Proceedings of the EuroGNC 2013, 2nd CEAS Specialist Conference
on Guidance, Navigation & Control, Delft University of Technology,
Delft, The Netherlands, April 10-12, 2013

# A New Approach for the Validation of Potential Pilot Gain Measures

**Ina Niewind**

German Aerospace Center, Institute of Flight Systems, Flight Test Branch

DLR c/o WTD 61, Flugplatz, 85077 Manching, Germany

Tel. +49 8459-80-2595

Fax +49-8459-80-3085

Email: Ina.Niewind@dlr.de

**Abstract** The term “pilot gain” essentially describes the way the pilot acts on the inceptor during flight. It is a key aspect of handling qualities research and related flight tests. Most test organizations have their famous high- and low-gain pilots and the term “pilot gain” is understood very well on an intuitive level. In spite of its importance for handling qualities flight test, however, there is no generally accepted verbal or mathematical definition of “pilot gain”. This paper summarizes an approach for the validation of multiple potential pilot gain measures in the time and frequency domain based on pilot models and the associated results. The validation is based on data from a simulator study which was performed with 12 experimental test pilots and 12 operational pilots who varied their pilot gain / aggressiveness on command during a closed-loop tracking task. The approach is based on the idea that the validity of a potential pilot gain measure is based on its ability to reflect the pilot gain the pilots intended to apply during the tests and its ability to identify known outliers.

## 1 Pilot Gain

Handling qualities flight test considers the resulting dynamics of an aircraft controlled by a human pilot who performs a closed-loop task with tight control. In this context often the term Pilot-Vehicle System (PVS) is used. It describes a closed-loop system with negative feedback where the pilot acts as a control element. The pilot’s task is to minimize the control error between actual and target aircraft parameters, e.g. the desired and actual bank or pitch angle.

In the same way an electronic controller alters the overall characteristics of a closed-loop system, the pilot alters the PVS. Its characteristics are hence different from the pure aircraft dynamics which are evaluated in the frame of flying qualities testing. One of the most important parameters which influence the results of handling qualities flight tests – if not the most important parameter – is pilot gain. Pilot gain essentially describes the aggressiveness the pilot uses on the inceptor. Fig. 1 shows an example for the pilot's stick inputs with an intentional pilot gain variation from low to high pilot gain while the task remained the same. The difference in the stick inputs is readily apparent.

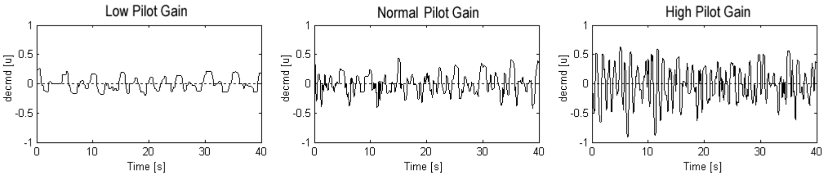

**Fig. 1** Time Histories of Low, Normal and High Pilot Gain

## *1.1 Pilot Model Gain*

Historically, the term "pilot gain" has its roots in control systems theory and is based on the idea that the pilot as part of the PVS can be represented by a pilot model which includes a gain factor. Common pilot models can range from the pure gain pilot up to highly complex control systems with numerous filters, spring-mass representations of limbs, time delays and switches. This inevitably leads to the question which model to chose when pilot model gain and pilot gain are used synonymously. But the pilot model is not the only factor. The pilot's control input can be primarily based on negative feedback, but may also include more or less significant portions of feedforward control. These portions are also part of the pilot's stick inputs and must be considered for pilot gain evaluations. In addition, common pilot models consider the pilot as a control element which intends to minimize a control deviation. Since the introduction of boundary avoidance tracking by Gray [7] a second type of control behavior based on the avoidance of unfavorable, often dangerous situations is also known. Conventional pilot models are not applicable for these situations. Finally, there is always a remnant in human control behavior which cannot be explained by conventional control systems theory. There are, hence, quite a few assumptions and limitations associated with the synonymous use of pilot gain and a gain factor of a pilot model.

Pilot models have been used by engineers in order to understand and explain how the pilot alters the PVS and why handling qualities differ from the aircraft's pure flying qualities. They have proven to be useful tools for aircraft manufacturers and handling qualities specialists during the design process and were the basis for the development of many common handling qualities criteria including the model-based Neal-Smith Criterion [11]. For the evaluation of pilot gain based on flight test data, however, the use of pilot models and the consideration of a pilot as an electronic controller in a negative feedback control loop may not be the preferred solution because of the large number of assumptions and limitations associated with this approach.

