Human Factors and Aerospace Safety 6(4), 383-394

The application of frequency analysis based performance measures as an adjunct to flight path derived measures of pilot performance

Matt Ebbatson, John Huddlestone, Don Harris
and Rodney Sears
Cranfield University, UK

Abstract

When evaluating pilot performance it is common to assess the magnitude of errors between the tracked parameter and a target value using measures such as the arithmetic mean error and standard deviation of error. These have strong validity when applied to parameters such as flight path or airspeed deviation especially when associated with a well-prescribed flight task that demands a high level of performance, such as flying an ILS-approach. However, there is a certain disassociation between the control input behaviour of the pilot and the flight path response of the aircraft, particularly in large transport types with relatively high inertia and stability. This study uses frequency-based metrics based on pilot control inputs as an adjunct to these commonly used measures to evaluate performance. Using both types of measure it can be seen how the performance of 12 cadet pilots changed while undertaking a 40-hour Jet Orientation course on a Boeing 737NG flight training device. The results show that variation in the flight path is reduced as the cadet pilots progress through the course. At the later stages of the course the control strategy used is characterised by more frequent but smaller amplitude control inputs. It is concluded that such measures would be

Correspondence: Matt Ebbatson, Flight Operations Research Centre of Excellence, Department of Human Factors, School of Engineering, Cranfield University, Cranfield, Bedford, MK43 0AL, UK or e-mail m.ebbatson@cranfield.ac.uk

suitable for future studies of the potential change in manual flying skills due to automation.

Introduction

In modern jet transport aircraft the pilot typically flies by programming commands into the Autoflight System, either tactically via the Mode Control Panel (MCP) or strategically via the Flight Management System (FMS), rather than through manual manipulation of the flying controls. In Wood's (2004) study of flight crews' dependency on automation it was noted that strong anecdotal evidence existed to suggest that pilots of highly automated aircraft may experience manual flying skills decay as a result of a lack of opportunity to practice during line operations (see also Curry, 1985; Veillette, 1995; Owen and Funk, 1997). It was suggested that this may pose a threat to flight safety, predominantly during periods of imposed manual flying, such as following a partial degradation (or even outright failure) of the aircraft's automation. However, there is a paucity of objective evidence available to substantiate this concern.

Young, Fanjoy and Suckow (2006), in a study of pilots during refresher training, observed that flight crew who made greatest use of the automatics demonstrated weaker manual control skills. However, their analysis of performance was based purely on observer assessments. The only study (to date) based on flight data was undertaken by Veillette (1995) who examined the manual tracking skills of crews operating analogue and 'glass cockpit' variants of a jet transport aircraft over a broad range of flight regimes. Significant differences were observed in performance between the groups. Pilots of highly automated aircraft exhibited greater deviations from their assigned course and greater variation in attitude parameters during normal and abnormal operations than did their counterparts. It was suggested that the level of performance observed presented a possible threat to safety. It was also noted that there was greater variability in performance within the 'glass cockpit' group. To this date there still exists some concern that the standard of manual flying skills may be declining. Since the study undertaken by Veillette the proportion of highly automated aircraft in the air transport fleet has grown enormously and the number of pilots with experience of less highly automated airliners has diminished.

When evaluating performance on any tracking task, such as flying an aircraft, it is most common to examine the end product of performance (i.e. measuring errors between the tracked parameter and a target value). Metrics such as the arithmetic mean error and standard deviation of error have strong validity when applied to parameters such as flight path or airspeed deviation especially when associated with a well-prescribed flight task that demands a high level of performance, such as flying an ILS-approach. The arithmetic mean error gives an indication of the overall flight path error (on a particular axis) and its associated standard deviation gives a measure of the ‘smoothness’ of the pilot’s performance. These two parameters are often used in preference to the Root Mean Square Error (RMSE). Taken in combination, the arithmetic mean error and the standard deviation of error completely define the root mean square error. Furthermore RMSE also has the additional disadvantage that it produces identical values for quite disparate performances. For example, being consistently high, consistently low, or at the correct mean height but with great variations in height keeping may all result in the same RMSE value (see Hubbard, 1987). Such measures have been used on many occasions to evaluate pilot performance, such as when comparing alternate training strategies (e.g. Rees and Harris, 1995); evaluating the effects of alcohol (Davenport and Harris, 1992) or evaluating the efficacy of emergency flight control systems (Demagalski, Harris and Gautrey, 2002).

