\frac{1}{t_d}\int_0^{t_d} \Delta TTC(t)dt + C_2\frac{1}{n}\sum_{k=0}^{n}|D(k)| \tag{3}$$

with D(k) the dimensionless level 1 detail coefficients of the discrete wavelet transform of ΔTTC using Daubechies 2 wavelets, and C_1 and C_2 constants equal to the inverse of the values of eqs. (1) and (2) respectively, averaged over all analysed flights.

2.3 Analysis of the Control Input Signal

During a landing, maintaining the 3 degree glide path is the major manual control task. The pitch angle, and therefore the elevator input command is therefore the most interesting to analyze. Although other controls are constantly being monitored, pressures on the rudder and ailerons or changes to the thrust settings are only applied occasionally. In line with this, inspection of the simulator time histories showed that the elevator deflection, as imposed by the pilot through the stick, is the control input that contains the most information on control actions inflicted on the airplane.

The main idea is that more control inputs mean a higher effort by the pilot. Therefore, the power of the elevator control signal is calculated. Plots of the power spectral density also appear to be a helpful tool for the interpretation of the control input data.

2.4 Heart Rate Variability Analysis

Heart rate (HR) and heart rate variability (HRV) are related to stress and effort. In particular the HRV power spectrum band from 0.06 to 0.14Hz is said to be suppressed in cases of high mental effort[3-4]. We used a ParamaTech EP-301 portable electrocardiogram (ECG) recorder with external electrodes to record ECGs at 250Hz (fig. 3). After various trials, we found that electrode placement on the left hand (control column), right neck, and right foot minimized noise due to the pilot's motions and gave clear, sharp R-wave peaks.

Fig. 3. The portable ECG recorder and electrode placement used when operating our fixed base flight simulator.

Heart beats (R-peaks) and inter-beat intervals (R-R intervals) were identified using the open source ecgBag software[5]. The automatically identified and filtered peaks were visually inspected and manually edited, as advised by cardiologic societies[6]. The PSD was calculated from the inter-beat intervals (ibi) after resampling at 8Hz to create an input with equidistant samples, as is common practice for the determination of the HRV.

2.5 Analysis of Blinks and Pupil Diameter

Blink rate and pupil diameter also depend on mental effort. They can both be measured using an eye camera such as the NAC EMR-8 (fig. 4) that we used in some of our experiments. Blink rate is highly suppressed duding high-demand visual tasks such as the flare maneuver,[7] and the pupil diameter changes not only with the brightness of the environment, but also with mental effort.[8]

Fig. 4. The flight simulator and the NAC EMR-8 eye mark recorder. On the monitor on the left, a thresholded image of the subject's pupil is visible.

3 Experiment Setup

We carried out a series of experiments to establish the usefulness of the analysis techniques described in the previous section. Since (partially automated or guided) curved approaches cannot be flown with our current simulator, and since a more basic first experiment seemed appropriate, we flew straight approaches and compared various factors, as described later in this section.

3.1 Subjects

Simulated flight data from three professional pilots —named PP1, PP2 and PP3— and two student pilots — named SP1 and SP2 — is available. All have agreed to the anonymous analysis of their data. The professional pilots are (current or retired) captains from All Nippon Airways or Japan Airlines. Among the group of student pilots that received some kind of basic flight training in the simulator, SP1 and SP2 were chosen because their training was by far the most extensive. Their flights were conducted at the end of three-months of twice-weekly training, when the instructing (ex-professional) pilot judged their performance to be adequate and, in this specific condition and flight procedure, on par with that of professional pilots.

3.2 Instruments

Flights are conducted in the fixed-base no-force feedback simulator of a Boeing 747-400, represented in figures 3 and 4. The aircraft's dynamic response is calculated by software developed in-house. The graphics are provided by Microsoft Flight Simulator. Cockpit sound is simulated; the pilot can adjust the volume to a comfortable level. A flight director for ILS approaches is available. Next to the runway a simulated PAPI visual aid is present. Different weather conditions can be created by setting a wind turbulence and cloud ceiling level. Flight data is recorded at 20Hz for all controls and aircraft response parameters.