## *1.2 Pilot Gain in Literature*

Many reports in the field of handling qualities which explicitly mention pilot gain – often in conjunction with pilot-induced oscillations – do not provide a definition of pilot gain. It appears that this expression is considered to be self-explanatory. Most reports which are related to the development of handling qualities criteria make synonymous use of pilot gain and pilot model gain (e.g. in refs. [11], [6] and [10]). The MIL-HDBK-1797A [1] and the FAA Flight Test Guide [2] use bandwidth in conjunction with pilot gain. Duda [4] provides a relation between the PVS phase margin and pilot gain. Gray introduced pilot inceptor workload, a two-dimensional representation of pilot gain based on the root mean square stick speed and the percentage of time the pilot moves the stick. In ref. [3] the time spent by the human pilot in different loops of a pilot model (error rate control and error control) is associated with pilot gain. The approach uses a structural model [5] and the concept of entropy from information theory. Field and Giese use the term “pilot control activity” in order to describe pilot gain. They relate different frequency regions to high and very high gain closed-loop control. Evaluated measures for pilot control activity are the root mean square of the inceptor deflection, Power Spectral Density (PSD) plots and a calculation of the area below the PSD plot in predefined frequency regions, the number of control reversals and aircraft energy (the energy in the response of the aircraft that results from the control input) [5].

## *1.3 Aim of this Paper and Rationale*

“Numerical measurement (of pilot gain) has historically been something of a black art.” [8]. No attempt has ever been made to validate the assumption that any of the abovementioned measures reflect “pilot gain” in the way it is interpreted intuitively by experimental test pilots during flight test.

The aim of this paper is to present a process and the results of the validation of potential pilot gain measures based on the intuition and experience of trained pilots. Potential pilot gain measures were selected and introduced based on literature and the comments of experienced test pilots and air crew members from the flight test community which were gathered in the frame of a dedicated survey with pilots and air crew members from the flight test environment [12].

Since the stability of the PVS is strongly influenced by pilot gain, an evaluation of the different levels of pilot gain applied during flight test can support explanations for the pilots' different impressions of the aircraft handling qualities. This is especially useful when pilots from different sides (e.g. manufacturer, customer and airworthiness pilots) are involved.

The correct and universally accepted definition and representation of pilot gain in the frame of one-dimensional values also is a powerful tool to support the training of future test pilots at test pilot schools and test organizations. It gives pilots and engineers alike the confidence in the correct application of different levels of pilot gain (low – medium – high) as it is often used in handling qualities flight test, and provides objective feedback based on flight test data.

Finally, validated pilot gain measures allow a comparison of different handling qualities flight test techniques with respect to their potential to increase pilot gain. The author's starting point was the question whether Gray's workload buildup flight test technique [8] can increase pilot gain more efficiently than an intentional increase of pilot gain performed on the engineer's command.

## 2 Test Setup and Participants

In order to analyze the phenomenon of pilot gain, a dedicated test setup was defined in ref. [12] and refined in ref. [13]. The idea is to use a tracking task which is performed first with the pilot's normal control behavior. In a second step the pilot is asked to perform the same task intentionally in a low-gain (non-aggressive) manner. The third step is the same task again, but intentionally performed in a high-gain (aggressive) fashion. The set of these three test points is called "pilot gain calibration". A reasonable pilot gain measure should be able to reflect the pilot gain the pilots intended to apply during the test. This principle is the key aspect of the validation process which is introduced in this paper.

### *2.1 Test Setup*

**Familiarization and Test Points** The tests were preceded by a thorough familiarization phase which included a free phase with open and semi-closed inputs and closed-loop tracking and a mandatory tracking phase. The order of the succeeding three test points was fixed and always started with the pilot's natural gain in order to avoid an influence of a previous test point with intentional pilot gain variation on the pilot's natural control behavior. The second test point was always performed in a low-gain manner because it is easier for test pilots to switch from low to high pilot gain than vice versa. Once high pilot gain has been applied on purpose, it is hard for a pilot to immediately calm down and reduce the stick inputs to the lowest possible pilot gain. In addition, this order (first low, then high pilot gain) reflects real flight test where a build-up approach from low to high pilot gain is performed.

**Briefing Strategy** A careful briefing strategy was of utmost importance for the pilot gain calibration. As part of the test card, the pilots were asked whether they are familiar with the term “pilot gain” prior to the pilot gain calibration runs. If the answer was “no”, the synonymous term “aggressiveness” was used. By no means terms which could imply the validity of potential pilot gain parameters had to be used. If the pilots were told that high pilot gain or aggressiveness means to move the stick frequently, quickly, with large amplitude etc., this would have implied a specific interpretation of pilot gain and thus compromised the results. In addition to a neutral briefing, an obviously wrong application of low or high pilot gain by the pilots was not commented by the flight test engineer in order to avoid an influence of the flight test engineer’s a priori opinion about pilot gain on the results.

**Analyzed Data** Because some pilots commented that they could memorize a few seconds of the target movement at the beginning and in order to allow the pilot to fully get into the loop, the first five seconds of each run were omitted for data analysis. In addition, some pilots commented that the last target movement was also clearly recognizable to them; as a consequence, the last five seconds of each run were also omitted for data analysis in order to avoid differences in tracking behavior due to the anticipated end of the task. The remaining 40 s were used for the calculation of potential pilot gain measures. The duration of 40 s was a good compromise between a duration which is short enough to avoid effects resulting from fatigue and reduced concentration and a duration which is long enough to allow the calculation of the Power Spectral Density (PSD), Bode diagrams and parameter identification of pilot models. Data were recorded with 50 Hz.