However, there is a certain disassociation between the control input behaviour of the pilot and the flightpath response of the aircraft. The series model of pilot control (McRuer, 1982) illustrates this. The pilot cannot directly observe the aircraft’s flight path and so instead effects control via a lower order surrogate, the aircraft’s attitude (e.g. pitch control) through use of the primary flight controls. The pilot is essentially trying to close an inner loop (related to aircraft attitude) as a surrogate for controlling the flight path (see figure 1). In a large conventional transport aircraft, the relationship between control input, aircraft attitude and flight path variation is mediated by factors such as inertia, control power and the relatively high stability of the machine. There is often a significant delay between control input and the aircraft’s response. Consequently further control inputs after the initial input may serve to cancel it out or reinforce the initial input before it has taken effect. As a result, significant control input activity may not be reflected in large changes in the aircraft’s attitude, and less so in the flight path. It is therefore unlikely that basic flight path measures alone will have the sensitivity required to investigate fine variations in manual flying skills. However, there is the potential for a direct assessment of the pilot’s control strategy by studying the inputs to the primary flight controls as an adjunct to the aforementioned performance metrics. It is argued that the evaluation of the pilot’s control inputs to the primary flight controls provide a direct measure of performance when closing the inner (attitude) control loop, whereas measures of flight path performance are a measure of success in closing the outer (flight path) control loop (see figure 1). Measures relating to the aircraft’s flight path can be regarded as ‘product’ measures, in contrast to measures derived directly from control inputs, which are essentially ‘process’ measures.

McDowell (1978) began the development of control movement power spectra (a description of the distribution and weighting of control input frequencies) based measures to evaluate pilot performance. McDowell used a series of filters to estimate the power spectra of control movements for novice, intermediate and experienced pilots flying a Cessna T-37 light military training aircraft. It was found that more experienced pilots generally used higher frequency control inputs, particularly in the roll axis. It was concluded that there were changes in pilot’s control movement power spectra as a function of skill level, and that measures of this property could be used effectively to discriminate pilot skill/experience level.

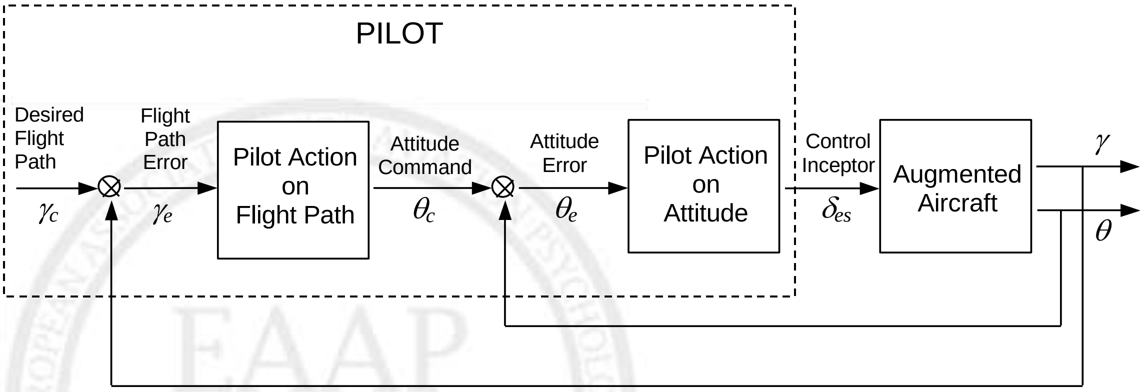

Figure 1 The series model of pilot control (adapted from McRuer, 1982)

More recently Rantanen, Johnson and Talleur (2004) regenerated interest in the use of frequency analysis based performance metrics, using modern computing techniques to perform Fourier transforms on flight path time series data. The result of the Fourier transform was to produce an estimate of the power spectra of the time series data from which various performance metrics were computed (see Method section). These metrics were applied in a series of flight trials. Data were collected from pilots performing an instrument proficiency check (IPC) in a light aircraft. The study demonstrated that the metrics were capable of discriminating between pilots who had passed or failed the IPC. However, in this case the metrics were applied primarily to flight path (outer loop) parameters, such as course deviation and glideslope deviation indications, rather than directly to primary flight control movement data.

