Novel Estimation of Pilot Performance Characteristics

Edward N. Bachelder[1] and Bimal Aponso[2]

San Jose State University/US Army Aviation Development Directorate/NASA Ames Research Center, Moffett Field, CA, 94035, USA

Two mechanisms internal to the pilot that affect performance during a tracking task are: 1) Pilot equalization (i.e. lead/lag); and 2) Pilot gain (i.e. sensitivity to the error signal). For some applications McRuer's Crossover Model can be used to anticipate what equalization will be employed to control a vehicle's dynamics. McRuer also established approximate time delays associated with different types of equalization – the more cognitive processing that is required due to equalization difficulty, the larger the time delay. However, the Crossover Model does not predict what the pilot gain will be. A nonlinear pilot control technique, observed and coined by the authors as 'amplitude clipping', is shown to improve stability, performance, and reduce workload when employed with vehicle dynamics that require high lead compensation by the pilot. Combining linear and nonlinear methods a novel approach is used to measure the pilot control parameters when amplitude clipping is present, allowing precise measurement in real time of key pilot control parameters. Based on the results of an experiment which was designed to probe workload primary drivers, a method is developed that estimates pilot spare capacity from readily observable measures and is tested for generality using multi-axis flight data. This paper documents the initial steps to developing a novel, simple objective metric for assessing pilot workload and its variation over time across a wide variety of tasks. Additionally, it offers a tangible, easily implementable methodology for anticipating a pilot's operating parameters and workload, and an effective design tool. The model shows promise in being able to precisely predict the actual pilot settings and workload, and observed tolerance of pilot parameter variation over the course of operation. Finally, an approach is proposed for generating Cooper-Harper ratings based on the workload and parameter estimation methodology.

Nomenclature

Y_p	=	pilot describing function
Y_c	=	controlled element transfer function
τ	=	pilot time delay
K	=	pilot gain
z	=	pilot lead frequency
δ	=	pilot control input
e	=	displayed error
CM	=	Crossover Model

I. Introduction

A myriad of factors affect pilot opinion of workload – cognitive capacity/capability, aircraft dynamics, proficiency, etc. To reduce the subjective variability in pilot opinion that results from these numerous determining factors, the Cooper-Harper handling qualities pilot rating (HQR) scale (Ref. 1) estimates workload on the compensation necessary to achieve the required performance in a flight task. The scale incorporates a decision-tree structure that is based on these two factors – compensation and performance attainment. An absolute measure of workload is difficult to obtain due to the multitude of influencing factors, the Bedford scale (Ref. 9) attacks the problem of assessing workload directly by asking the pilot to rate workload in terms of spare capacity available. These structured scales, when coupled with carefully constructed and executed pilot-in-the-loop

---

evaluations/experiments can provide pilot opinion data with reduced variability (particularly in handling qualities assessments) that allows statistically viable results that may be used as a premise for setting requirements on aircraft behavior and performance (Ref. 2). However, understanding the fundamental bases and processes that form pilot opinion remains elusive. Initial work by McRuer (Ref. 3) correlated pilot opinion to pilot model parameters that reflected the performance and compensation, and coalesced to the Cooper-Harper scale. McDonnell looked at workload in terms of spare capacity using the division of attention between a primary tracking task (usually the aircraft being evaluated) and an increasing unstable element that the pilot needs to stabilize (Ref. 4). The premise was that the maximum level of instability the pilot could manage in the secondary axes while maintaining performance in the primary axis was indicative of the pilot's maximum 'spare capacity'. This assessment, however, relied on a secondary task and could not readily be applied to the general situation of evaluating workload in a realistic operational setting.

In a 1982 study, Hess (Ref. 5) indicated that the combination of error and control rate can be used as a metric giving an indication of subjective rating, where error captures the effects of performance and control rate captures the effects of workload on the rating. Good correlation was demonstrated between the metric and actual ratings, however this approach was subject to the following limitations:

1. A precise mathematical description of aircraft and display dynamics driving the tracking task was required to scale the control rate and error
2. The tasks investigated were restricted to attitude tracking, and required that the vehicle dynamics could be represented in the form typical of aircraft pitch response shown in Eq. 1, where higher-order aircraft dynamics were linearized to fit this form

$$\frac{\theta}{\delta} = \frac{K_\theta(s + \frac{1}{T_L})e^{-\tau_D s}}{s(s^2 + 2\zeta_n \omega_n s + \omega_n^2)} \tag{1}$$

3. The metric was extremely sensitive to the derived equivalent time delay τ_D shown in Eq. 1
4. An absolute rating needed a 'calibration factor' which was obtained using an actual pilot rating for the task. Thus calibration was subject to one data point's bias, noise, etc.
5. Display and/or vehicle nonlinearities such as mode switching, saturation, etc. cannot be accounted for.

There are numerous other manual control tracking tasks besides attitude tracking (and of greater mission importance) that can occur in the cockpit, such as station-keeping (hovering or formation flight), velocity and glideslope tracking, etc. In light of this and the limitations 1-5 listed above, an approach that is more widely applicable and easier to implement may be desirable.

Hoh (1998, Ref. 6) proposed to fit visual cue rating (VCR) data to HQR data from various rate-command helicopters to offer a relationship between HQR and VCR, and attentional demand and VCR. However, VCR is itself a subjective measure, and the findings are mostly relevant to the task and vehicle used in this study, and therefore, not generalizable.

More recent and practical objective measures of workload are those that attempt to relate pilot rating to stick activity by assessing the frequency of the stick activity in a tracking task (Ref. 7). An improved version of this measure is power frequency that was suggested by Lampton & Klyde (2012, Ref. 8). Both these objective measures were developed to provide a basis for assessing pilot opinion for handling qualities where workload is not explicitly rated. These measures, however, are task-specific, and as will be shown later, do not correlate well with the Bedford workload rating scale.

As will be demonstrated, pilot technique can exert a large influence on workload and performance. Technique has been ignored until now, which may be one reason why assessment and prediction of workload (and prediction of pilot settings) has been so elusive. Pilot style can introduce significant nonlinearities, which if not taken into account can distort identification of pilot parameters. Much previous research has focused on linking workload to some of these identified or assumed parameters (pilot lead, time delay, open-loop stability), but the problem of such an approach - besides using parameters that might be incorrect due to the unaccounted-for nonlinearities, and can undergo significant change over the course of a maneuver - is that workload appears to be influenced by many of these parameters. Any objective metrics are inevitably confounded except for very restricted cases that do not carry over to the generalized scenario.

The work documented in this paper represents the initial steps to developing a novel, simple objective metric for assessing pilot workload and its variation over time across a wide variety of tasks. Additionally, it offers a tangible, easily implementable methodology for anticipating a pilot's operating parameters and workload, that could also be used as an effective design tool. The model shows promises in being able to precisely predict the actual pilot settings

and workload, and observed tolerance of pilot parameter variation over the course of operation. Finally, an approach is proposed for generating Cooper-Harper ratings based on the workload and parameter estimation methodology. This study used data from a piloted simulation experiment that was designed to examine factors affecting pilot's cognitive workload associated with vehicle control.

II. Experimental Protocol and Results

A. Protocol

Various command vehicle dynamics (proportional, rate, acceleration, jerk), vehicle gains (vehicle sensitivity to input), and display gains (display sensitivity to error) were used with a lateral station-keeping using a compensatory display, where a random forcing function continuously disturbed ownship's position.

Figure 1a shows a schematic of the station-keeping task and the display, pilot and vehicle components of the closed-loop system. Figure 1b gives the range of conditions within each component that were tested. The jerk condition for the vehicle dynamics (fourth condition listed for Y_V), contains a pole p whose location was varied. Twenty-three display configurations were tested with each subject using various combinations of the conditions shown in Figure 1b. Since pilot proficiency with any test condition was not a factor in this experiment, pilots were given two practice of each vehicle dynamic type (proportional, rate, acceleration, jerk) prior to testing.

One may be tempted to think that the display gain and the vehicle gain are effectively interchangeable and the same gains from the pilot's vantage point. The following example will serve to disprove this common misnomer. If the pilot's input is zero, the disturbance is perceived through the display gain – the vehicle gain does not come into play at all. Based on his/her control activity and quiescence, a pilot learns to decouple the effects of the display gain from the vehicle gain – thus decoupling aircraft motion due to disturbance from pilot-commanded vehicle motion.
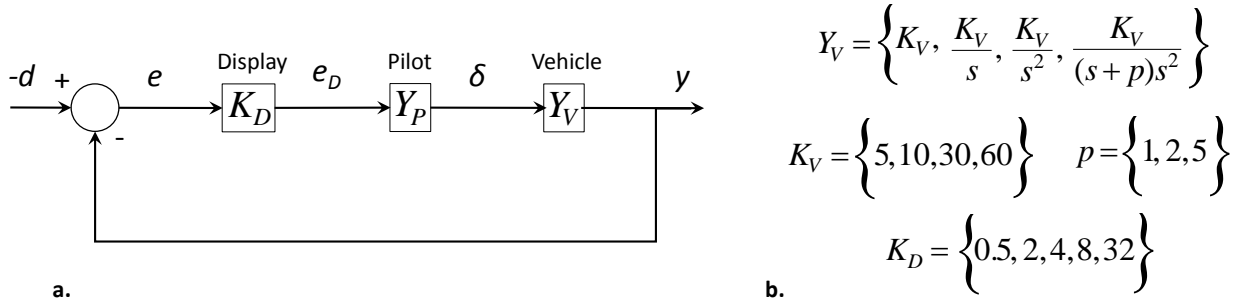

$$Y_V = \left\{ K_V, \; \frac{K_V}{s}, \; \frac{K_V}{s^2}, \; \frac{K_V}{(s+p)s^2} \right\}$$

$$K_V = \{5,10,30,60\} \qquad p = \{1,2,5\}$$

$$K_D = \{0.5,2,4,8,32\}$$

Figure 1. Station-keeping task. *a) Pilot, display, and vehicle elements; b) Range of conditions for display and vehicle elements.*

The components of the pilot element *Yp* of Figure 1 are shown in Figure 2. Pilot visual noise is added to the displayed error (Ref. 3), the sum is operated on by the Crossover Model (CM, Ref. 3) pilot. The output of this is then sent through a limiter.
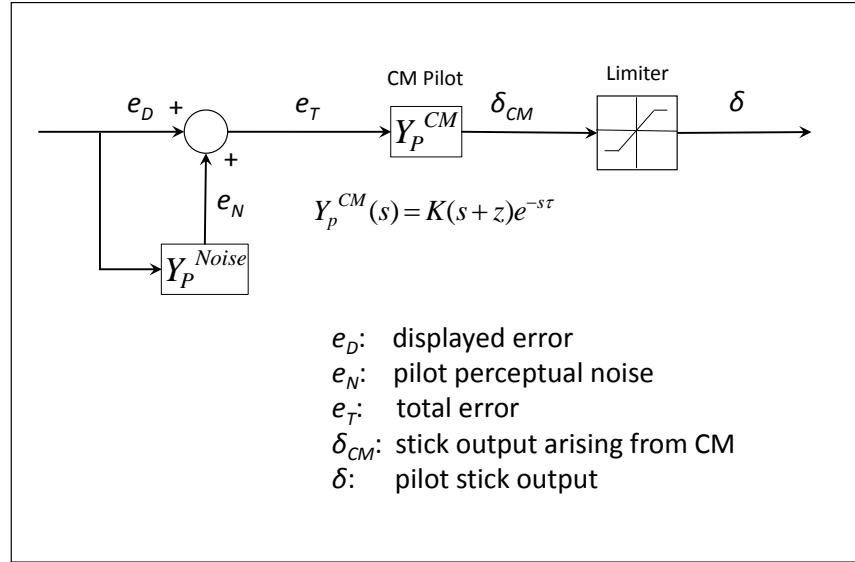
Figure 2. Components of Pilot element Y_p shown in Figure 1a.