In the experiments where ECGs were recorded, the above-mentioned ParamaTech EP-301 was used. For the eye camera recordings, the NAC EMR-8 was mostly used, although for a few flights the similar NAC EMR-9 was used.

3.3 Flight procedure

The simulation starts with the aircraft headed for a landing on runway 34R of Haneda Airport in Tokyo. The initial altitude is 1800 feet AGL and initial indicated airspeed is 150 knots. The flight director indicates that the aircraft is on the ILS glideslope and localizer. The elevator is nearly fully trimmed (the aileron and rudder are not equipped with trim tabs). Each pilot performs two of the described flights in visual meteorological conditions (VMC; good visibility) with light turbulence, followed by a short rest with a duration to the pilot's liking. Then, two flights are conducted in instrument meteorological conditions (IMC, 800m visibility) with the same turbulence setting. The visibility in IMC is such that the approach lights of the runway just become visible at 500 feet above ground level, when the aircraft is inbound on the ILS flight path.

The pilots are asked to perform the task to the best of their ability. Prior to the first recorded flight, the pilots have all had the opportunity to practice with at least one extra flight in VMC. Two pilots have performed the experiment, consisting of four flights, a second time on a different day.

3.4 Data Analysis

In our analysis we compare the two weather conditions (VMC vs. IMC) and two phases of the approach, the long and short final. Since the long and short final segments of the approach are not well-defined in terms of a flight parameter such as altitude, we choose the end of the long final and beginning of short final to coincide with the time at which the pilot in IMC will transition to using visual references for the landing; this transition marks a significant change in flying strategy. This moment occurs some seconds after the runway has first come into sight under IMC. With the IMC weather conditions as described, the runway becomes visible at about 500feet above ground level for an aircraft on a stabilized track and, in general, 40 to 35 seconds before touchdown. We have decided to call the last 35 seconds of the flight the “short final”, and everything before that the “long final”.

4 Results

4.1 Time-to-Crash Analysis

The TTC indices given in equations (1) and (2) have been applied to the short final (last 35 flight seconds) and long final (until the last 35 flight seconds) segment of all 14 VMC and 14 IMC approaches. The results are given in tables 1 and 2, respectively.

Table 1. Prolonged time-to-crash index for long and short final, per pilot and weather condition. (In seconds).

	SP1, day 1		SP1, day 2		SP2		PP1, day 1		PP1, day 2		PP2		PP3	
	long	short	long	short	long	short	long	short	long	short	long	short	long	short
VMC	8.05	1.68	7.01	3.36	8.22	3.70	9.52	4.81	11.88	6.48	14.08	5.71	10.84	5.19
	4.88	3.73	11.46	2.39	12.69	5.01	9.50	2.77	10.08	4.72	14.17	7.06	7.58	4.05
IMC	11.63	5.67	11.53	7.73	7.94	1.80	10.86	2.80	10.86	4.31	17.28	5.16	19.46	8.06
	11.84	6.17	10.73	6.94	9.20	3.82	9.18	1.44	9.59	3.33	8.91	2.14	14.29	3.73

Table 2. Temporary time-to-crash index for long and short final, per pilot and weather condition. (Dimensionless)

	SP1, day 1		SP1, day 2		SP2		PP1, day 1		PP1, day 2		PP2		PP3	
	long	short	long	short	long	short	long	short	long	short	long	short	long	short
VMC	0.34	0.66	0.90	0.61	0.74	0.81	1.08	0.30	0.21	0.80	0.77	0.56	0.77	0.76
	0.69	1.51	0.79	2.25	0.46	0.27	0.39	0.53	0.47	0.85	0.37	0.41	0.35	0.52
IMC	1.08	2.21	1.09	2.00	0.91	1.96	0.64	0.93	0.59	1.40	0.36	0.58	0.95	0.44
	1.13	0.91	1.54	1.28	0.31	1.42	0.43	0.75	0.23	1.47	0.44	1.08	0.38	0.69

These numbers consistently show that the TTC index for prolonged deviations is larger on the long final than the short final, for all pilots and both weather conditions. From a one-way ANOVA it follows that this is a statistically significant result with $F(1,54) = 89.62$, $p<0.01$. The TTC index for temporary deviations, on the other hand, is statistically significantly larger on the short final, for all pilots and both weather conditions, with $F(1,54) = 7.33$, $p<0.01$.