The aforementioned studies by McDowell (1978) and Rantanen, Johnson and Talleur (2004) all used light, responsive aircraft. As noted previously, though, the relationship between control input, aircraft attitude and flight path variation is mediated by a number of other factors, especially in a large transport aircraft with significant lags in the flight control system. To evaluate fully a pilot’s manual flying skills it is suggested that both ‘product’ (outer loop) and ‘process’ (inner loop) measures are required. This study evaluates if frequency-based metrics provide a sensitive adjunct to conventional, flight path derived measures of performance when assessing the performance of novice pilots undergoing initial conversion training to a large jet-engined airliner.

Method

Participants

The manual handling performance of 12 cadet pilots (all male, aged between 18 and 25) was evaluated whilst they undertook a 40-hour Jet Orientation course on a Boeing 737NG flight training device. All students had similar levels of flying experience (approximately 180 hours in light singles and twins) at the start of the course. None had any experience on large jet transport types prior to commencing the conversion course.

Equipment

The Jet Orientation training comprised of a series of line orientated scenarios conducted on a fixed base Boeing 737NG flight training device. The flight training device was approved for jet orientation training by the regulator and was equipped with 180 degree projected visual display and electrically back driven flying controls. A data-logging computer was integrated into the training device. The data logger sampled 96 selected flight parameters (including all primary control inputs) at a sampling rate of 4Hz. Triggering conditions were specified in the logging software so that data recording would commence automatically at the beginning of an approach segment. Time and date encoded output data files were produced in comma separated variable format.

Task

The jet orientation syllabus required students to fly manual precision instrument approaches at a number of intervals throughout the course. The orientation course immediately followed the students' initial ATPL training. Each student's performance was sampled twice, once within the initial period of training and once during the final stages of the training programme.

The approaches were conducted in a standardised form. Students were asked to fly a manual ILS approach (i.e. without autopilot, flight director or autothrottle assistance) in IMC to a minimum decision height. In the development of the frequency-based metrics by Rantanen et al. (2004) this proved to be the task that gave best discrimination of tracking performance between pilots who passed or failed an instrument proficiency check flight.

Measures

For each approach, ILS tracking error (on both localizer and glideslope), airspeed error and control input data were recorded for the segment between passing 2,500ft established on the ILS and the stabilised approach point at 500ft above aerodrome level (AAL). The data from the trials were analysed using Flightscape's Insight™ flight data analysis software.

The flight path performance metrics computed were the arithmetic mean (ME) and standard deviation (SD) of glideslope and localizer angular error, as well as the ME and SD of airspeed error relative to the target approach speed (See Hubbard, 1987). A Matlab™ M-file was then used to compute the Fourier transform of the control input data and in turn determine the weighting of each frequency in the series, the power spectral density (PSD).

The Fourier coefficients, $\tilde{Y}_j$, in the Fourier transform, Y_k, of the time series data were given by:

$$\tilde{Y}_j = \sum_{k=1}^{N} Y_k e^{\frac{-2\pi_i}{N}(k-1)(j-1)}$$

From the Fourier transform it is possible to then calculate how the power of the time series is distributed across frequencies (the power spectral density), given by:

$$\frac{\left|\tilde{Y}_j\right|^2}{N}$$

where N is the number of points in the time series.

This distribution essentially describes the weighting of control input frequencies and from it the performance metrics developed by Rantanen et al (2004) were computed. These are summarised in table 1.

Measures were computed for pitch, roll and yaw control inputs. As with the previous study a critical value was derived to ensure that only the significant control-related frequency components were analysed and background noise was removed from the data file (Johnson, Rantanen and Talleur, 2004). This critical value was set at 5% of the power of the largest spectral component.

Results

Flight path performance measures (Product measures)

Using a paired t-test no significant differences were observed in performance early and late in the training course on the mean tracking error on the ILS glideslope, t(11) = 0.481, p>0.05. However, the performance later in the training course showed significantly smaller standard deviations in glideslope tracking error, t(11) = 3.548, p<.05 indicative of smoother control of the aircraft's profile (see table 2).