Four male participants took part in the study. Three were Experimental Test Pilots (graduates of Navy Test Pilot School) with 1,900, 1,900, and 2,450 rotary wing flight hours. The fourth participant had logged 800 hours of rotary wing flight time. Ownship error relative to the hover location was displayed on a laptop monitor, and the pilot attempted to minimize the error using a gaming joystick. The Bedford rating scale (Ref.9) was used to subjectively score each pilots spare capacity at the end of each 60-second tracking run. Dependent variables were: stick position, rate and acceleration, stick position reversals, display error, rate, and acceleration. The positional disturbances imposed on the helicopter were designed to be both realistic and a diagnostic probe for pilot control behavior. Composed of a sum of 11 non-harmonically-related sine waves (the amplitude spectrum is shown in Figure 3a), the disturbance was perceived by the pilot as a random process – the result, however, was that the pilot's control response power resided largely at the same frequencies contained in the input disturbances. The disturbance time history is shown in Figure 3b.
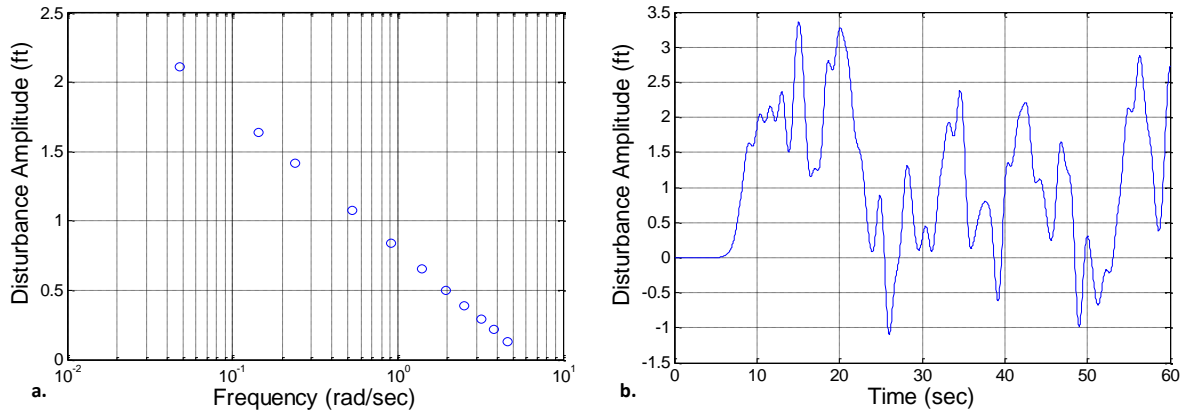

Figure 3. Sum-of-sines positional disturbance. *a) Amplitudes spectrum; b) Summation of 11 sine waves.*

B. Results

Before presenting the key results, the data are analyzed using two quantitative HQR metrics previously proposed by Lampton & Klyde in 2011 and 2012 respectively: cutoff frequency (Ref. 7) and power frequency (Ref. 8). Figure 4a shows cutoff frequency as a function of the actual Bedford ratings, where the coefficient of determination is effectively zero. Figure 4b shows power frequency vs. Bedford rating, and the coefficient of determination is 0.51. Coefficient of determination provides a measure of how well observed outcomes are replicated by a linear model,

American Institute of Aeronautics and Astronautics

based on the proportion of total variation of outcomes explained by the model (Ref. 10). While it can be argued that these metrics were designed to be used with Cooper-Harper HQR scale, the fact that the metrics are based solely on stick output should make them more applicable for use with Bedford since Cooper-Harper reflects pilot compensation to maintain a given level of performance. Neither cutoff nor power frequency accounts for performance.
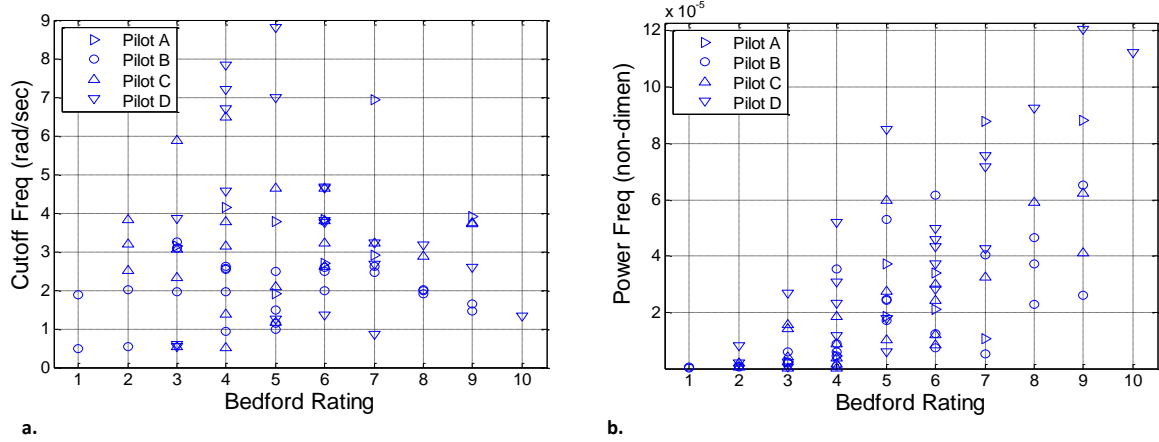

a.   **b.**

Figure 4. Comparison of previously proposed handling qualities metrics with experiment's Bedford ratings (coefficient of determination given in parentheses). *a) Cutoff frequency (0.01); b) Power frequency (0.51).*

In this station-keeping experiment the variables that correlated best with subjective rating were stick rate and displayed error rate. Figure 5a shows root-mean-square (RMS) stick rate versus rating (it has a coefficient of determination of 0.51, which matches that of power frequency in Figure 4b), and in Figure 5b a clear trend is seen between the RMS display error rate and subjective rating. Figure 5c gives the operations conducted on stick position and display error to produce the Bedford estimate. An empirically-derived relationship using these two measures yields the 92 data points shown in Figure 5d, where the coefficient of determination is 0.89.
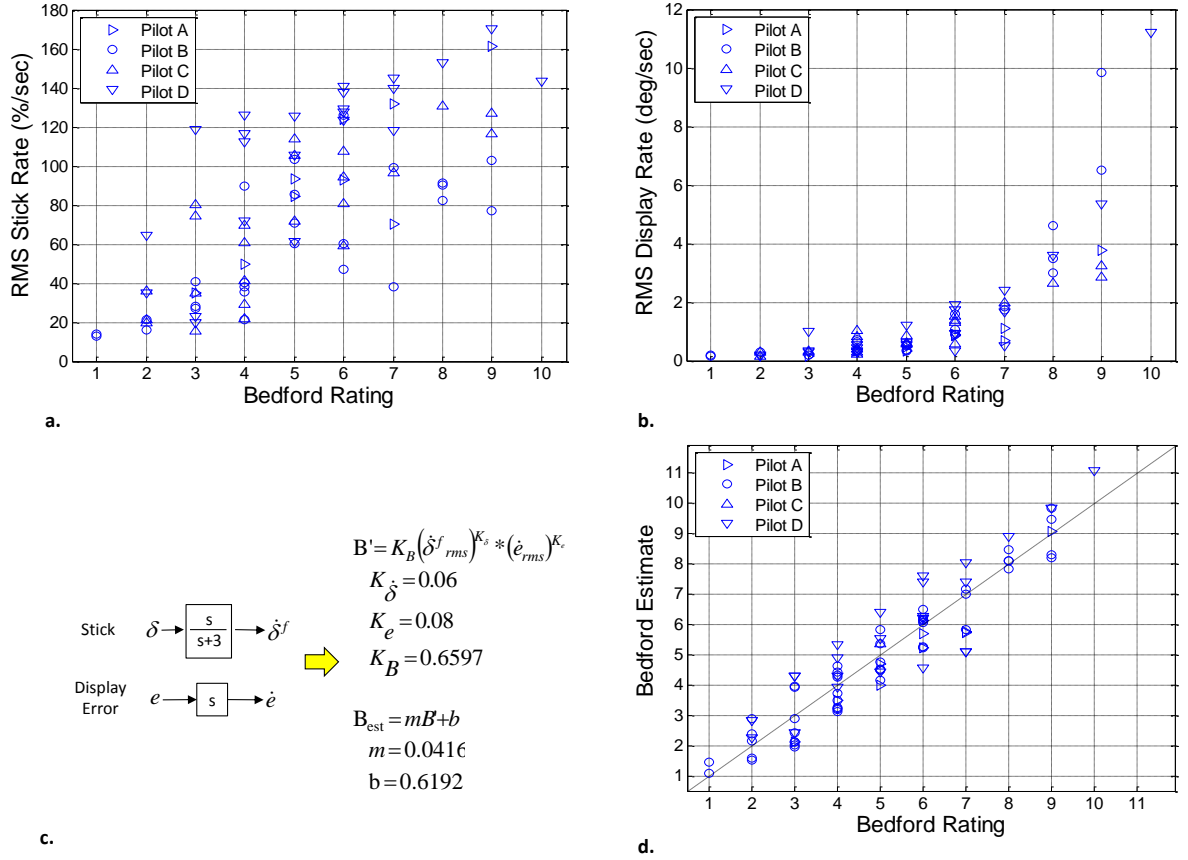
Figure 5. Elements of Bedford Estimation (coefficient of determination given in parentheses). *a) RMS stick rate (0.51); b) RMS display error rate; c) Computation of Bedford Estimate; d) Bedford Estimate (0.89).*

The Bedford rating scale is based on descriptors and a decision tree similar to what the Cooper-Harper HQR uses, and likewise is ordinal. The Bedford rating scale allows the ordering of workload from insignificant (1) to uncontrollable (10). It would be desirable if the rating was interval, allowing differences to be measured as a continuum. In Ref. 12 Mitchell and Aponso demonstrated that the use of the ordinal Cooper-Harper HQRs as if they were interval was both appropriate and reasonable. Given the similarities between the Bedford and Cooper-Harper ratings, it is assumed that the same argument for treating the Bedford rating as interval can be made.

To test reasonableness of the Bedford estimator it was applied to helicopter multi-axis flight. Sixty-seven helicopter approaches – in actual degraded visual environment (DVE) flight - were executed by five Army evaluation pilots using a head-down display to track a commanded approach profile. Upon completion of an approach the pilot provided a Bedford rating. Each control axis (lateral cyclic, longitudinal cyclic, collective, pedals) was normalized by the largest displacement encountered in all the data. Lateral and longitudinal cyclic motion were combined since the display presented error to the pilot as a single distance and direction. Display error was converted to degrees as viewed from the pilot's distance. The Bedford estimate for each control axis (cyclic, collective, pedals) was computed over a sliding window of eight seconds, and the maximum of the three estimates at every time increment was used to create the combined history shown at the bottom of Figure 6a. The mean of the entire flight segment of interest gives the estimate of the overall rating for the segment. Figure 6b gives the distribution of ratings (% of total collected), Figure 6c shows the data and the best-fit line through them, and Figure 6d plots the mean and standard deviation bars of data overlaid on the line representing ideal correspondence between the estimated and actual pilot rating.

Given that the Bedford estimator was developed from a single-axis simulation task experiment, and that the DVE flight test conditions presented the pilots with multi-modal cueing (tactile, audio) and a mixed visual environment (head-down display and a distracting out-the-window scene of actual blowing dust), these preliminary results are encouraging. Applying a linear correction (based on the best-fit line intercept and slope) to the original Bedford estimator results, Figure 7b shows the means of the revised data overlaying well with the idealized line. It is possible

American Institute of Aeronautics and Astronautics

that the linear skewing is a function of the number of axes controlled, or arises due to some or all of the cueing factors just listed. Nonetheless, with a simple linear correction the proposed Bedford estimator appears to be effective for this complex task in the UH-60.
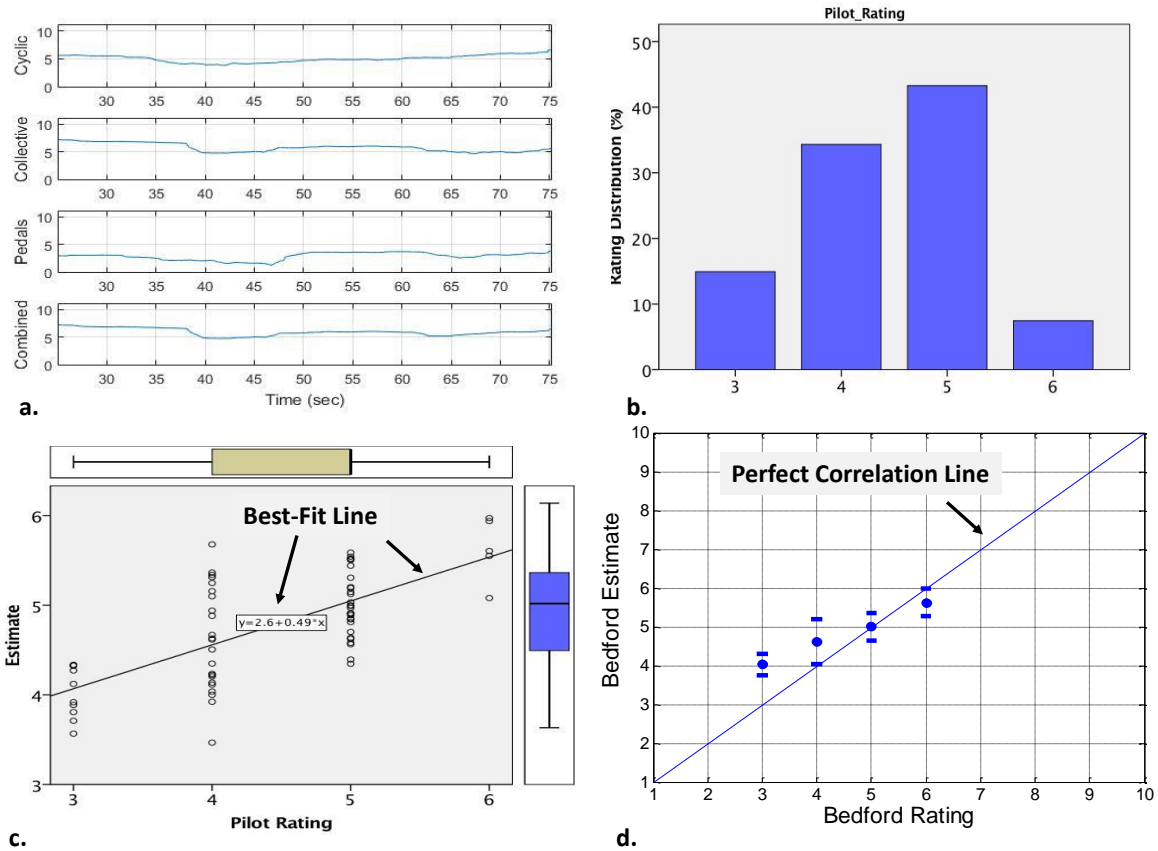

Figure 6. Application of Bedford estimator to multi-axis helicopter flight. *a) Time history of Bedford estimates for cyclic, collective, pedals, and aggregate; b) Distribution of ratings collected (%, sample size was 67 flights); c) Best-fit line through the data; d) Mean and standard deviation bars of data overlaid on the line representing ideal correspondence between the estimated and actual rating.*
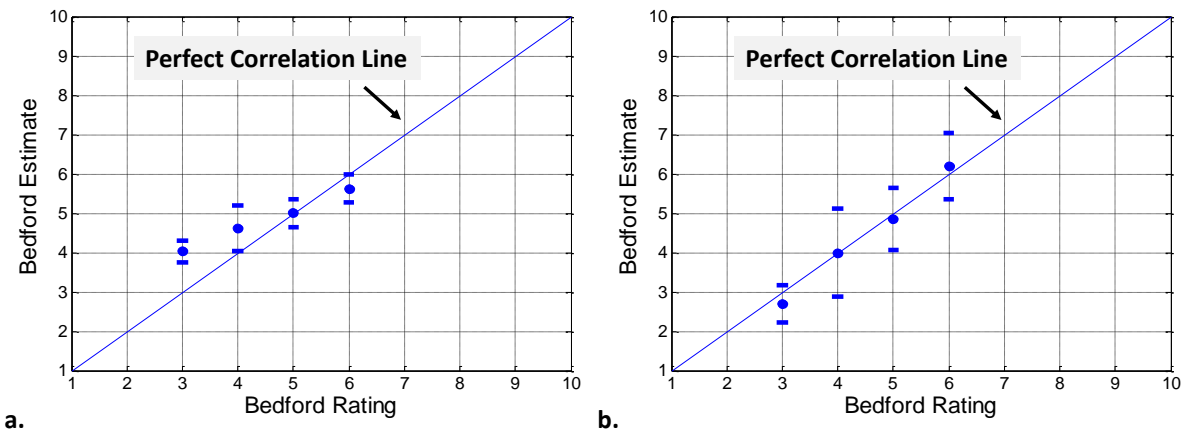

Figure 7. Bedford estimator flight test results. *a) Original data, shown in Figure 6c; b) Data after applying linear correction.*

C. Amplitude Clipping

Large differences can arise between the response predicted by McRuer's *CM* and actual pilot data, especially when the vehicle being controlled has dynamics that require significant lead compensation by the pilot. A nonlinear pilot control technique, observed and coined by the authors as 'amplitude clipping', occurs when a pilot limits his/her control input to some fixed amplitude over a short period of time rather than responding continuously to the error signal. Thus the pilot responds to an error similar to what the *CM* predicts up to a certain stick amplitude that remains fixed until the error signal reverses and returns, at which time the pilot resumes active tracking. The amplitude at which the control input is capped can vary over time.