**Simulator** Simulator tests were performed in a fixed-base simulator which consisted of a seat equipped with a throttle on the left hand side and a joystick plus armrest on the right hand side (Figure 2 left).

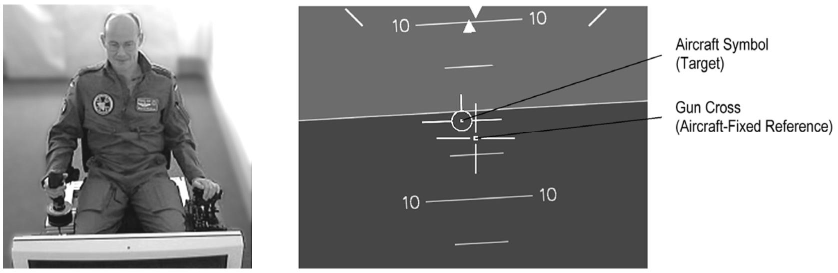

**Fig. 2** Fixed-Base Simulator (left) and Tracking Display (right)

Multiple ergonomic parameters of the simulator seat were adjustable within a wide range. Both throttle and joystick were from Thrustmaster (“HOTAS Warthog”). The joystick is sold as a replica of the A-10C stick, which is, however, used as a center stick whereas the simulator used the joystick as a modern side stick. The stick inputs were provided with a resolution of 16 bit. The throttle was not used in the frame of this study.

**Aircraft Dynamics** Appropriate aircraft dynamics were determined by Niewind and Opel in the frame of simulator tests with military test pilots from the German flight test center WTD 61 [15]. A linear model of the short period motion was used for the tests in order to keep the test setup as simple as possible and to avoid an additional influence on the test results caused by speed or altitude maintenance and induced roll movements. Four different agile and four different sluggish aircraft models were presented to the pilots who were asked to evaluate their suitability for several types of tracking tasks. The pilots could change the damping coefficient and the control power. The resulting aircraft dynamics were based on wind tunnel data of the Lockheed Jetstar aircraft [16]. Modifications were applied to control power and damping.

$$\frac{q(s)}{\eta(s)} = -1.05 \frac{2.1633s + 1.9379}{s^2 + 3.3789s + 3.1801} \tag{1}$$

**Target Movement** Figure 2 right shows the display used for the tests. Suitable target movements were evaluated by test pilots from WTD 61. A suitable movement was defined by a target movement which keeps the pilot constantly in the loop, is unpredictable, not too easy but also not excessively challenging and has the potential to get the pilot into boundary avoidance tracking events, i.e. moments when the pilot's input is solely based on the avoidance of a displayed boundary. The latter requirement was tied to a second aspect of the simulator study which is beyond the scope of this paper. Figure 3 shows an excerpt from the resulting tracking task which consists of 9 sines with a maximum frequency of 0.7 Hz.

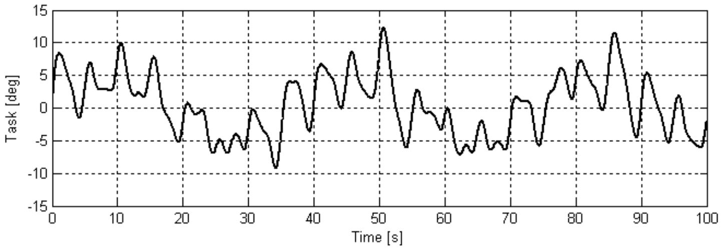

**Fig. 3** Sum of Sines Task [13]

## 2.2 Participants

Overall 24 pilots conducted the simulator study. 12 pilots were experimental test pilots from the German flight test center WTD 61 in Manching (8 pilots) and Cassidian (4 pilots). All participating test pilots were male and between 36 and 54 years old with an experience ranging from 1,550 to 6,700 flight hours. 11 out of 12 test pilots had a full course at the test pilot school (1 year duration); one attended a short course (4 months duration). 12 participants were operational fighter pilots from the 73rd Fighter Wing in Laage. All of them were male and between 25 and 35 years old with an experience ranging from 260 to 1,700 flight hours.

## 3 Potential Pilot Gain Measures

This section introduces potential pilot gain measures which are grouped in three different classes: time domain-based measures, frequency domain-based measures and pilot model-based measures. Measures in the time domain are more tangible while frequency domain-based measures are more abstract, but often include in-depth information about the system which is not readily apparent in the time domain. Measures based on pilot models require assumptions and simplifications, but they intend to reflect the pilot's control actions by means of control systems theory.

### *3.1 Time Domain-Based Measures*

One result of the pilot gain survey in ref. [12] is that pilots prefer time domain-based measures like stick deflection, stick speed, stick acceleration and the number of stick inputs. These parameters are experienced in real life and are therefore more tangible. Measures in the time domain do not require transformations, assumptions or parameter identification; they use readily available signals.