Significant differences were observed in the mean tracking error of ILS localiser course early and late in the training course, t(11) = 1.808, p<.05.

Performance in the later stages of training showed a smaller degree of error, tracking closer to the centreline. Similarly, performance later in the training course exhibited significantly lower standard deviation of tracking error, t(11) = 2.332, p<.05, indicative of smoother control of the aircraft's track (see table 2).

Table 1 Fourier analysis based performance metrics as developed by Rantanen et al (2004).

Metric	Description
Mean Magnitude of Spectral Components (MSC)	Mean of the $\frac{\lvert\tilde{Y}_j\rvert^2}{N}$, relating to the amplitude of deviations in the time series
Number of Spectral Component Greater than Cut-Off (NCGC)	Number of $\frac{\lvert\tilde{Y}_j\rvert^2}{N}$ greater than a minimum cut-off value to remove noise
Mean Frequency of Spectral Components Greater than Cut-Off (FMGC)	Mean frequency of the $\frac{\lvert\tilde{Y}_j\rvert^2}{N}$ greater than a minimum cut-off value
Standard Deviation of the Frequency of Spectral Components Greater than Cut-Off (FDGC)	SD of frequencies of the $\frac{\lvert\tilde{Y}_j\rvert^2}{N}$ greater than a minimum cut-off value
Median Frequency of the Spectral Components (MEDF)	Median frequency of the power spectrum

There was no significant difference between the mean airspeed errors early and late on the training course, t(11) = 0.324, p>.05. However, performance later on the course did demonstrate significantly lower standard deviations of airspeed error, t(11) = 3.710, p<.05. This was indicative of greater stability in the control of the target approach speed (see table 2).

Control input measures (process measures)

The mean magnitude of the pitch control spectral components (MSC) was significantly lower for pilots late in their training, t(11) = 1.863, p<.05, indicating that the control inputs were generally of lower amplitude (see table 3). The mean and median frequencies of pitch control spectral components (FMGC and MEDF) were significantly higher late in training, t(11) = -1.802, p<.05 and t(11) = -1.804, p<.05, respectively, indicating that pilots later in training generally made a greater number of pitch inputs when attempting to control the profile. However, there was no significant difference between either the number or spread of significant frequency components (NCGC and FDGC) early and late in training, t(11) = -1.667, p>.05 and t(11) = 1.239, p>.05, respectively. The complexity and range of the pitch control strategies did not differ significantly.

Table 2 Arithmetic mean error and standard deviation for ILS product performance parameters, broken down by early or late course assessment.

			ILS Tracking Error			
	Glideslope (deg)		**Localiser (deg)**		**Airspeed (kts)**	
	M	**σ**	**M**	**σ**	**M**	**σ**
Mean Error (Early Course)	0.064	0.140	0.513	0.507	3.246	4.647
Mean Error (Late Course)	0.040	0.064	0.191	0.269	2.719	3.880
Standard Deviation (Early Course)	0.199	0.111	0.411	0.178	6.656	2.333
Standard Deviation (Late Course)	0.084	0.052	0.212	0.118	3.078	2.208

There were no significant differences in the mean magnitude of roll control spectral components (MSC), t(11) = 1.529, p>.05, early or late in the training course, suggesting no significant differences in the amplitude of control wheel deflections. Likewise there were no significant differences between either the mean frequency (FMGC), t(11) = 0.086, p>.05, or the median frequency (MEDF) of roll control inputs. Additionally, there were no significant differences in the number of significant spectral components (NCGC), t(11) = 1.161, p>.05, or the spread of their frequencies (FDGC), t(11) = -0.102, p>.05, indicating no significant differences between the complexity or range of the roll control strategies employed during early or late training.

For the yaw axis, there were no significant differences in the mean magnitude of yaw control frequency spectra (MSC), t(11) = 1.036, p>.05, early or late in the training course indicating no significant differences in the amplitude of rudder pedal inputs between early and late training. Similarly there were no significant differences between the mean frequency (FMGC), t(11) = 0.662, p>.05, or the median frequency (MEDF), t(11) = 0.646, p>.05, of yaw control inputs. Finally there was no significant differences between number of significant spectral components (NCGC), t(11) 0.527, p>.05, and the spread of their frequencies (FDGC), t(11) = -0.113, p>.05, indicating no significant differences in the complexity or range of the yaw control strategies utilised.