Amplitude clipping lowers the effective gain of the pilot, as it is the equivalent of saturation. Ref. 11 graphs the gain attenuation for a sinusoid of amplitude A that saturates at amplitude a, shown in Figure 8.
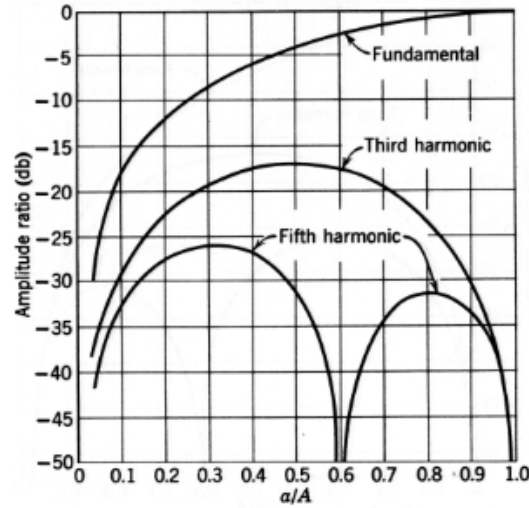

Figure 8. Sinusoidal describing function for saturation (Ref. 11).

Figure 9 compares the pilot's actual stick response to a simulated pilot stick response generated by the pilot model, and clips this stick signal using the actual pilot's stick min/max segments as a template. The pilot parameters associated with the *CM* (time delay, gain, and lead) used in the pilot model are given in Eq. (2). Actual pilot stick response and station-keeping error are shown in Figure 10. In Figure 10a the pilot is seen to clip at approximately 50%.
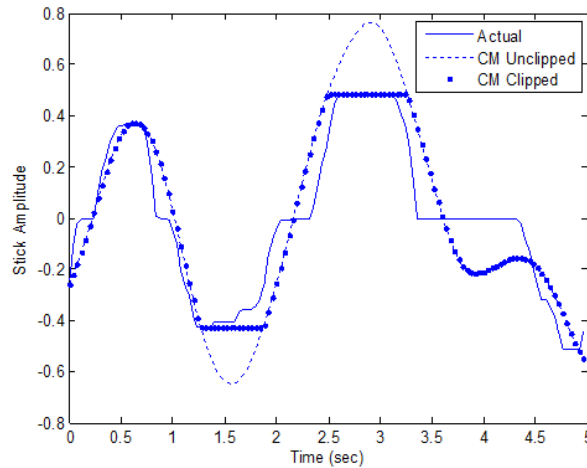

Figure 9. Comparison of actual stick data, unclipped Crossover Model (CM) response, and clipped CM response.

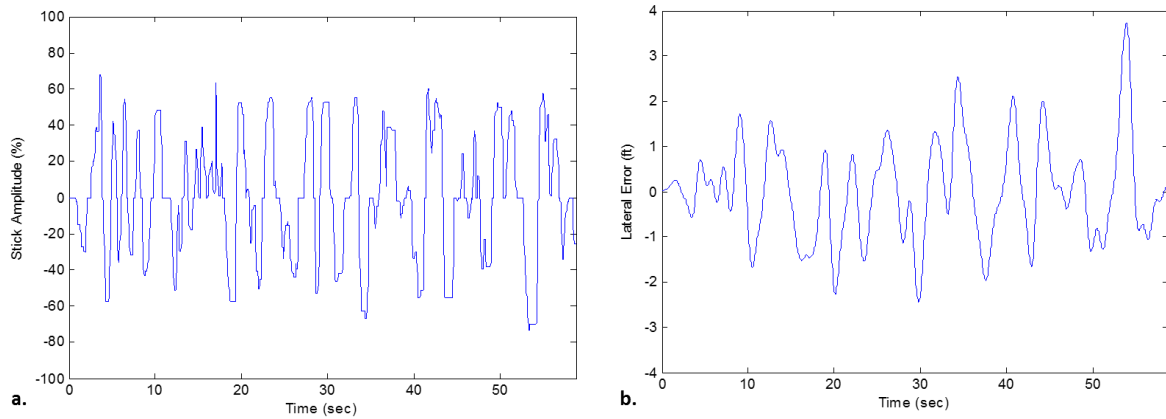$$Y_P(s) = K(s + z)e^{-s\tau} \tag{2}$$

Figure 10. Actual pilot data from station-keeping task. *a) Stick position; b) Lateral error.*

D. Computation of Pilot Parameters

Using the clipped simulated pilot response, the three parameters of the pilot model are iteratively varied and filtered until a best match is computed between the actual and simulated/clipped stick. This parameter identification is conducted over a sliding time window, allowing near real-time measurement. In Figure 11 the values of the pilot parameters used in a simulation of the pilot were fixed (denoted as truth by dash-dot lines) over the course of the run. The solid line shows the identification when clipping is accounted for, and the dashed line denotes the identification when clipping is not taken into account. In all three cases accounting for clipping produced very good matching with the actual parameters. In Figure 11a the non-clipped gain falls to almost 50% of the true CM gain – this represents the effective reduction in open-loop gain due to clipping. Non-clipped time delay has a transient excursion from truth, and non-clipped lead rises to twice the actual value before decreasing. Thus if clipping is not considered, the values computed can be misleading as to what is actually occurring in the pilot.
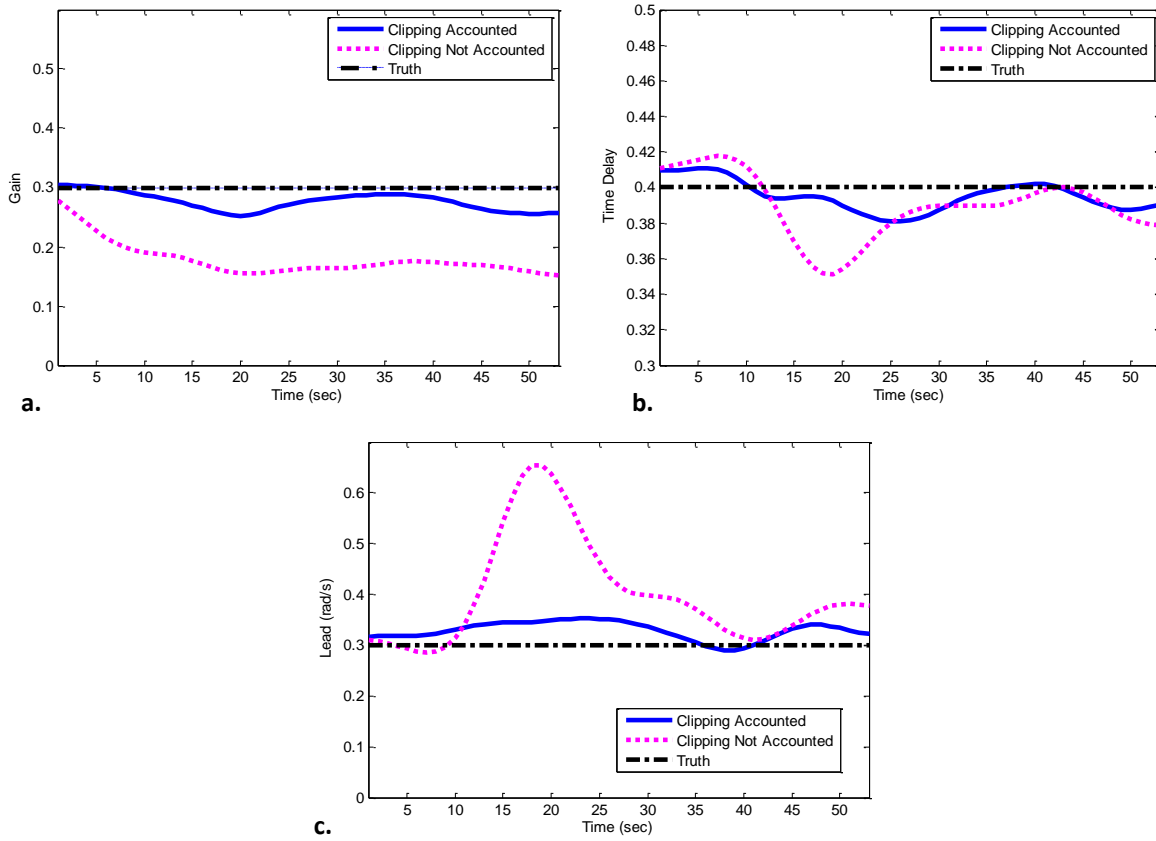
American Institute of Aeronautics and Astronautics

Figure 11. Effect of accounting/not accounting for clipping on pilot parameter identification . *a) Pilot gain; b) Pilot time delay; c) Pilot lead frequency.*

To test the method's efficacy during dynamic pilot changes, parameters of an actual pilot were identified over time. These parameter time histories drove the pilot simulation that was used in Figure 11, and in turn the parameters identified from the simulation were compared with the parameters originally obtained from the pilot data. After shifting the observed histories in time by half of the sliding time window that was used (the window was eight seconds, the average lag would be half of this), Figure 12 shows near-perfect correspondence between the original and recovered pilot values.
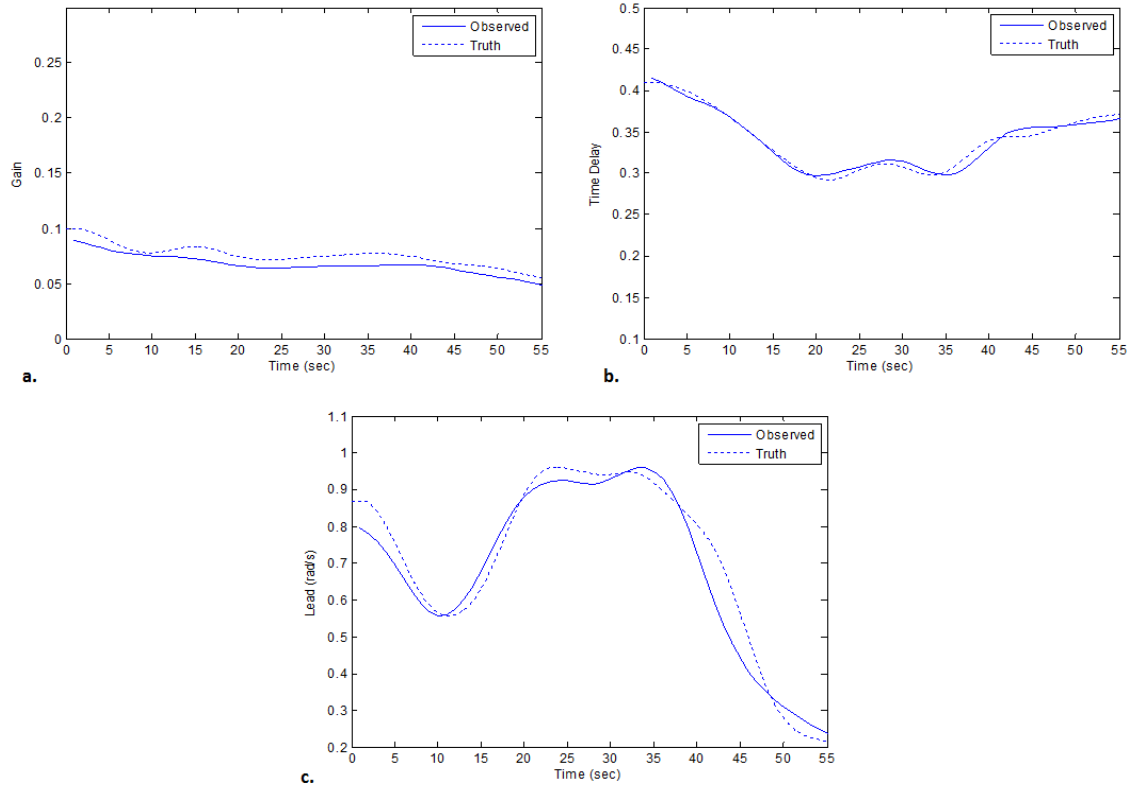
Figure 12. Pilot parameters identified and cross-checked. *a) Pilot gain; b) Pilot time delay; c) Pilot lead frequency.*