**Stick deflection, speed and acceleration** The stick deflection is directly determined from simulator signals. Stick speed and acceleration can be calculated based on the discrete derivative. The range of the stick deflection signal was set to [-1; 1]. The signal is unitless, but for clarity the the unit "u" (stick unit) is used for stick signals. For all three parameters and each test point, the mean value, the root mean square (rms), the ratio rms/mean, the maximum value and the percentage of large stick deflections and high stick speeds or accelerations are calculated.

Mean and rms value are two ways to express an average; the rms is more sensitive towards momentarily high values. The mean value is calculated as the 1-norm, i.e. the mean of the absolute value. Both represent the data of the complete test point.

The local maximum is a parameter which is only representative of a single instance of the test point. Especially for derivatives (speed and acceleration), noise can induce artifacts in the resulting maximum value.

The percentage of large/high values depends on a reasonable choice of limits. If the threshold is set too high, many pilots and runs achieve results of 0 % and the parameter has no potential to differentiate between medium, low and very low pilot gain. If the threshold is set too low, many pilots and runs achieve results of almost 100% and the parameter has no potential to differentiate between medium, high and very high pilot gain. In the frame of the tests, a threshold was chosen based on the first few completed tests which covered amongst others a well-known high- and low-gain pilot. The thresholds were set to 0.25 u stick deflection, ±1 u/s stick speed and ±30 u/s² stick acceleration.

**Number of Control Reversals** The number of control reversals represents the frequency of stick inputs as a tangible measure in the time domain. It depends on the duration of the test point and is thus normalized to the duration of one minute. A control reversal is defined as the change of direction of the stick movement. The change in direction has to exceed a noise threshold of 0.035 u in order to avoid erroneous detections.

**Duty Cycle** Like the number of control reversals, the duty cycle represents the frequency of stick inputs without using Fourier transforms. It is defined as "the percentage of time the pilot is changing his input on the stick" [8]. If the stick is held in its maximum position, this also contributes to the duty cycle because it is assumed that the pilot would move the stick even further if he could. The duty cycle is a part of Gray's pilot inceptor workload, a two-dimensional pilot gain representation. The disadvantage of the duty cycle is that it only provides information about the percentage of time the pilot moves the stick, meaning that the pilot could move the stick fairly slowly and still achieve a duty cycle of 1. The concept of pilot inceptor workload intends to cure this problem by adding information about the rms stick speed and combining the information in a two-dimensional plot [8].

**Tracking Error** The tracking error is defined as the difference between the target attitude and the current attitude of the aircraft. It is not expected to be a valid pilot gain measure, but was included because of the negative connotation associated with "high gain". Descriptive terms from the pilot gain survey in ref. [12] were "greedy", "agricultural", "rough" and "ham-fisted". In addition there was a strong association with the words "overcontrol", "overcorrection", "poorly coordinated", "mistimed" and "PIO". The preliminary study in ref. [12] gave an indication that many pilots achieve their best tracking performance when they use their natural pilot gain for tracking. This effect was independent of the pilots' natural gain, i.e. his individual preference to generally perform tasks with rather low, medium or high pilot gain.

## *3.2 Frequency Domain-Based Measures 1: Power Spectral Density*

The Power Spectral Density (PSD) provides information about the frequency content of a signal. PSD plots are often represented in a logarithmic form which allows the representation of a wide frequency and spectral density range. In the frame of this paper a linear representation is pursued because a logarithmic representation masks effects which are readily apparent with a linear scale.

Fig. 4 shows three PSD plots of the pilot gain calibration runs of a high-gain pilot. The grey area shows the range of frequencies of the sum of sines task (individual frequencies represented by dotted lines). The peaks in the PSD of the pilot's stick inputs (solid line) are sharp and defined in this area and coincide with task frequencies. The white area represents the frequency range above the Highest Task Frequency (HTF) up to 3 Hz. The PSD peaks in the area beyond the HTF are more scattered and distributed over a frequency range. They are often a result either of high pilot gain or flaws in the aircraft dynamics (or both). This can e.g. include oscillations resulting from overshoots which are aggressively corrected in both directions. When the aircraft dynamics are kept constant, pilot gain is the main contributing factor for differences in this area. This is also apparent in Figure 4.: the frequency content beyond the task frequencies is very low for the low-gain test point and increases with a pilot gain increase.

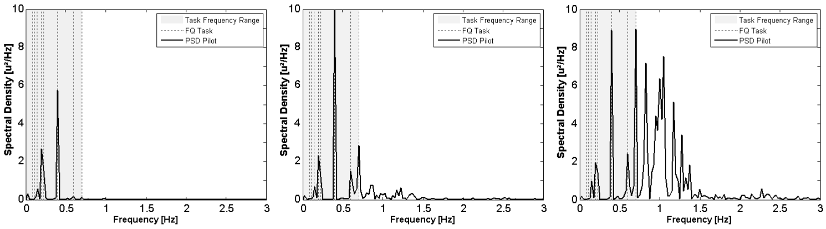

**Fig. 4:** PSD of a Typical High-Gain Pilot

**Signal Power and Signal Power Ratios** In ref. [12] it was shown that low- and high-gain pilots have different interpretations of pilot gain during an intentional pilot gain variation. Fig. 4 is representative of a high-gain pilot. Low-gain pilots tend to have a significantly lower frequency content in the range above the HTF even when high pilot gain is applied on purpose. The peaks in the task frequency range, however, increase with increasing pilot gain, while these peaks rather decrease or remain at the same level for high-gain pilots. A reasonable pilot gain measure must be able to identify pilot gain variations for low-, normal- and high-gain pilots alike. Because of the inherent differences in the frequency regions, the signal power of the following frequency regions is evaluated: 0 Hz-HTF, HTF-3 Hz and 0 Hz-3 Hz. Also, the ratios between the three signal powers are considered.