Table 3 **Frequency analysis metrics for primary flight control inputs (process parameters) during the ILS tracking task broken down by early or late course assessment.**

	Control Input Spectral Analysis					
	Pitch Input		Roll Input		Yaw Input	
	M	σ	M	σ	M	σ
MSC[a] (Early Course)	382.882	745.349	4551.011	4814.001	1686.278	3151.009
MSC[a] (Late Course)	106.501	169.342	3820.362	5227.165	668.354	746.155
NCGC (Early Course)	14.000	9.667	24.130	8.467	4.170	3.713
NCGC (Late Course)	19.000	7.711	20.130	8.593	3.670	1.231
FMGC[b] (Early Course)	0.041	0.023	0.120	0.446	0.019	0.026
FMGC[b] (Late Course)	0.054	0.023	0.119	0.035	0.014	0.010
FDGC[b] (Early Course)	0.034	0.018	0.078	0.034	0.016	0.020
FDGC[b] (Late Course)	0.041	0.015	0.080	0.023	0.016	0.015
MEDF[b] (Early Course)	0.017	0.013	0.090	0.030	0.004	0.009
MEDF[b] (Late Course)	0.028	0.019	0.087	0.037	0.003	0.008

[a] PSD (W/Hz) - power of individual frequency components in the spectral distribution.

[b] frequency – Hz. See table 1 for metrics description.

Discussion

Measures of flight path performance

When measuring performance from a traditional perspective (by examining error in outer-loop parameters – see McRuer, 1982) the results showed a general increase in performance in all measured parameters over the period of the training course. The student's standard deviation of error values, indicative of their smoothness in tracking, for localiser, glideslope and airspeed targets all decreased over the period of training showing an improvement in performance. In addition the accuracy of localiser tracking improved over the period, with mean error more than halving. This confirms that these product-related metrics are still sensitive, even when flying a large aircraft and provide useful information in the investigation of manual flying performance (cf. Davenport and Harris, 1992; Rees and Harris, 1995; Demagalski, Harris and Gautrey, 2002).

Measures of control movement

With regard to the inner-loop control performance of pilots the frequency analysis based metrics successfully discriminated between the performance of students early and late on the conversion course on a number of dimensions. In keeping with the results of the earlier work performed by McDowell (1976), over the period of the training course there was observed an increase in the mean pitch input frequency in the spectral components, as well as an increase in the central frequency of the spectral distribution. In addition, the mean PSD decreased over this period, indicating a decrease in the amplitude of pitch inputs. These results would also seem to complement the findings of Rantanen et al (2004), although it should be noted once again that this study examined outer-loop (product) based performance, rather than inner-loop control behaviour (see McRuer, 1992, figure 1). The results from the present study indicate that more experienced pilots have a finer touch on the controls and perhaps perform more iterations of the outer control loop to contain the flightpath. These measures of the control input process are used as an adjunct to the flight path (product) measures and allow the performance to be described in greater depth. The data suggests that variation in the flightpath is reduced when frequent but small control inputs are employed. Large variations occur when less frequent, but larger magnitude, control inputs are used. The latter would seem to be the control strategy employed by novices.

The frequency analysis-based metrics were not as sensitive with regard to identifying differences between early and late training performance when applied to roll or yaw input data. This may be an attribute of the task setting, with the lateral tracking component requiring relatively low performance levels compared to the vertical tracking which was presented. As demonstrated by McDowell (1976), performance metrics typically discriminated more effectively when applied to complex tasks rather than simple tasks, as they give the performer a greater opportunity to demonstrate their expertise and ‘stand out from the crowd’. In this study the vertical tracking component was highly coupled to changes in thrust and airspeed parameters and could be considered to be more complex than the lateral tracking task, which with no external disturbances and symmetrical thrust, is relatively straightforward. Therefore in the context of this ILS task there may have been insufficient opportunity to evaluate performance in this respect. A more demanding lateral tracking task, such as an asymmetric or crosswind approach, may generate greater variation on this axis.