Figure 13 defines clipping area as the portion of the *CM* stick output that is clipped relative to the total *CM* stick output. Figure 14 presents time histories of the computed Bedford estimate and percentage clipping area for a low and a high workload run. It appears that the workload slightly dips when clipping increases, and increases when clipping is reduced. When a pilot employed clipping, mean clipping area was observed to correlate with the overall workload rating. McRuer had noted that a rise in pilot lead generation was accompanied by an increase in pilot perceptual 'noise' that was uncorrelated with the error, which he called 'control remnant'. The preliminary data from this experiment indicates that part of this remnant can be due to performance-enhancing nonlinear pilot techniques such amplitude-clipping and pulsing.
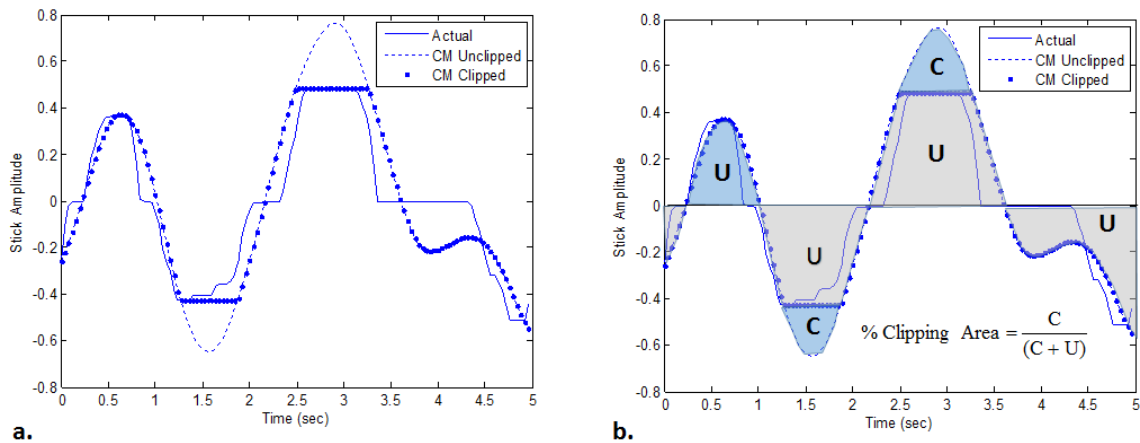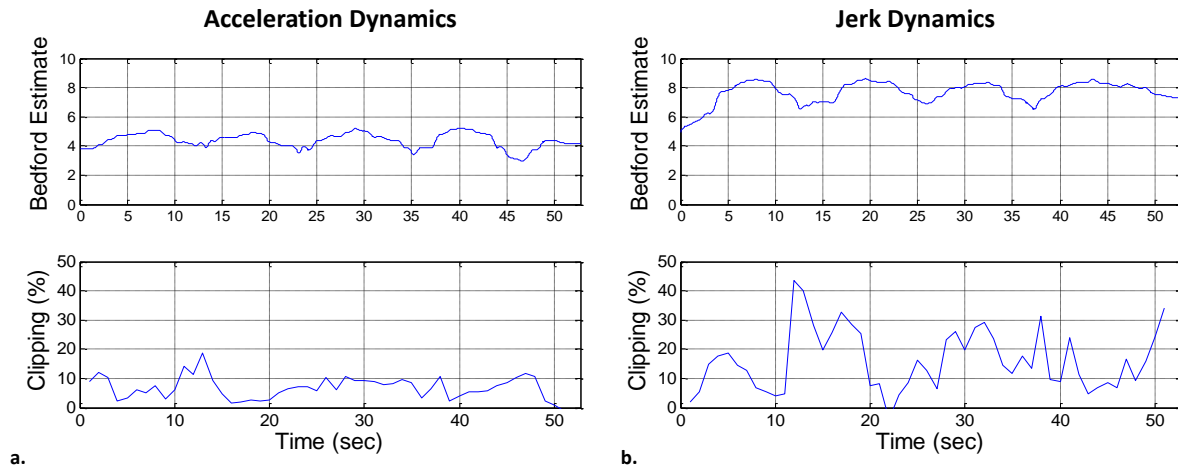

Figure 13. % Clipping area defined.

11

Figure 14. Bedford estimate and clipping area time histories. *a) Acceleration dynamics (overall Bedford rating 4); b) Jerk dynamics (high workload, overall Bedford rating 8).*

The effects of pilot technique and changes in pilot parameters are now illustrated. Figure 15 shows the simulated time responses and bode plots associated with increasing pilot time delay from 0.4 seconds to 0.9 seconds. The response for 0.9 seconds delay when amplitude clipping is present (Figure 16) gives an RMS error that is approximately half the RMS error for when continuous tracking is employed (for the no clipping case, stick output is limited to 100% when it reaches the limit of throw). In Figure 17 and Figure 18 the pilot gain is doubled for the no clipping and clipping cases. Note that the nominal pilot gain was set to 0.3 for the clipping examples vs. 0.24 with the continuous tracking examples in order to give approximately the same crossover frequencies for both pilot techniques. Table 1 compares these conditions, with clipping consistently producing the lower Bedford rating. While a pilot would not be expected to persist in maintaining parameters that give poor performance, these examples serve to show how amplitude clipping would allow a system to absorb transient excursions in pilot parameters and permit continued satisfactory performance.

The oval points in the bode plots of Figure 15 - Figure 18 were computed at the frequencies of the sum-of-sines (see Figure 3) using power spectral density (psd) ratios of the display input (aircraft positional error) and the aircraft position response due to pilot control.

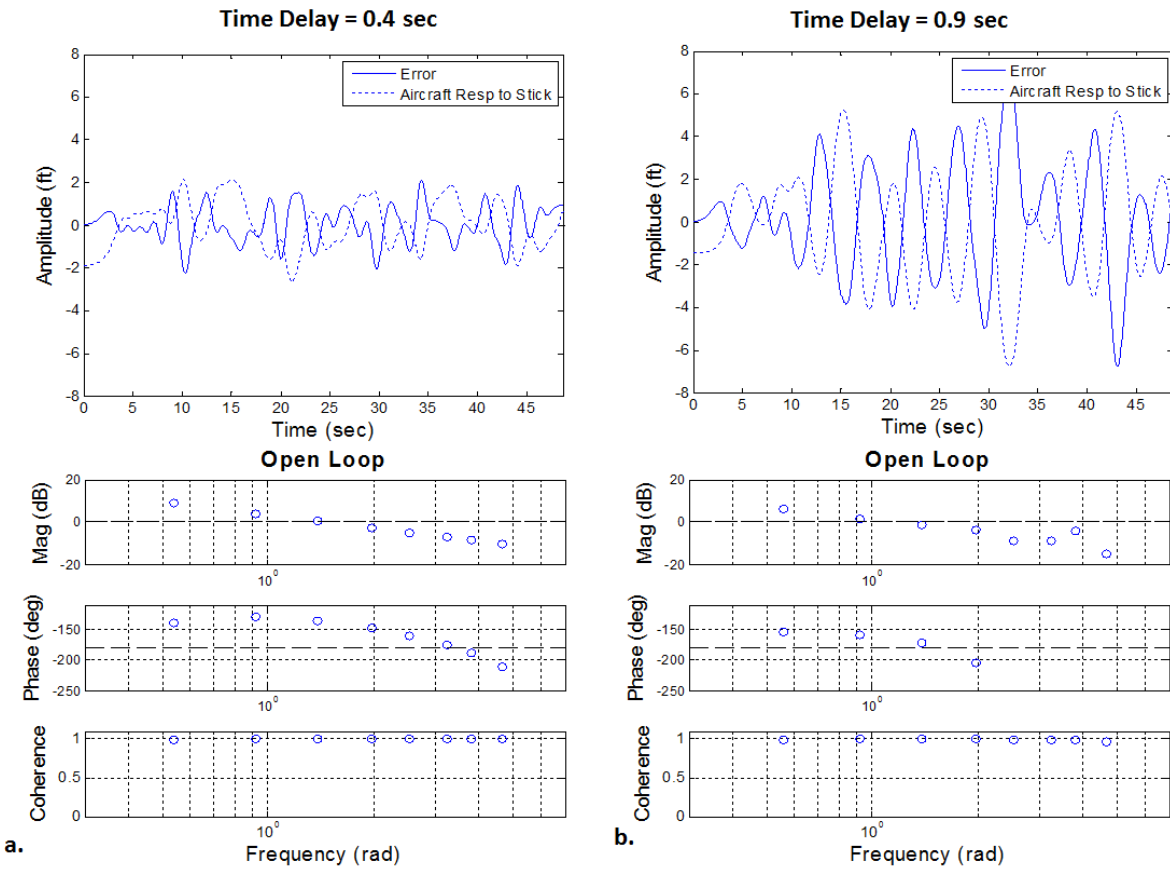American Institute of Aeronautics and Astronautics

No Clipping

Time Delay = 0.4 sec

Time Delay = 0.9 sec

Figure 15. Effects of increased pilot time delay (simulated, no clipping). *a) 0.4 second delay; b) 0.9 second delay.*

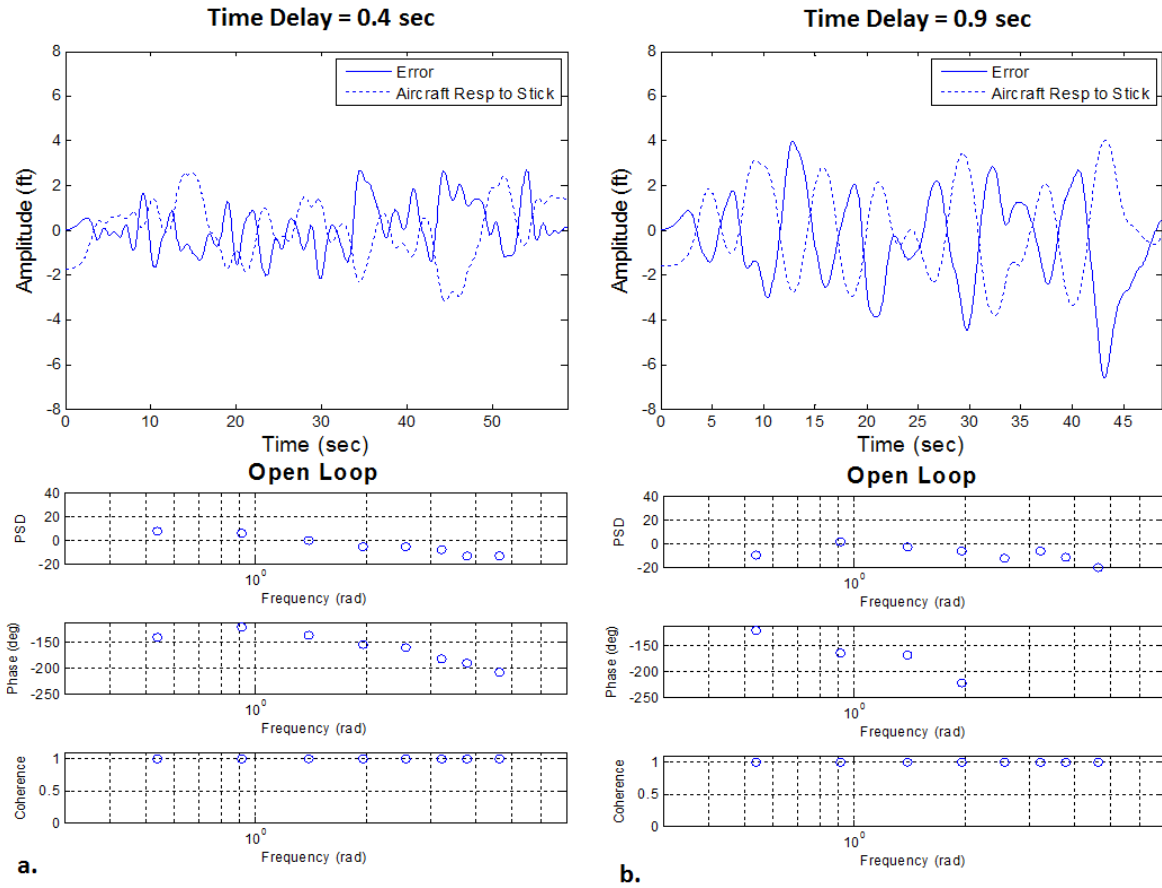# With Clipping

Time Delay = 0.4 sec

Time Delay = 0.9 sec

Figure 16. Effects of increased pilot time delay (simulated, with clipping). *a) 0.4 second delay; b) 0.9 second delay.*

No Clipping

Pilot Gain = 0.24

Pilot Gain = 2*0.24

a.

b.