**Bandwidth** The bandwidth of a signal is defined as the frequency range with a signal power above 0. Because all measured signals include noise, a noise threshold of 0.2 $u^2$/Hz was used instead of 0. Another way to determine the bandwidth is by defining the frequency up to which 95% of the overall signal power are achieved. The suitability to represent pilot gain is evaluated for both variants.

### *3.3 Frequency Domain-Based Measures 2: Bode Diagram*

Bode diagrams show the frequency response of linear, time-invariant systems. In handling qualities research, Bode diagrams of the open-loop PVS and of the closed-loop PVS are commonly used. It is unlikely that a real flight test will be specifically tailored for Bode diagram generation if the sole purpose of the Bode diagram is the calculation of pilot gain. For similar reasons the target movement of the simulator study was not specifically tailored to the demands of Bode diagram generation. As a consequence, additional challenges apply to the use of related parameters. The solid black line in Fig. 5 shows a typical closed-loop PVS Bode diagram from the simulator study. The grey area marks the frequency region where the coherence is < 0.6 and thus too low for adequate results. Both, the magnitude and the amplitude plot, show a stepwise change of the graphs. To explain this behavior, a simulation with a pilot model consisting of gain, time delay and lead/lag compensation (see also Pilot Model 1 further below) was performed with parameters based on parameter identification of the pilot's run. The time delay was linearized based on a fourth-order Padé approximation. The results are shown in the dotted line in Fig. 5. They show the same stepwise change of magnitude and phase and a reasonable agreement with the pilot's data in the frequency range with adequate coherence. This result supports the hypothesis that the stepwise increase of magnitude and phase in the frequency response of the simulator study is a result of the test setup. It also provides confidence that the pilot model presents a reasonable representation of the data. In order to avoid erroneous results due to the stepwise change in magnitude and phase, the theoretic linear shape of the Bode diagrams is used for the determination of key parameters in the Bode plot. The theoretic data are shown in the grey line in Fig. 5.

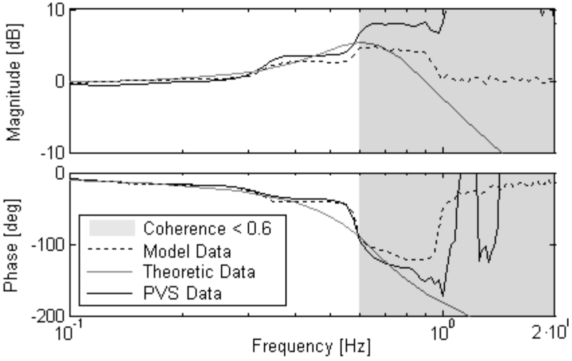

**Fig. 5** Example Bode Diagram (PVS, Model and Theoretic Data)

**Open-Loop Parameters** Open-loop parameters of the PVS which may be related to pilot gain are the crossover frequency $\omega_{cr}$ and the phase margin $\Phi_{mar}$ which is defined as the difference between the phase at the crossover frequency $\Phi_{cr}$ and -180 deg. (see also Fig. 6). In ref. [4] Duda states a typical value of $\Phi_{mar} = 45$ deg and accounts for a variation of ± 25 deg. A lower phase margin is associated with high-gain pilots who are known to destabilize the PVS.

**Closed-Loop Parameters** Closed-loop parameters of the PVS which may represent pilot gain are the resonance amplitude $A_{res}$ which is closely tied to the damping of the closed-loop system and the closed-loop bandwidth $\omega_{bw}$ which is defined as the frequency after the resonance peak when the amplitude drops below -3 dB as a measure of the range of frequencies which can pass through the system without significant attenuation (see also Figure 7).