Table 3 compares the weighed TTC indices as given in eq. (3) for the VMC and IMC. Though not consistently the case for every pilot, it appears that the penalty in IMC is higher than in VMC, confirming the expectation that the IMC (degraded visibility) would be riskier. A one-way ANOVA demonstrates that this result is statistically significant, with $F(1,26) = 5.74$, $p<0.05$.

Table 3. Weighted prolonged and temporary time-to-crash function values for VMC and IMC, per pilot. (Dimensionless; note that the overall average is 2 as per eq. 3)

	SP1, day 1		SP1, day 2		SP2		PP1, day 1		PP1, day 2		PP2		PP3	
	VMC	IMC	VMC	IMC	VMC	IMC	VMC	IMC	VMC	IMC	VMC	IMC	VMC	IMC
Flight 1	1.26	2.96	1.74	2.95	1.77	2.29	2.04	1.90	1.65	2.09	2.23	2.10	2.05	2.87
Flight 2	1.74	2.55	2.59	3.01	1.70	1.69	1.41	1.46	1.71	1.64	1.86	1.60	1.26	1.87

4.2 Analysis of the Control Input Signal

Figure 5 shows the typical elevator deflection and power distribution. The instantaneous elevator control power (green) has been averaged with a 5 seconds window (black) to better see the trend. A separation line (blue dashed) indicates the end of the long final segment (until 35 seconds to touchdown) and beginning of the short final segment (the last 35 flight seconds). Clearly, the average power during the short final is higher.

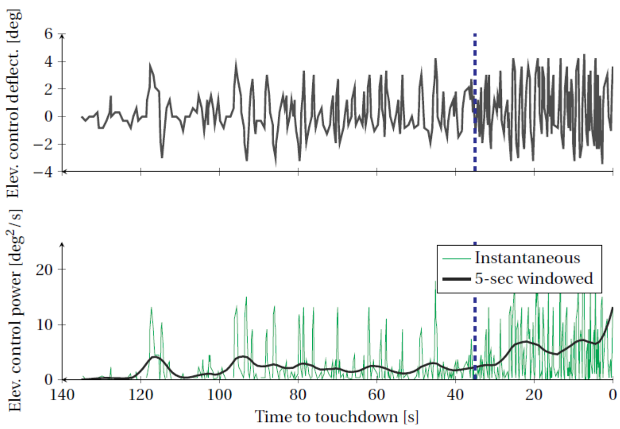

Fig. 5. Power and magnitude of elevator deflection for PP1, day one, IMC flight 2.

The values in table 4 represent the elevator control power averaged on the long and short final segment for all flights. It can be seen that for nearly all flights the short final contains more elevator power than the long final segment. A one-way ANOVA shows that this is a statistically significant result ($F(1,54) = 16.98$, $p<0.01$).

The difference in elevator control power between VMC and IMC in the long final was not significant. This was to be expected, since the control in the long final is mostly based on the cockpit instruments, so outside visibility is not likely to influence the control.

Table 4. Elevator control power [deg^2/s] on long and short final per pilot and weather condition.

	SP1, day 1		SP1, day 2		SP2		PP1, day 1		PP1, day 2		PP2		PP3	
	long	short	long	short	long	short	long	short	long	short	long	short	long	short
VMC	5.71	9.05	5.97	9.86	1.66	2.92	1.11	2.39	0.35	1.56	1.69	2.00	2.12	9.02
	5.25	12.43	6.61	7.80	1.96	1.53	0.70	2.12	0.62	1.21	1.19	6.17	2.87	5.23
IMC	8.19	17.54	4.06	12.57	3.22	10.53	1.38	3.48	0.93	1.94	1.66	7.53	0.93	2.12
	5.00	17.09	6.14	16.87	3.34	11.14	1.90	5.76	0.41	2.84	2.06	7.39	1.67	8.24