Figure 17. Effects of increased pilot gain (simulated, no clipping). *a) Nominal pilot gain for a crossover frequency of 1.4 rad/sec; b) Nominal gain doubled.*
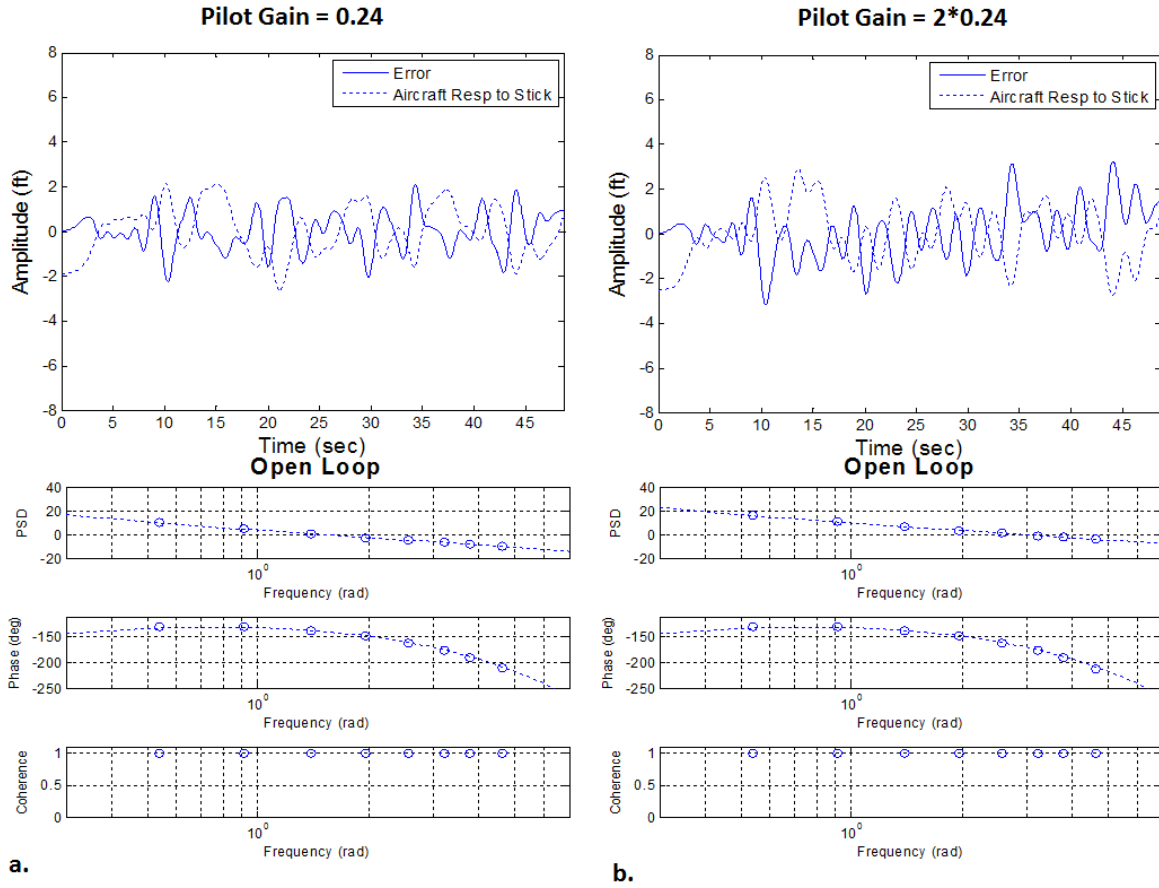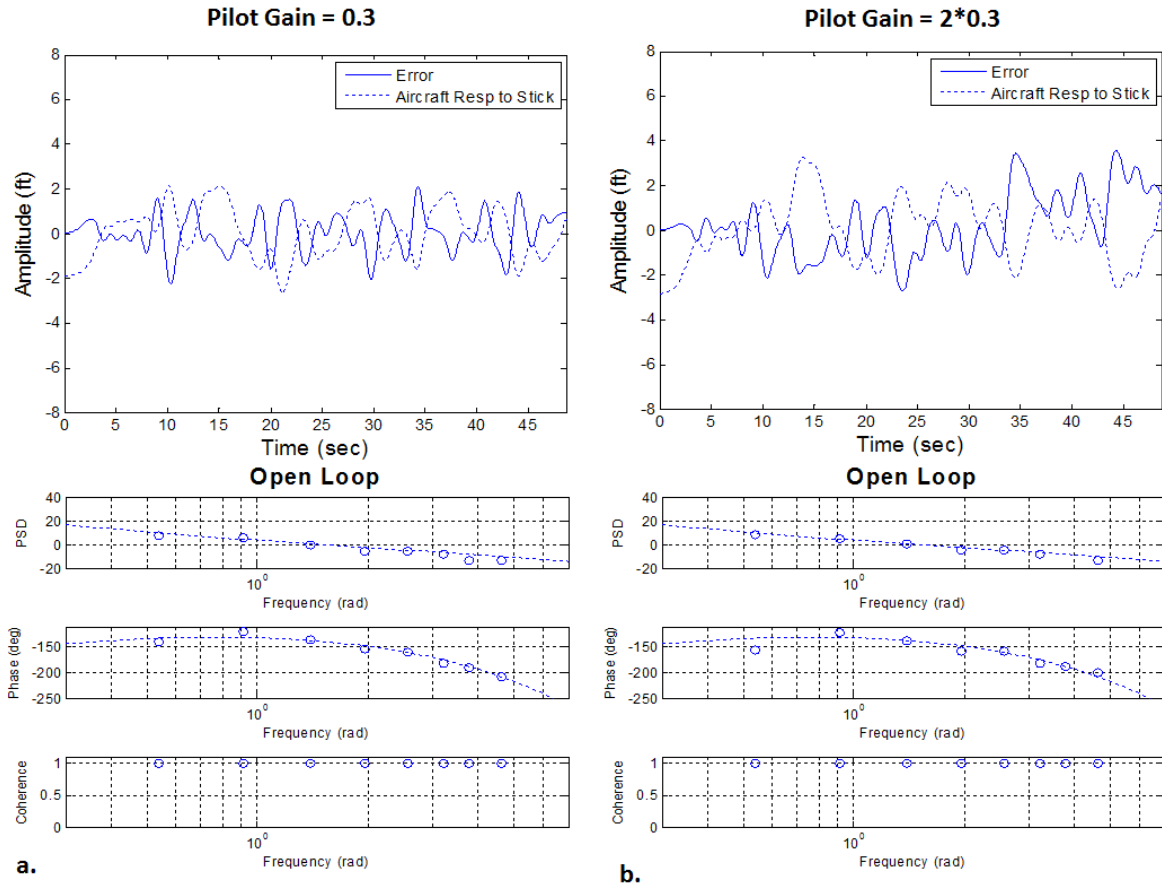
Figure 18. Effects of increased pilot gain (simulated, with clipping). *a) Nominal pilot gain for a crossover frequency of 1.4 rad/sec; b) Nominal gain doubled.*

Table 1. Comparison of RMS error and Bedford estimates with and without clipping.

τ (sec)	K	Clipping	RMS Error (ft)	Bedford Est
0.4	0.2	no	0.9	6.5
0.4	0.2	50%	1.0	6.1
0.4	0.4	no	1.2	8.0
0.4	0.4	50%	1.3	6.7
0.9	0.2	no	2.2	7.7
0.9	0.2	50%	1.6	6.4

The pilot model shown earlier in Figure 2 was used with and without the presence of pilot perceptual noise, and subjective ratings using the Bedford estimator were generated. Figure 19 shows the % increase in RMS error and associated % decrease in Bedford rating (referred to as an elasticity analysis in business) for both amplitude clipping and continuous tracking. The curve for each technique started with the same point (same gain, no clipping), yielding an initial RMS error and Bedford estimate. To generate the clipping curve, clipping was progressively increased with gain held constant, and the percent changes in RMS error and Bedford estimate relative to the starting values produced the ordinate and abscissa pairs. To generate the continuous tracking gain adjustment curve, gain was decreased (no clipping used). Figure 19a indicates that when perceptual noise is absent, then for the same decrement

of performance the task would be easier using the continuous tracking gain reduction compared to when clipping reduces gain. However, with the addition of perceptual noise, Figure 19b shows that amplitude clipping can generate a faster decrease in workload than continuous tracking for the same increase in RMS tracking error. For this example, clipping amplitude at 50% of full stick motion results in less workload than if continuous tracking were used to generate the same RMS error (that 50% amplitude clipping produces). Another important advantage clipping offers is the opportunity for the pilot to scan other visual cues or events in the environment while the clipping is occurring. Furthermore, it appears that clipping enhances system robustness to transient changes in pilot time delay and CM gain, which could allow more relaxed operation.
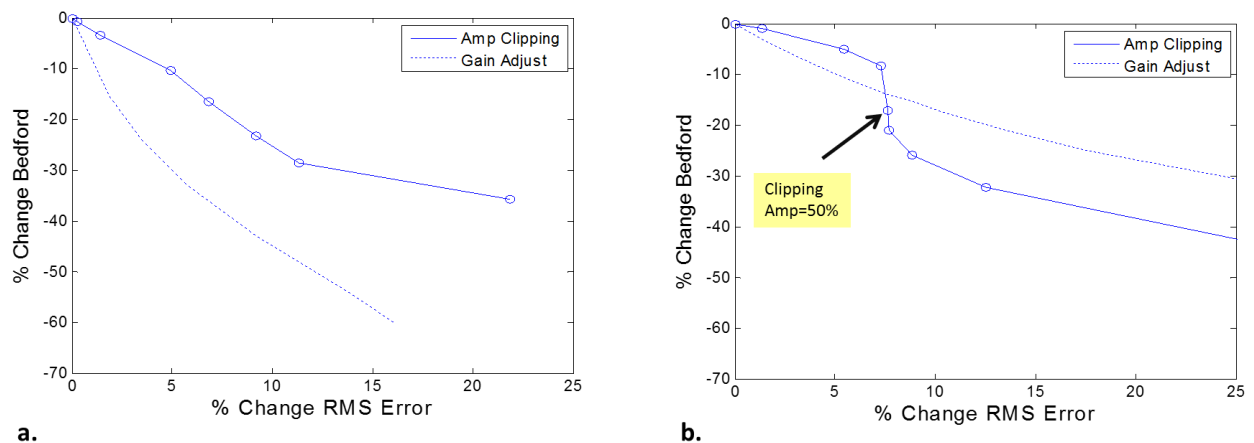

a.   b.

Figure 19. Elasticities using amplitude clipping and continuous tracking gain adjustment (acceleration dynamics). *a) Without pilot perceptual noise; b) With pilot perceptual noise.*

E. Estimation of Pilot Lead and Gain Selection

In Figure 20 the pilot parameters for actual runs were tracked for two different vehicle systems: acceleration and jerk-command (the jerk command had a pole at 2 rad/sec). Note that the pilot's lead for the lower-workload system (acceleration command, given an overall Bedford rating of 4 for this run) undergoes a significant change, whereas for the jerk system (given a Bedford rating of 8) the lead stayed essentially constant. Similar trends in lead were observed for other pilots flying the two vehicle types.
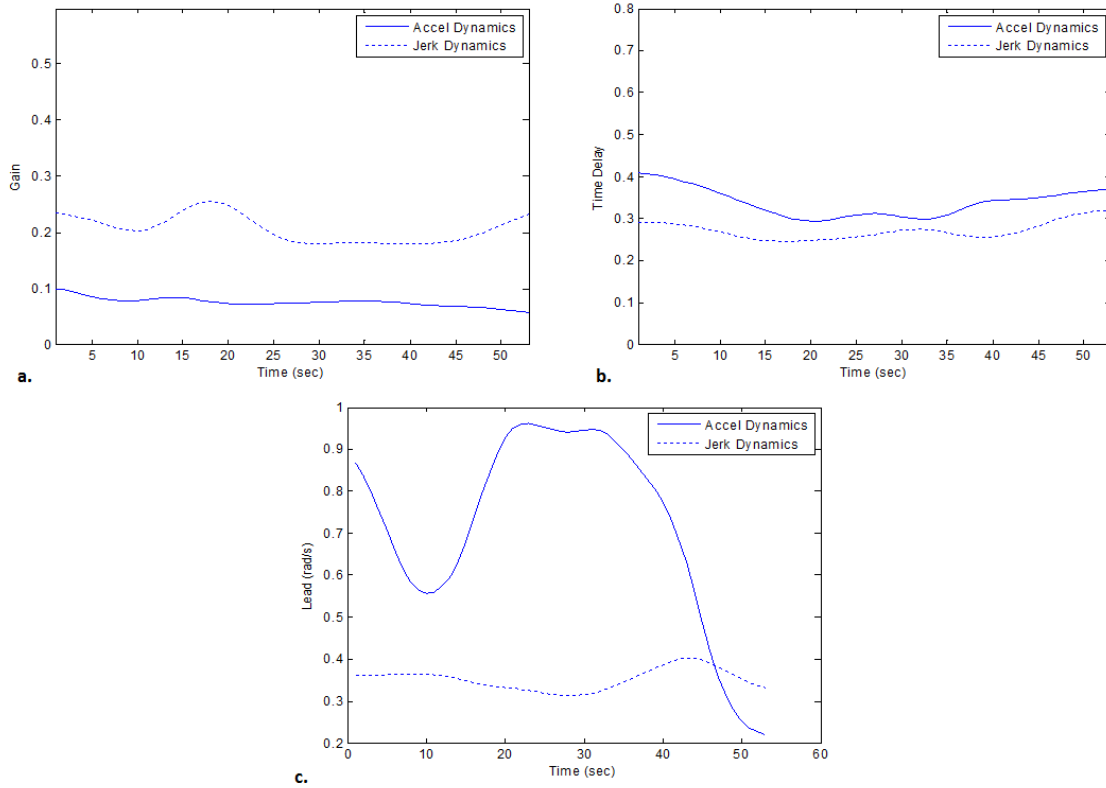
Figure 20. Comparison of identified pilot parameters histories when pilot is controlling acceleration and jerk dynamics. *a) Pilot Gain; b) Pilot time delay; c) Pilot lead frequency.*

The question as to what drives an operator's gain and lead settings is now investigated. The approximate pilot gains that were identified for the acceleration and jerk systems shown in Figure 20 (0.1 and 0.2, respectively) were used in the pilot model, and pilot lead varied. Figure 21a shows that for the acceleration dynamics, aircraft position error is fairly insensitive to lead, whereas lead greatly affects position error for the jerk dynamics. Figure 21b plots Bedford estimates vs. lead, and in general lower lead frequencies (i.e. increasing lead compensation) yield lower workload for both systems. Note this appears to contradict the conclusions of Ref. 3, where pilot rating decrement is presented as increasing with the slope of pilot lead compensation at the crossover frequency (Figure 22a). In Figure 22b the results from one pilot (Ref. 12) likewise do not corroborate Ref. 3– all six pilot exhibited similar behavior. Indeed, there is an increase in rating decrement with increasing lag. Ref. 12 states that the vehicle dynamics associated with the lag and high workload (i.e. the left-side of Figure 22b) had a low-damped mode that forced the pilots to employ lag in order to avoid operation in the vicinity of the oscillatory mode. While in very general terms it can be said that the relationship plotted in Figure 22a is correct (i.e. a second order system is generally more difficult to control than first order dynamics), it is not sufficiently specific to allow workload prediction for many applications.