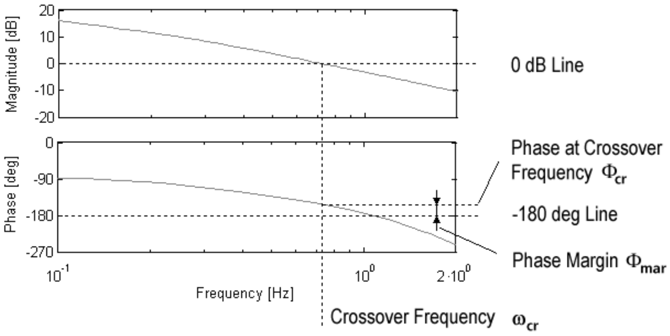

**Fig. 6** Characteristic Parameters of the Open-Loop Bode Diagram

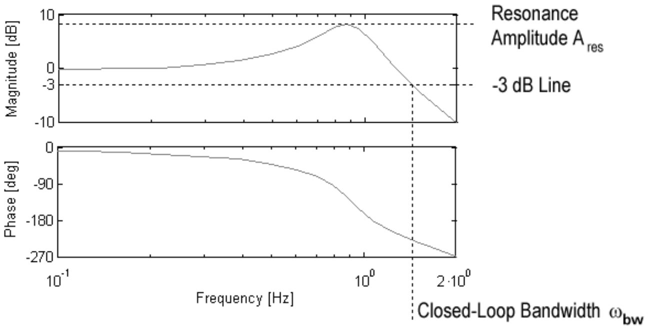

**Fig. 7:** Characteristic Parameters of the Closed-Loop Bode Diagram

### *3.4 Pilot Model-Based Measures*

Historically, the term "pilot gain" has its roots in control systems theory and is based on the idea that the pilot as part of the PVS can be represented by a pilot model. This paper uses two different pilot models.

$$\text{Pilot Model 1}: F_1(s) = K_1 e^{-T_{e1}s} \frac{T_D s + 1}{T_I s + 1} \tag{2}$$

$$\text{Pilot Model 2}: F_2(s) = K_2 e^{-T_{e2}s} \tag{3}$$

Pilot Model 1 is identical with the model used in the Neal-Smith criterion [11]. Instead of using the assumptions of the Neal-Smith criterion, parameter identification was performed with the DLR tool Fitlab [17]. Pilot Model 2 is a fairly simple model consisting of a pilot model gain and a time delay. Like for Pilot Model 1, parameter identification was used in order to derive the parameters of Pilot Model 2.

In general, the use of pilot models requires assumptions about the suitability of the pilot model and it requires parameter identification – two steps which can lead to a much higher difference between the potential pilot gain measure and the pilot's real control behavior than those measures which use parameters that can be measured directly or only require simple calculations, e.g. measures in the time domain.

**Pilot Model Gain** Quite often in literature, pilot gain is used synonymously with pilot model gain, the gain factor K of a pilot model. Because of this, the pilot model gain is evaluated as a potential pilot gain measure. The two pilot models were used to demonstrate a problem of the synonymous use of pilot gain with pilot model gain: Pilot Model 1 implicitly includes the simpler Pilot Model 2; they are identical for $T_D = T_I$. Pilot Model 1 thus creates a better model fit because it accounts for lead and lag compensation applied by the pilot. However, when the pilot model gain K is used as a representation of pilot gain, different parameter sets for lead and lag compensation make no difference to the resulting pilot model gain, but they can significantly change the pilot's stick inputs. The simpler Pilot Model 2 is thus expected to be a better representation of pilot gain even though its model fit is worse.

**Model Fit** A simple measure of the model fit was also calculated based on the rms value of the deviation between the pilot's and the pilot model's stick input. Because the deviation is based on the baseline stick deflection, it is normalized by the rms stick deflection.

**Other Components** Lead and lag time constants for Pilot Model 1 and the time delay for both models were also evaluated with no specific expectation about a relation with pilot gain.

## 4 Validation of Potential Pilot Gain Measures

### *4.1 Validity Plots*

Based on the experience that pilot gain is understood quite well on an intuitive level, the pilot gain calibration runs are used for the validation of potential pilot gain measures. The idea is that a valid pilot gain measure should be able to reflect the pilot gain the pilot intended to apply. The test point which was intentionally performed in a low-gain manner should result in a lower value than the normal-gain test point which should result in a lower value than the high-gain test point. For monotonically decreasing pilot gain measures the opposite applies. In any case, hence, a valid pilot gain measure must be monotonically increasing or decreasing with the applied pilot gain and the result must be consistent for all or at least most pilots. Figure 4 shows an example for valid and invalid measures.

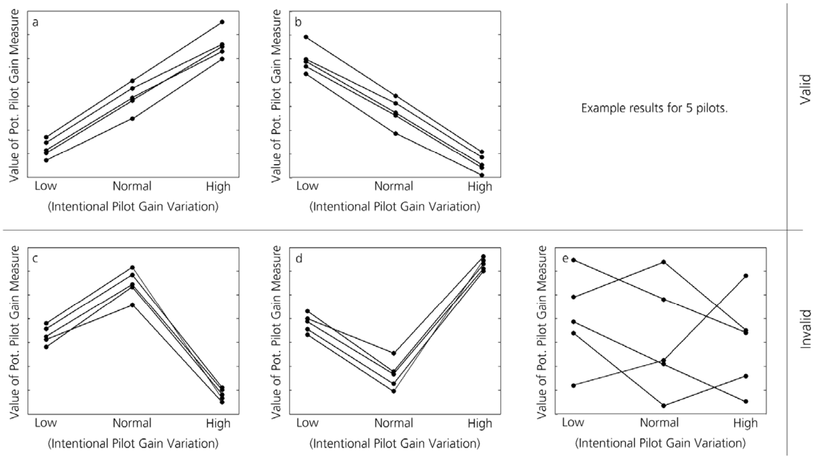

**Fig. 8** Validity Plots: Examples for Valid and Invalid Measures

## *4.2 Validity Index*

While validity plots provide an effective and tangible tool to visualize the validity of potential pilot gain measures, a numerical validity index allows a quantification of the validity of the measures under evaluation.