Conclusions and recommendations

The results of this initial study suggest that the frequency metrics proposed by McDowell (1978) and Rantanen et al. (2004) can provide a sensitive measure of pilot performance in air transport aircraft when applied directly to control input data, and find similar results to studies based on light aircraft using different methodologies. These metrics give an extra dimension to performance measurement, relating directly to the inner loop of the series control model and describing the process by which pilots exercise their control. As a result these metrics are considered to be a useful adjunct suitable for use in future studies of manual flying skill in large jet transport aircraft. They enable a more complete picture to be developed for the future study of manual flying skills.

It is suggested that further research should be conducted to investigate the effects of a more demanding lateral tracking task to help assess the sensitivity of these metrics, for example a backing crosswind may be introduced, or asymmetric flight conditions may be studied. The effects in other related parameters such as thrust lever position should also be considered.

References

Curry, R.E. (1985). *The Introduction of New Cockpit Technology: A Human Factors Study. NASA Technical Memorandum 86659, 1-68.* Moffett Field, CA: NASA Ames Research Center.

Davenport, M.D. and Harris, D. (1992). The Effect of Low Blood Alcohol Levels on Pilot Performance in a Series of Simulated Approach and Landing Trials. *International Journal of Aviation Psychology, 2,* 271-280.

Demagalski, J.M., Harris, D. and Gautrey, J.E. (2002). Flight control using only engine thrust: development of an emergency display system. *Human Factors and Aerospace Safety 2,,* 173-192.

Hubbard, D.C. (1987). Inadequacy of root mean square error as a performance measure. In, R.S. Jensen (ed.), *Proceedings of the Fourth International*

Symposium on Aviation Psychology (pp. 698-704). Columbus, OH: Ohio State University.

Johnson, N.R. Rantanen, E.M. Talleur, D.A. (2004). Criterion Setting for Objective Fourier Analysis Based Pilot Performance Metrics. *Proceedings of the Human factors and Ergonomics Society 48th Annual Meeting*, Santa Monica.

McDowell, E.D. (1978). *The Development and Evaluation of Objective Frequency Domain Based Pilot Performance Measures in ASUPT*. Air Force Office of Scientific Research, Bollings AFB, DC.

McRuer, D.T. (1982). *Pitfalls and progress in advanced flight control systems (AGARD CP-321)*. Neulliy-sur-Seine; AGARD/NATO.

Owen, G. and Funk, K. (1997). *Flight Deck Automation Issues: Incident Report Analysis*.http://www.flightdeckautomation.com/incidentstudy/incidentanalysis.aspx. Corvallis, OR: Oregon State University, Department of Industrial and Manufacturing Engineering

Rantanen, E.M. Johnson, N.R. Talleur D.A. (2004). *The Effectiveness of a Personal Computer Aviation Training Device, a Flight Training Device, and an Airplane in Conducting Instrument Proficiency Checks. Volume 2: Objective Pilot Performance Measures*. AHFD-04-16/FAA-04 6. FAA Civil Aerospace Medical Institute, Oklahoma City, OK

Rees, D.J. and Harris, D. (1995). The Effectiveness of ab initio Flight Training Using Either Linked or Unlinked Primary Axis Flight Controls. *International Journal of Aviation Psychology 5*, 291-304.

Veillette, P. (1995). Differences in Aircrew Manual Skills in Automated and Conventional Flight Decks. *Transport Research Record 1480*, 43-50.

Wood, S. (2004) *Flight Crew Reliance on Automation*. CAA Report 2004/10. Civil Aviation Authority, Gatwick.

Young, J.P. Fanjoy, R.O. Suckow, M.W. (2006). Impact of Glass Cockpit Experience on Manual Flight Skills. *Journal of Aviation Aerospace Education and Research 15*, 27-32.

Acknowledgements

The Flight Operations Research Centre of Excellence (FORCE) is co-funded by the UK Civil Aviation Authority (CAA) the Engineering and Physical Sciences Research Council (EPSRC) and Cranfield University. The Loss of Manual Flying Skills project is also part sponsored by the Guild of Air Pilots and Air Navigators (GAPAN). European Pilot Selection and Training (EPST) of Holland kindly provided access to their training course and facilities for the collection of data. We would also like to thank Professor Esa Rantanen and Nick Johnson for their assistance in the production of the Matlab files used within this work.