In the absence of specific performance objectives, as was the case for the experiment of this paper, it is reasonable to assume a pilot bases his/her operating parameters on some combination of performance and workload (these two variables form the basis for the Cooper-Harper HQR scale). Comparing Figure 21a and Figure 21b, it is not readily apparent how performance and workload might be each weighted in selecting the lead operating range for each system. For the purpose of creating a preliminary metric of 'undesirability', these two are multiplied together, since this yields trends that are consistent with tradeoffs between performance and workload. Figure 21c plots the undesirability metric against lead frequency. When controlling the acceleration dynamics, this plot indicates that the choice of lead frequency does not make much of a difference (although there should be some preference toward the lower frequencies). Referring back to Figure 20c, the pilot's observed lead is seen to span most of the frequencies tested in Figure 21. The time trace for lead in Figure 20c appears to show probatory behavior, where lead frequency goes from high to medium, back to high, and then makes a decisive descent toward 0.2 rad/sec. This appears to indicate that when a pilot parameter has minimal effect on the combination of workload

and performance, the pilot may exercise loose control over that parameter. Conversely, with the jerk dynamics, Figure 21c shows a relatively narrow range of lead values (0.25 – 0.40) which minimimize the workload/performance combination. Looking again at Figure 20c, this window does indeed bound the lead observed for the jerk system.
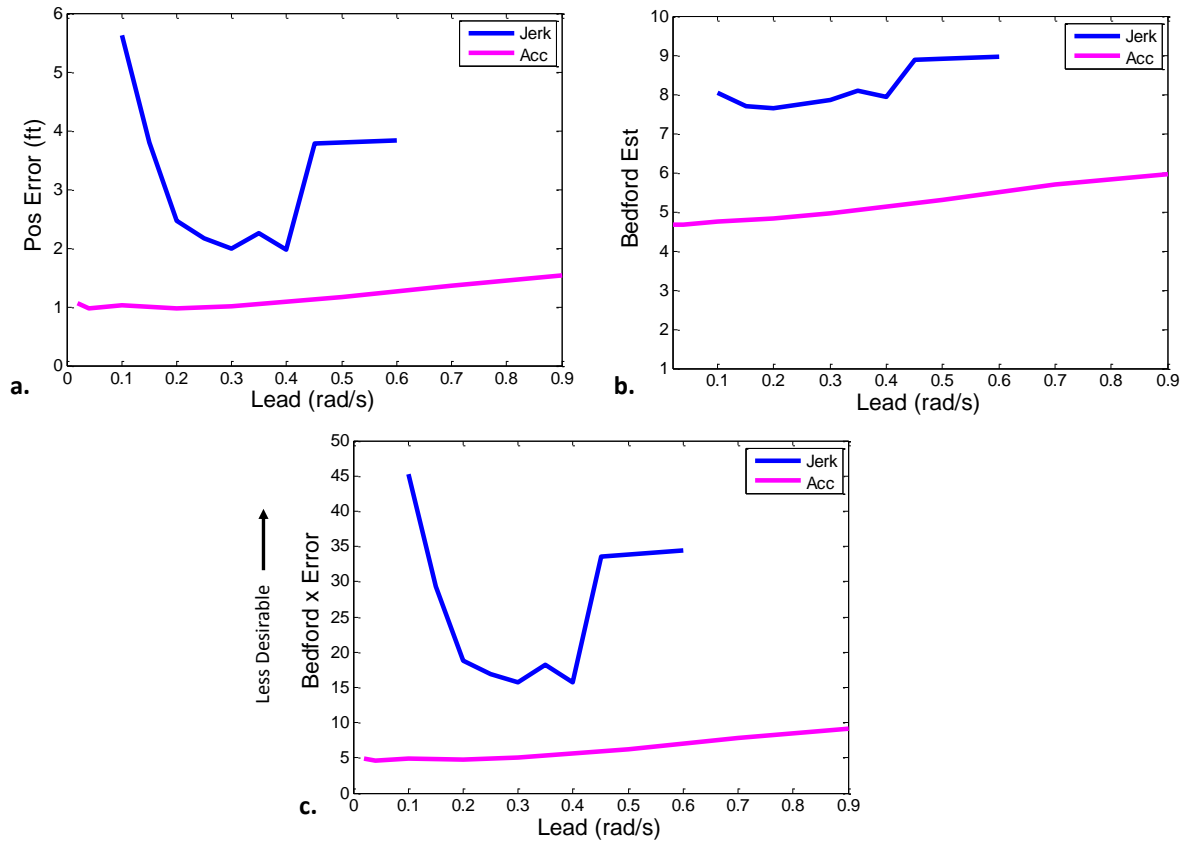

Figure 21. Modeled responses as a function of pilot lead frequency (acceleration and jerk vehicle dynamics). *a) Aircraft position error; b) Bedford estimate; c) Product of position error and Bedford estimate.*
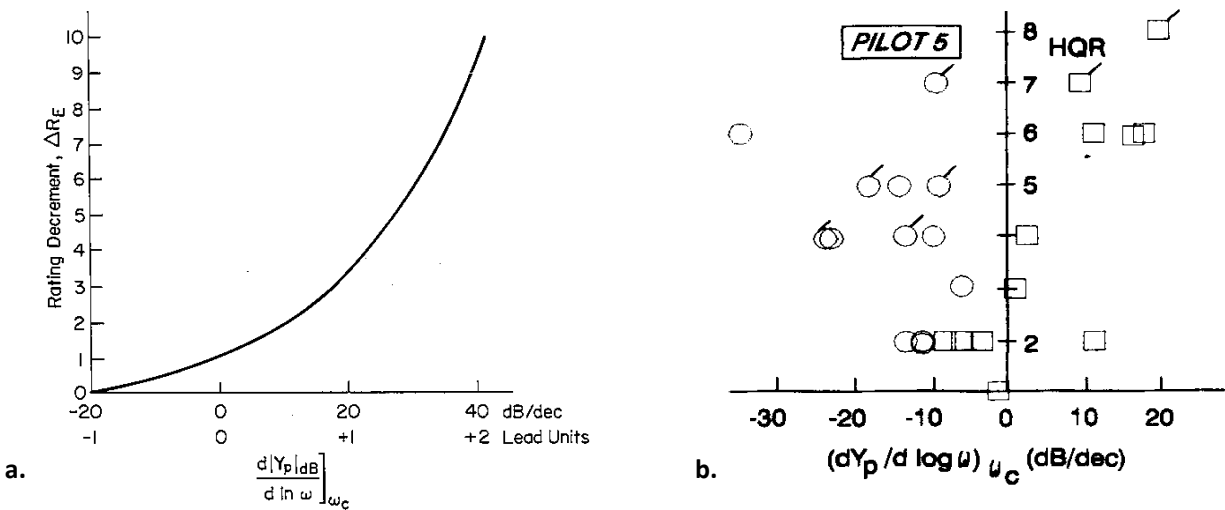

Figure 22. Conflicting observations of pilot rating (Cooper-Harper) decrement. *a) Ref. 3; b) Ref. 12.*

The same approach is employed to investigate favored regions of gain associated with the jerk system. In Figure 23 position error, Bedford estimate, and the product of the two is plotted for various gains across lead frequency. The gains 0.15 and 0.20 yield roughly the same position error (Figure 23a) in the region 0.25 − 0.40 rad/sec, but the the workload associated with the 0.15 gain in Figure 23b is roughly an entire point below the neighboring gain curve (from 0.25 − 0.35 rad/sec). In Figure 23c the workload/performance combination shows the undesirability curves for gains 0.15 and 0.20 to be the lowest of the four gain curves in the region of 0.25 − 0.40 rad/sec, which should indicate a preference for operating at these gains and lead frequencies (actual lead frequency agreed with this, as was discussed above). In Figure 20c the actual gain varied from 0.18 − 0.25, with the most time spent at 0.18. Thus it appears that when the pilot employs amplitude clipping, the pilot model presented herein can be used to anticipate both pilot lead and gain with good precision.
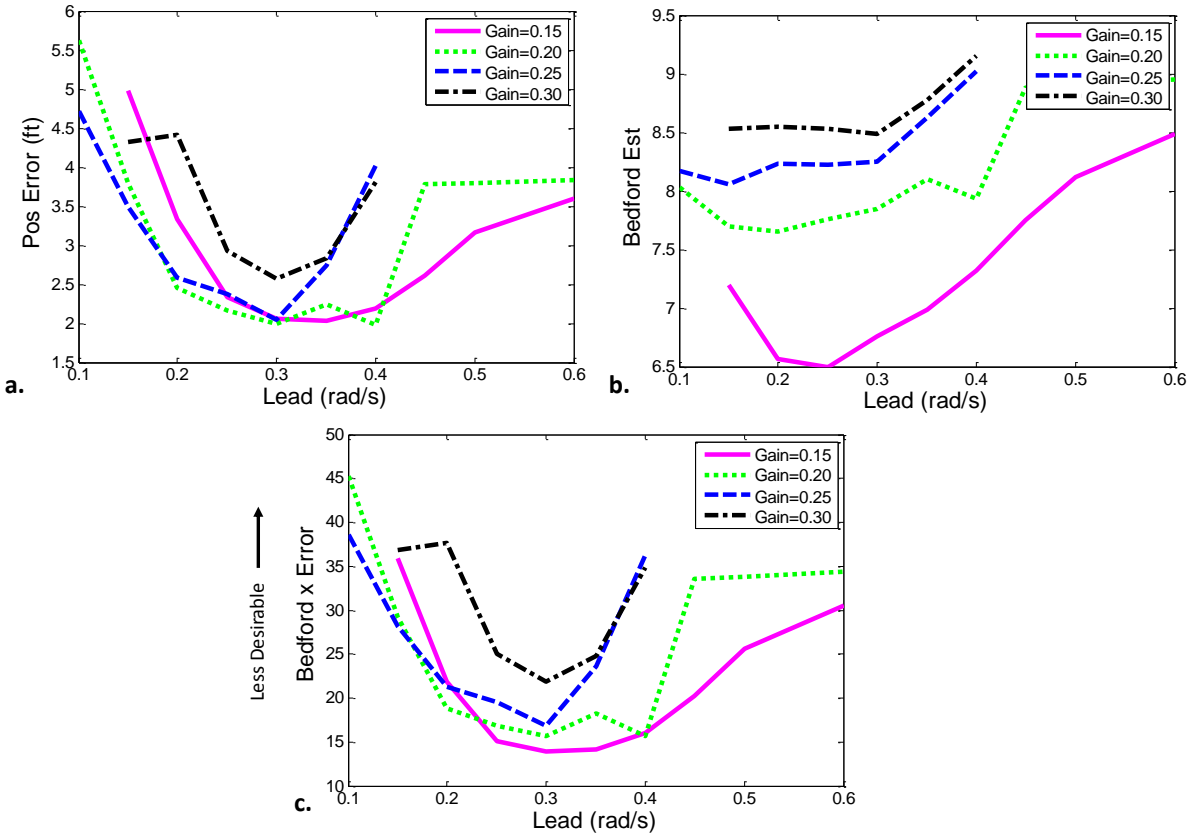

Figure 23. Modeled responses as a function of pilot lead lead and gain (jerk vehicle dynamics). *a) Aircraft position error; b) Bedford estimate; c) Product of position error and Bedford estimate.*

Now examining the effect of pilot technique on the pilot's operating parameters, Figure 24 compares the modeled response using continuous tracking and clipping during control of the acceleration dynamics. Continuous control yields better performance for lead frequencies greater than 0.2 rad/sec, but note that for all lead frequencies clipping produces less workload. Combining performance and workload in Figure 24c we see that above 0.2 rad/sec both control techniques yield approximately the same increase in undesirability, whereas below 0.2 rad/sec there is a rapid rise in undesirability when using continuous tracking. Assuming a pilot were skilled at both control styles, these results could be interpreted as follows. If the task demands minimum tracking error, the pilot can employ continuous tracking at the cost of slightly higher workload (½ a Bedford point compared to if clipping is used), and the lead will be restricted to frequencies at or above 0.2 rad/sec. If 10% more tracking error is acceptable, clipping can be employed and the Bedford rating will improve by ½ a point. Furthermore, lead frequency can range from almost zero on up. To determine the likely regions of gain operation (in Figure 24, the pilot gain for continuous tracking was 0.07, and for clipping 0.1), an analysis similar to what was done in Figure 23 could be conducted.
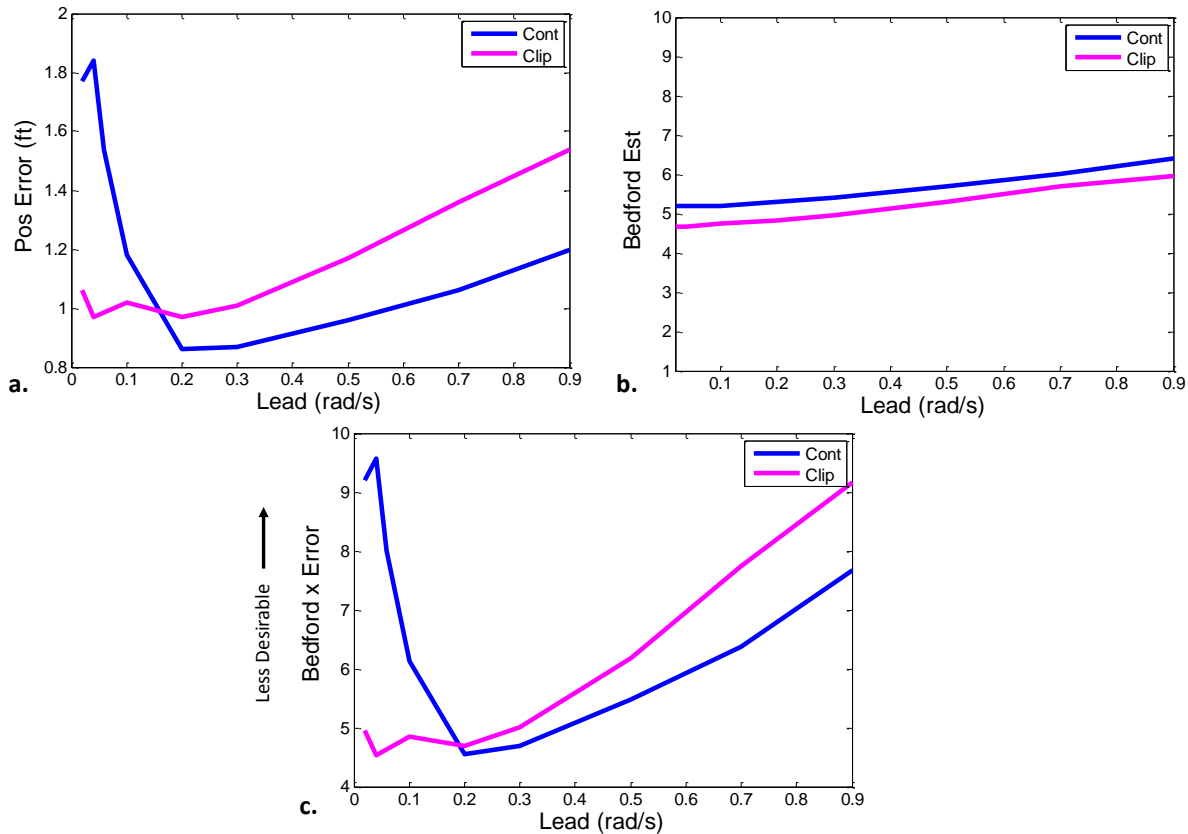
Figure 24. Effect of pilot technique on modeled responses as a function of pilot lead (acceleration vehicle dynamics). *a) Aircraft position error; b) Bedford estimate; c) Product of position error and Bedford estimate.*

F. Open-Loop Stability Margin Identification

Figure 25 defines key frequency domain metrics: for phase margin, gain margin, and crossover frequency.