A suitable validity index has to be able to distinguish between valid and invalid potential pilot gain measures and preferably also between monotonically increasing and decreasing valid pilot gain measures. As a baseline, a numerical concept similar to the regression coefficient was the target for the validity index:

- Validity Index (VI) = -1: monotonically decreasing valid pilot gain measure.
- Validity Index (VI) = 1: monotonically increasing valid pilot gain measure.
- Validity Index (VI) = 0: invalid measure with no relation to pilot gain.

To achieve the abovementioned conditions, the validity index VI is calculated as follows:

$$\mathrm{VI} = \frac{\sum_{i=1}^{n} P_i}{n} - 1 \qquad (4)$$

with $P$: test statistic for each pilot (value range: 0 - 2; unitless parameter)
$n$: number of pilots who successfully performed the test
$i$: index (value range: $1 - n$)

The test statistic for each pilot is calculated as follows:

$$P = P_{LG} + P_{HG} \tag{5}$$

with

$$P_{LG} = \begin{cases} 1, LG < NG \\ 0, \text{otherwise} \end{cases} \tag{6}$$

$$P_{HG} = \begin{cases} 1, NG < HG \\ 0, \text{otherwise} \end{cases} \tag{7}$$

and $LG$ = low gain value, $NG$ = normal gain value, $HG$ = high gain value of the potential pilot gain parameter (intentionally applied pilot gain during the pilot gain calibration runs).

## *4.3 Validation Based on Outlier Detection*

During the conduction of the tracking task, three pilots created significant outliers. All outliers were already apparent during the task conduction. This first impression could be supported by PSD plots and time histories [14].

Pilot A performed his high-gain test point in an unusually aggressive manner which clearly exceeded the limits of what can be considered reasonable tracking behavior. This test point is probably a simulator artifact since it is unlikely that the pilot would have shown the same tracking behavior in flight. Pilot B evidently used higher pilot gain for his low-gain run than for the normal-gain run. A similar problem occurred during the pre-study with another pilot. Back then the pilot commented that he allowed a larger error during the low-gain task. As a consequence, he often reached a point when the tracking error was so large that he felt obliged to apply a quick and large input. This input then ended up being more aggressive than what he had applied during the normal-gain task [12]. The same probably happened to Pilot B. Pilot C did not succeed in varying his pilot gain. As he commented after the complete test, he would recommend to repeat the pilot gain calibration runs again at the end of the complete test battery since he would then have performed it differently.

Even though the existence of outliers is in general unfavorable, they can serve a practical purpose. A good pilot gain measure should be able to identify the outliers.

- Pilot A is correctly identified when the high-gain data point creates a significantly higher value than the high-gain data points of the other pilots.
- Pilot B is correctly identified when he is the only pilot who has his lowest pilot gain value for the normal-gain data point.
- Pilot C is correctly identified when there is no significant difference between the values of all three data points from the pilot gain calibration runs.

All three outliers are omitted for the calculation of the validity index in order to avoid misleading differences in the values of the validity index for different measures based on the outliers. Especially the very small differences between the different runs of Pilot C can lead to differences in the results which do not reflect higher or lower validity of the measure under evaluation.

## 5 Results from the Simulator Study

### *5.1 Validity Plots*

Figure 9 shows four different types of results for potential pilot gain measures from the simulator tests. The plots are based on the data of all 24 pilots; outliers are shown with black lines and squares, valid data in grey and with circles.

The validity plot of the rms stick speed in Figure 9 left above represents a valid pilot gain measure which is monotonically increasing with pilot gain. The validity index is 1.00 and all three outliers are correctly identified. The validity plot of the signal power in the region above the highest task frequency of 0.7 Hz (HTF - 3 Hz) in Figure 9 right above represents another valid pilot gain measure. While the increase in the rms stick speed is approximately linear with intentionally applied pilot gain, this measure is nonlinear with a steeper increase between normal and high gain than between low and normal gain. The validity index is 1.00 and all three outliers are correctly identified. The validity plot of rms/mean stick acceleration ratio in Figure 9 left below represents a potential pilot gain measure which is monotonically decreasing with pilot gain. The validity index is -0.95, but none of the three outliers is correctly identified. The validity plot of the rms tracking error in Figure 9 right below represents an invalid pilot gain measure. Overall, there was a V-shaped trend for 12 pilots, an A-shaped trend for 1 pilot and a monotonically decreasing trend for 8 pilots. The resulting validity index is -0.38, indicating a slight tendency for increased tracking performance with increasing pilot gain for some pilots. Based on the poor validity index, outlier detection is not possible.

The preliminary result in ref. [12] which indicated that most pilots have their best tracking performance when using their natural pilot gain is not confirmed by these results.