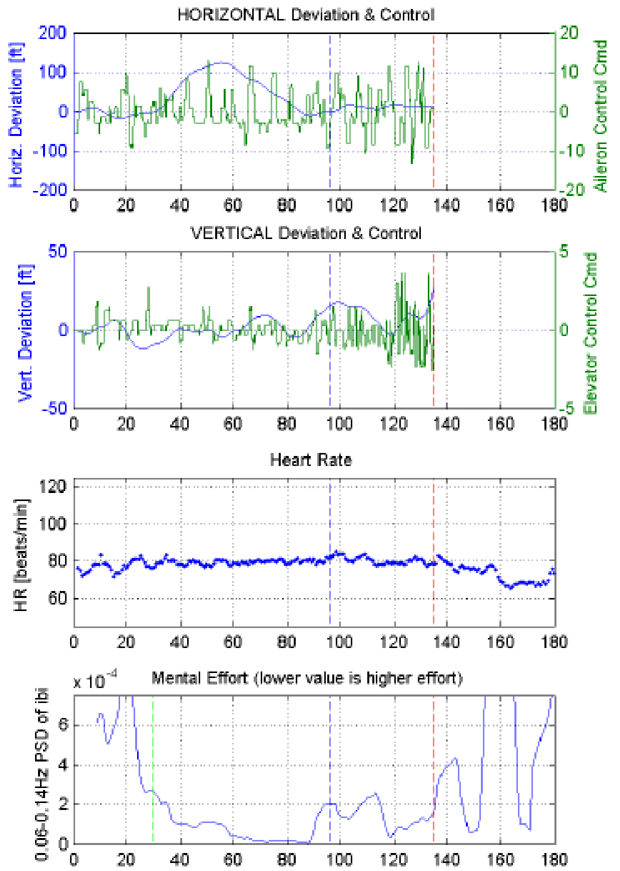

Fig. 6. Analysis of heart rate and HRV for a flight of the retired professional pilot.

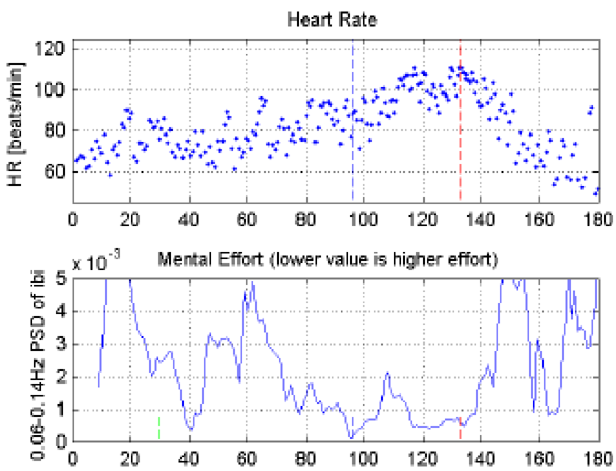

Fig. 7. Analysis of heart rate and HRV for a flight of a student pilot.

In the short final the IMC flights require significantly more control action than the VMC ones ($F(1;26)=4.23$, $p<0.05$), supporting our hypothesis that IMC flights require more pilot effort.

4.3 Heart Rate Variability Analysis

Figure 6 shows the HR and HRV analysis. It is interesting to note that the Heart rate of the professional pilot is very constant and quite high. Changes in the HRV can be related to the deviation. Around 90s into the flight, the aircraft is horizontally aligned, and the pilot seems to relax a bit (increase of PSD means decrease of mental effort). At the final part of the landing, in the flare, the PSD goes down again, meaning that the pilot has a high mental workload again, which is common during the flare and often self-reported.

As a comparison, fig. 7 shows the HR and HRV of a student pilot. Whereas the HR of the captain is extremely constant, the student's HR shows high variability and the PSD is about an order of magnitude higher. This remarkable difference may have to do with personal differences such as age, but probably also with training. More research is needed to clarify this.

It is clear for all subjects that the PSD, indicating metal effort, is much more suppressed during the flight, than just after. This confirms our expectations that operating the flight simulator requires much effort. Differences between flight phases and conditions have to be studied in more detail before hard conclusions can be drawn.

4.4 Analysis of Blinks and Pupil Diameter

At the moment, only a very crude analysis of blinks and pupil diameter can be made. The main problems are that both analyses rely heavily on proper calibration and filtering of the measured signal. Especially blink filtering (removing half blinks, noise, position out-of-range, etc.) is challenging.