- Crossover frequency, ω_c, denotes the system bandwidth (i.e. upper frequency that an input is effective at reducing tracking error).

- Phase Margin refers to the difference between phase that exists at ω_c and -180 degrees (how much phase can be lost and still maintain negative feedback)

- Gain Margin refers to the gain at -180 degrees (how much gain can be lost and still maintain negative feedback)
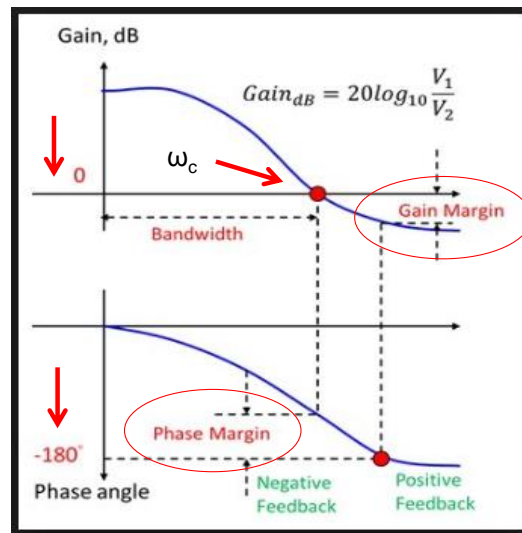

Figure 25. Definitions of crossover frequency, gain margin, and phase margin.

There is often scatter in both magnitude and phase when using psd ratios to compute a describing function (DF) from actual pilot-in-loop data. This scatter reduces confidence in metrics obtained from the DF such as crossover frequency, phase margin, and gain margin. As an alternative to using the DF, an open-loop transfer function for the acceleration system in Figure 20 was created by substituting the identified pilot parameters (computed using a time window equal to the duration of the run) into Eq. 1 and using the known vehicle dynamics. This transfer function was plotted in Figure 26a as a dashed line. Since it reflects the aggregate characteristics for the entire run, the magnitude and phase obtained from the DF (using psd ratios) should roughly overlay on the transfer function, which they do. A bode plot of the two methods for the jerk command system is shown in Figure 26b.
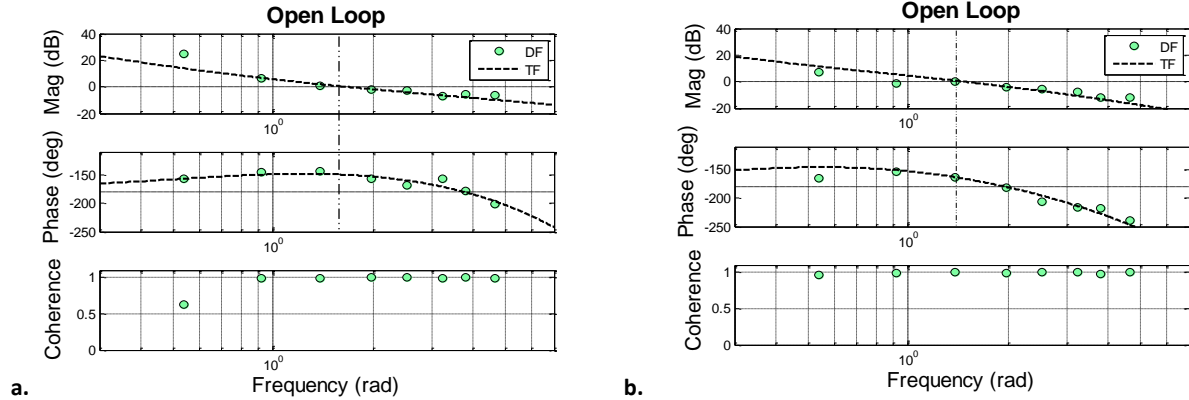
Figure 26. Open-loop bode plots generated by the describing function (DF) and transfer function (TF) for the two systems shown in Figure 20. *a) Acceleration command dynamics; b) Jerk command dynamics.*

Figure 27 shows the time histories of phase margin, gain margin, and crossover frequency for the acceleration system used in Figure 20. The identified time-varying pilot parameters and the known vehicle dynamics were used to compute the time traces of the metrics in Figure 27. Since stability margins and crossover frequency are based on effective open-loop gain, crossover used the pilot time delay, lead, and gain computed when not taking clipping into account.

Figure 27. History of frequency-domain metrics for acceleration example in Figure 20. *a) Phase margin; b) Gain margin; c) Crossover frequency.*

G. Cooper-Harper HQR Estimation

Figure 28 shows a potential method for estimating Cooper-Harper HQR using the Bedford estimator and pilot model. The pilot's pure time delay is esti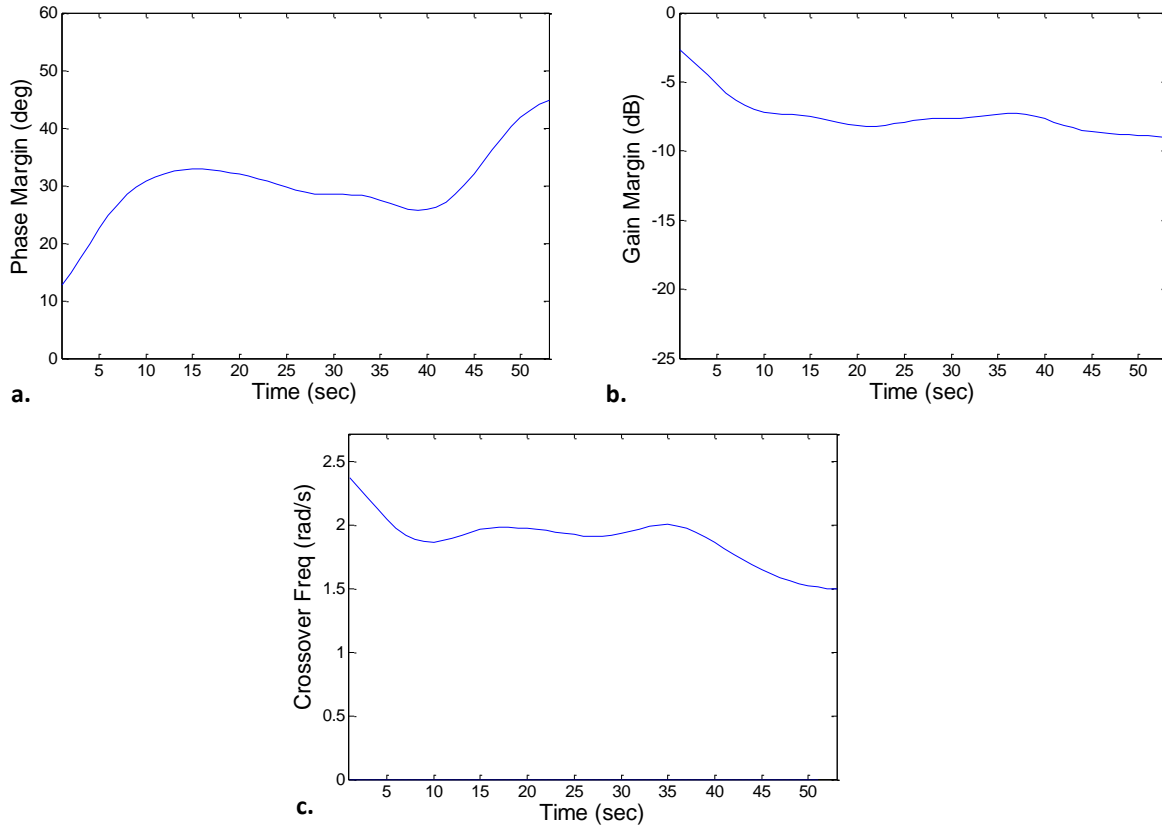mated using the assumed disturbance bandwidth, control feel system, and pilot lead compensation at crossover as recommended in Ref. 3. The pilot's technique (continuous tracking, amplitude clipping, pulsing) is specified, and simulations are run iterating on the pilot parameters. The performance of each run is checked for meeting the defined performance criteria of the task. If any runs satisfy the desired performance, the check for meeting adequate performance is not conducted. The minimum Bedford estimate associated with the error and control rates of the desired set is used as the HQR value (can be rounded for ordinal values). A Bedford estimate greater than 4 is capped at 4 when performance meets desired. If the desired performance check yields the null set, the check for adequate performance is conducted. Out of this set the minimum Bedford estimate is assigned as the HQR (values are limited to greater than or equal to 5 and less than or equal to 6). If the adequate performance check yields the null set, (which means the desired check produced the null set as well), all the runs are used to compute the minimum Bedford estimate, which is capped at 10 and not allowed to drop below 7.
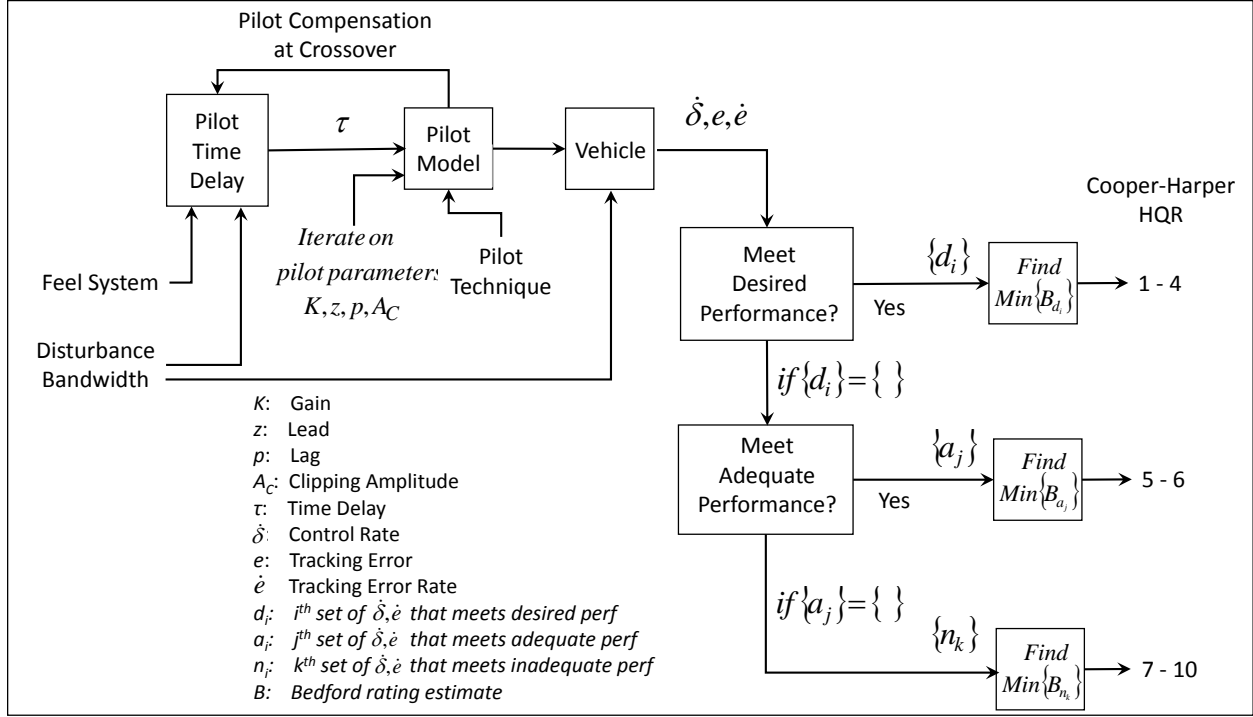
Pilot Compensation at Crossover

Pilot Time Delay $\xrightarrow{\tau}$ Pilot Model $\rightarrow$ Vehicle $\xrightarrow{\dot{\delta}, e, \dot{e}}$

Iterate on pilot parameters K, z, p, A_C

Pilot Technique

Feel System

Disturbance Bandwidth

Cooper-Harper HQR

Meet Desired Performance? — Yes $\{d_i\}$ → Find $Min\{B_{d_i}\}$ → 1 - 4

$if\{d_i\}=\{\ \}$

Meet Adequate Performance? — Yes $\{a_j\}$ → Find $Min\{B_{a_j}\}$ → 5 - 6

$if\{a_j\}=\{\ \}$

$\{n_k\}$ → Find $Min\{B_{n_k}\}$ → 7 - 10

K: Gain
z: Lead
p: Lag
A_C: Clipping Amplitude
τ: Time Delay
$\dot{\delta}$: Control Rate
e: Tracking Error
$\dot{e}$: Tracking Error Rate
d_i: i^{th} set of $\dot{\delta}, \dot{e}$ that meets desired perf
a_j: j^{th} set of $\dot{\delta}, \dot{e}$ that meets adequate perf
n_j: k^{th} set of $\dot{\delta}, \dot{e}$ that meets inadequate perf
B: Bedford rating estimate

Figure 28. Proposed methodology for estimating Cooper-Harper HQR.

III. Discussion

Based on the experimental results presented, it appears that pilots attempt to minimize the error they control while using a minimum of control exertion – a combination of tracking and control economy. Intuitively this makes sense: error rate is the stimulus to the controller, and stick rate reflects economy of the controller's output. The Bedford estimation algorithm proposed herein thus reflects 'efficiency of pilot's input relative to the physical output'. The interesting question is what process occurs for the pilot to maintain the observed relationship between stick rate and display error rate.

A key aspect of the experiment conducted was that there was no performance expectation for the pilots, they were simply told to track at whatever level of effort/error they felt comfortable with. Significantly, skill, learning effects, fatigue level, and pilot technique were not factors of importance, since the only focus was workload given a pilot's set of resources.

Amplitude-clipping produces a number of effects that are relevant to pilot performance and workload. First, it reduces the effective crossover frequency (i.e. bandwidth) of the open-loop system. While slightly degrading performance, moderate clipping can increase system stability, improving robustness to variations or increases in pilot time delay during divided-attention (such as multi-axis) or high-lead tasks. This in turn should produce an effective pilot gain that is less time-variant than if the pilot were to continuously track the error signal. The result should be more consistent performance with less effort, since the actual variations in pilot gain and time delay during active tracking will have less overall effect on performance (the pilot does not have to focus as much on gain and time delay constancy).

Pilot technique such as amplitude-clipping, pulsing, and continuous tracking can have a significant influence on workload and performance, and a pilot's ability to maintain satisfactory performance in the presence of sudden external and internal changes. This presents the potential for test, evaluation, and even fleet pilots to learn different control strategies and optimally match and apply them to different tasks and environments.

Compensation in terms of pilot lead is important, but it doesn't offer a comprehensive means for assessing workload. As the order of the vehicle dynamics increases it is generally more difficult to control, yet there can be subtle but important differences between what the pilot does with lead and what lead the task *requires*. It was seen that vehicle dynamics and task may combine such that the pilot is forced to generate lead compensation that would normally be associated with low workload – but in fact operates at high workload to avoid intolerable workload and/or performance that would arise if different lead was employed. It is significant that no satisfactory correlation

was observed between frequency-domain measures (i.e. crossover frequency, stability margin, pilot time delay, pilot compensation, etc.) and pilot spare capacity.

To compute overall workload for the inflight helicopter approach data, Bedford estimates for each channel of control were computed using a sliding window, and the highest estimate of all the channels at any time was used to form an aggregate Bedford estimate. This method yielded the best correlation with the actual pilot ratings, as opposed to an approach that considers all of the channels to contribute in some weighted fashion to the overall perceived workload. This implies that during multi-axis operation (where each axis may affect the other's performance), the Bedford estimate for any given axis reflects the combined effects of all axes on that axis - and the individual estimates do not additively (even partially) contribute to the overall workload perception. The channel with the highest workload at any instant is responsible for the overall workload at that instant.

To illustrate, Figure 29a shows the Bedford estimates of each control channel for a run taken from helicopter flight data shown earlier. Automatic heading control was enabled throughout all of the approaches, however, during some runs extremely gusty conditions required that the pilot actively control the pedals to maintain desired heading. At the 47-second mark in Figure 29a the pedal workload increases sharply. For most of the run collective has the highest Bedford estimate of the three channels, since the Combined estimate at the bottom closely repeats the collective profile. At 47 seconds the collective workload is seen to increase, and then decrease at 62 seconds synchronously with a decrease in pedal workload. While it might be argued that collective is driving pedal effort, the fact that the for this task pedal use is the exception rather than the norm lends to the assertion that pedal workload is influencing collective. Investigating this further, in Figure 29b the two variables comprising the RMS Bedford estimate, control rate and display error rate. At 47 seconds the pedal control rate jumps about 20%, whereas the collective rate increases less than 10%. The display error rate for pedals rises approximately twice the display error rate associated with collective. Thus it seems that the pedals are driving collective rather than the converse. While in this example pedal workload does not directly contribute to the overall Bedford estimate (it would if it had values that were higher than the other channels), the workload algorithm appears to capture an axis' influence on a channel that is the main driver of workload. This is significant in that it infers that overall workload perception and estimation is essentially an issue of rank.
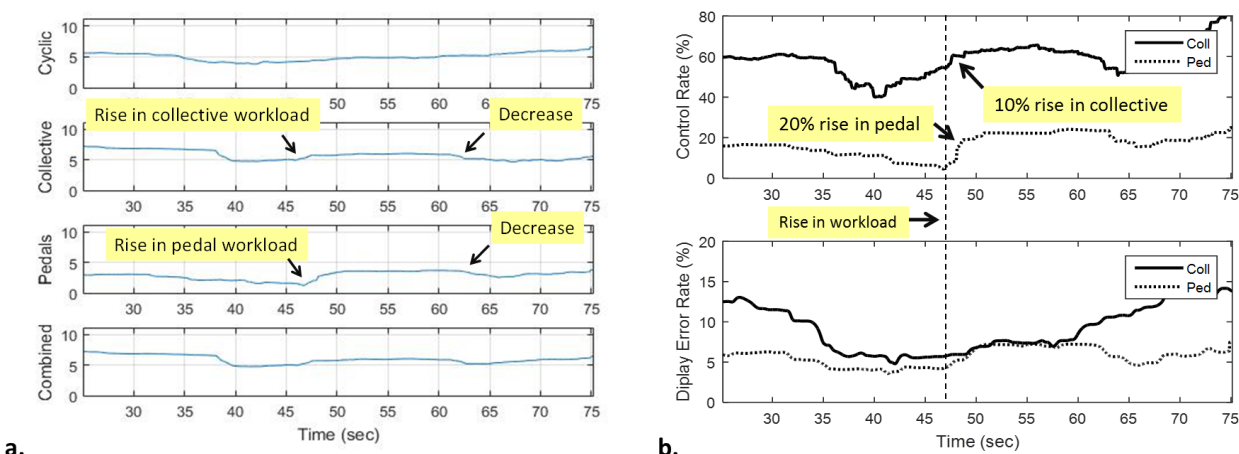

Figure 29. a) Bedford estimates for cyclic, collective, pedals, and aggregate; b) Normalized control and display rates for collective and pedals.

In 2000, Bachelder (Ref. 13) investigated the effect of translation cues on manual control. In this study a helicopter simulation experiment employed a number of unique techniques designed to create an environment where differences in performance between axes would be due primarily to visual perception and control strategy differences. In the experiment depth cues were artificially magnified so that the angular sensitivity of motion cues were equal in all three axes of translation (fore/aft, left/right, up/down). Each axis of vehicle translation was governed by the same dynamics, and the controls were all spring-centering. During single-axis control (where one axis was active and the other two disabled), stability margins and tracking performance were roughly the same for all three axes, but the longitudinal axis was given the highest workload rating (looming cues are more difficult to control than transverse cues). Multi-axis control yielded statistically significant interaction effects and differences between axes. By varying motion cue sensitivity in one axis and observing performance change in the others, it was

concluded that the depth axis receives more attentional resources than the others, i.e. the axis with the poorer handling qualities will draw a disproportionate amount of attention.

Referring again to Figure 29, a quick comparison will now be made between the UH-60's vertical control dynamics (collective-to-displayed sink rate) and heading control (pedal-to-displayed yaw) dynamics to establish which would exhibit the least favorable handling qualities during single-axis operation. Since the displayed sink rate for this task had automatic lead compensation, the dynamics would roughly appear as proportional to the pilot. However, the the dynamics associated with heading control would approximately appear as a second-order system - considerably more difficult to operate. Without examining the neuromuscular effects of the collective and rudder control systems, it thus appears that the pedal control would tend to affect collective, rather than the converse.

IV. Conclusion

A pilot model is proposed whose elements include the standard *CM* components (lead, time delay, and gain), perceptual noise, and a limiter that clips the CM output. A nonlinear pilot control technique, observed and coined by the authors as 'amplitude clipping', is shown to improve stability, performance, and reduce workload when employed with vehicle dynamics that require high lead compensation by the pilot. Combining linear and nonlinear methods a novel approach is used to measure the pilot control parameters when amplitude clipping is present, allowing precise measurement in real time of key pilot control parameters. It is hypothesized that it is easier for the pilot to clip amplitude (similar to bang-bang control) than to modulate the *CM* gain (without amplitude clipping) in response to changing internal and external variables.

Based on the results of an experiment designed to probe workload primary drivers, a method was developed that estimates pilot spare capacity (Bedford rating scale) from readily observable measures. From this experiment it appears that pilots attempt to minimize the error they control while using a minimum of control exertion − a combination of tracking and control economy. This relationship observed appears to be largely task-generic. To test reasonableness of the Bedford estimator it was applied to helicopter multi-axis flight. Given that the Bedford estimator was developed from a single-axis simulation task experiment, and that the DVE flight test conditions presented the pilots with multi-modal cueing (tactile, audio) and a mixed visual environment (head-down display and a distracting out-the-window scene of actual blowing dust), the degree of correspondence between the actual and estimated pilot ratings is encouraging.

From the inflight test results it appears that during multi-axis operation (where each axis may affect the other's performance), the Bedford estimate for any given axis reflects the combined effects of all axes on that axis - and the individual estimates do not additively (even partially) contribute to the overall workload perception. The channel with the highest workload at any instant is responsible for the overall workload at that instant.

Compensation in terms of pilot lead is important, but it doesn't offer a comprehensive means for assessing workload. As the order of the vehicle dynamics increases it is generally more difficult to control, yet there can be subtle but important differences between what the pilot does with lead and what lead the task *requires*.

Pilot technique such as amplitude-clipping, pulsing, and continuous tracking can have a significant influence on workload and performance, and a pilot's ability to maintain satisfactory performance in the presence of sudden external and internal changes. This presents the potential for test, evaluation, and even fleet pilots to learn different control strategies and optimally match and apply them to different tasks and environments.

The work documented in this paper represents the initial steps to developing a novel, simple objective metric for assessing pilot workload and its variation over time across a wide variety of tasks. Additionally, it offers a tangible, easily implementable methodology for anticipating a pilot's operating parameters and workload, and an effective design tool. It is shown that the model can predict very precisely the actual pilot settings and workload, and observed tolerance of pilot parameter variation over the course of operation. Finally, an approach is proposed for generating Cooper-Harper ratings based on the workload and parameter estimation methodology.

Future work will include applying the Bedford estimator to additional flight data and flight tasks, and expanding the repertoire of nonlinear pilot techniques that the pilot model can identify and employ.

American Institute of Aeronautics and Astronautics

Acknowledgments

References

[1] Cooper, G. E. and Robert P. Harper, J., \The Use of Pilot Rating in the Evaluation of Aircraft Handling Qualities," Tech. Rep. NASA TN D-5153, National Aeronautics and Space Administration, April 1969.

[2] Anon., Aeronautical Design Standard Performance Specification Handling Qualities Requirements for Military Aircraft, ADS-33E-PRF, 21 March 2000.

[3] McRuer, D. T. and Krendel, E. S., "Mathematical Models of Human Pilot Behavior," No. AGARDograph No. 188, November 1973.

[4] McDonnell, J., Pilot Rating Techniques for the Estimation and Evaluation of Handling Qualities, AFFDL-TR-68-76, May 1968.

[5] Hess, R., "Prediction of aircraft handling qualities using analytical models of the human pilot", NASA Technical Memorandum 84233, 1982.

[6] Hoh, R., "ACAH Augmentation as a Means to Alleviate Spatial Disorientation for Low Speed and Hover Helicopters", Heli-Japan – AHS International Meeting on Advanced Rotorcraft and Disaster Relief, 21 – 23 April, 1998, Gifu, Japan.

[7] Lampton, A., Klyde, D., et al, "Evaluation of a Precision Hover Task with Added Rotorcraft Dynamics Using Time-Varying Cutoff Frequency," presented at 67th Annual Forum of the American Helicopter Society, Virginia Beach, VA, May 3-5, 2011. (STI-P-762).

[8] Lampton, A., Klyde, D., "Power Frequency - A New Metric for Analyzing Pilot-in-the-Loop Flying Tasks", AIAA Atmospheric Flight Mechanics Conference, Guidance, Navigation, and Control, Vol. 35, September 2012.

[9] Roscoe, A. H., & Ellis, G. A. (1990). A subjective rating scale assessing pilot workload in flight. A decade of practical use. Royal Aerospace Establishment. Technical Report 90019. Farnborough. UK: Royal Aerospace Establishment.

[10] Draper, N. R.; Smith, H. (1998). Applied Regression Analysis. Wiley-Interscience. _ISBN_ _0-471-17082-8_.

[11] Graham, D., "McRuer, D., Analysis of Nonlinear Control Systems", New York, Wiley, 1961.

[12] Mitchell, D., Aponso, B.,"Reassessment and extensions of pilot ratings with new data", 17th Atmospheric Flight Mechanics, Portland,OR, 1990.

[13] Bachelder, E.: Perception-Based Synthetic Cueing for Night Vision Device Hover Operations, MIT Ph.D. Dissertation. (2000)