The analysis of the change in pupil diameter is very sensitive to both blinks and to the changing scene brightness (between instrument panels and outside view, and even between different instruments on the panel and areas in the outside view). We have therefore not yet been able to obtain conclusive results.

One result is that we can confirm that in the last part of the short final (i.e., the visual approach), pilots almost never blink. This is in line with earlier research.

5. Conclusions

New technology and automation lead to more efficient and safer operations in general, but many pilots fear a simultaneous degradation of their manual flying skills. As Capt. Drappier put it: *"The transition between smooth easy flying on [AutoPilot] and being challenged by hair-raising situations can be very abrupt in the modern cockpits. In some respects, automated aircraft may require a higher standard of basic stick and rudder skills, if only because these skills are practiced less often and maybe called upon in the most demanding* *emergency situations.*[9)] "

Indeed, automation will probably work correctly in standard operations, but fail or face its limitations in challenging situations. It is in these already high-workload conditions that the pilot will suddenly have to assume the full control authority.

Our research proposed various methods to analyze risk. The time-to-crash index showed that prolonged deviations from the ideal TTC are larger on the long final than the short final, whereas temporary deviations are larger on the short final. This indicates that the pilot needs time to stabilize his approach. Then, during the short final, he focuses and makes many small corrections. Having sufficient time to stabilize the aircraft is therefore important, and this may be difficult when the final straight segment in RNP-AR curved approaches gets too short.

The elevator control power analysis and ecg analysis were also show to be beneficial for discriminating between easy and difficult situations, and are therefore promising for the future analysis of the challenges that pilots face when they fly RNP-AR curved approaches. Data obtained using the eye camera might be useful, in particular blink rate, but proper filtering is still an issue and no hard conclusions can be drawn at the moment.

Especially with highly automated operating procedures such as RNP-AR, sufficient training to recognize and mitigate problems is essential to maintain flight safety.

Acknowledgments

The authors would like to thank Mr. Daiichiro Matsunaga for his assistance with operating the simulator and other equipment. This research is financially supported by the Japan Aerospace eXploration Agency (JAXA) under the next generation air-traffic management system "*DREAMS*" program, and we like to express our gratitude to Dr. Kohei Funabiki of JAXA for his guidance.

References

1) Causse, M., Baracat, B., Pastor,J., Dehais, F.: Reward and Uncertainty Favor Risky Decision-Making in Pilots: Evidence from Cardiovascular and Oculometric Measurements, *Appl Psychophysiol Biofeedback*, Vol. 36 (2011), pp. 231–242.

2) Gawron, V.J.: *Human Performance, Workload, and Situational Awareness Measures Handbook*, Second Edition. Taylor & Francis, (2008).

3) Aasman, J., Mulder, G., and Mulder, L.J.M.: Operator Effort and the Measurement of Heart-Rate Variability, *Human Factors*, Vol. 29 (1987), Nr. 2, pp. 161-170.

4) Vicente, K.J., Thornton, D.C., and Moray, N.: Spectral Analysis of Sinus Arrhythmia: A Measure of Mental Effort, *Human Factors*, Vol. 29 (1987), Nr. 2, pp. 171-182.

5) Clifford, G.D.: ecgBag software. (2010) MIT website. Available at: www.mit.edu/~gari/CODE/ECGtools/ecgBag/.

6) Task Force of The European Society of Cardiology and The North American Society of Pacing and Electrophysiology: Heart rate variability - Standards of measurement, physiological interpretation, and clinical use, *European Heart Journal*, Vol. 17 (1996), pp. 354–381

7) Entzinger, J.O., and Suzuki, S.: Inside the Pilot's Mind: Visual Information Use in Manual Aircraft Landings, *Proceedings of the 28th Congress of the International Council of the Aeronautical Sciences*, 23-28 September 2012, Brisbane, Australia.

8) Palinko, O., Kun, A.L., Shyrokov, A., and Heeman, P.: Estimating Cognitive Load Using Remote Eye Tracking in a Driving Simulator, *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications* pp. 141-144

9) Drappier, J.: Maintaining Manual Flying Skills, *Proceedings of the 2012 World Aviation Training Conference and Tradeshow (WATS 2012)*, 17-19 April 2012, Orlando, Florida, USA.