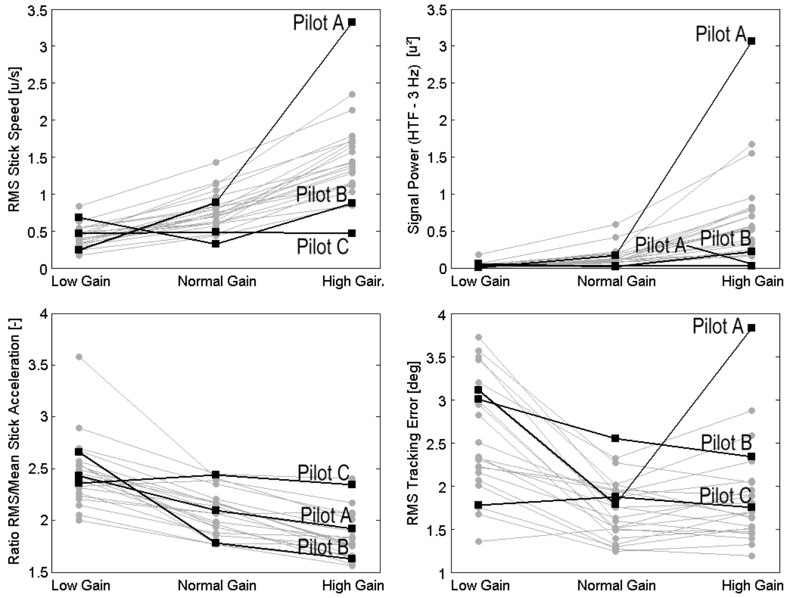

**Fig. 9** Validity Plots from the Simulator Study

## *5.2 Validity Index*

11 out of 42 potential pilot gain measures achieved a validity index of 1 or -1:

- Time Domain
  - stick speed: mean, rms and percentage of high speeds
  - stick acceleration: mean, rms and percentage of high accelerations
  - number of control reversals per minute

- Frequency Domain 1 (PSD)
  - signal power in the region between HTF and 3 Hz
  - signal power ratio (0 Hz – HTF)/(0 Hz – 3 Hz), monotonically decreasing
  - signal power ratio (HTF – 3 Hz)/(0 Hz – 3 Hz)
  - signal power ratio (HTF – 3 Hz)/(0 Hz – HTF)

Parameters based on the Bode diagram and based on pilot models did not achieve a validity index of 1.

10 measures achieved a validity index of -0.9 or 0.9 or better (but not 1):

- Time Domain
  - stick deflection: maximum, mean, rms, percentage of large deflections
  - stick accelerations: ratio rms/mean, monotonically decreasing
  - duty cycle
- Frequency Domain 1 (PSD)
  - signal power 0 Hz – 3 Hz
  - bandwidth (> noise threshold, 0.2 $u^2$/Hz)
  - bandwidth (95% of signal power)
- Frequency Domain 2 (Bode)

  none
- Pilot Models
  - pilot model 2: pilot model gain K

The parameters based on the Bode diagram did not achieve significant validity indices. This may be a result of the fact that a large number of calculations has to be performed in order to achieve the resulting values. In addition, because of the influence of the sum of sines task frequencies on the Bode diagram, the data of a pilot model were used instead of the pilot's original data. Furthermore, the coherence of the signal in the higher frequency ranges which are most interesting for pilot gain evaluations was poor. All these effects are expected to be present in regular flight tests which are not specifically tailored for frequency response generation. Because a number of other pilot gain measures with a perfect validity index is available, changes in the sum of sines task which aim at better frequency response generation but may at the same time limit the operational representativeness of the task are not recommended. The best validity index achieved by a Bode diagram-based parameter was 0.86 for the closed-loop bandwidth. The crossover frequency and the phase margin achieved validity indices of 0.76 and 0.71, respectively.

For the pilot model-based parameters, only the pilot model gain of pilot model 2 (gain and time delay) achieved a reasonable validity index of 0.95. The pilot model gain of pilot model 1 only achieved a validity index of 0.86.

## ***5.3 Outlier Detection***

Four examples for outlier detections were already discussed in the preceding subchapter. Out of the 11 measures with a validity index of 1 or -1, 7 clearly show the outliers. The remaining 4 measures are the PSD ratios and the number of control reversals per minute. They identified 2 of 3 outliers and it was always Pilot A they failed to identify. While this is not a strong reason to exclude these measures, the remaining measures which were able to identify all outliers should be preferred.

### *5.4 Relation between Potential Pilot Gain Measures*

It is obvious that many of the investigated measures are closely related. The most obvious example is given by the rms and mean value of the stick deflection, speed or acceleration - both are just two different ways to express an average over the test point. The stick speed is also calculated based on the stick deflection; however, theoretically a mean stick deflection of 1 u (maximum stick position throughout the test point) and a stick speed of 0 u/s (no movement) are possible. These measures are tied together by the sum of sines task. The same applies not only for closely related parameters, but also parameters from different classes (e.g. time and frequency domain). Figure 10 shows a scatter plot for two validated pilot gain measures, the rms stick speed vs. the signal power in the frequency range above the task frequencies (HTF – 3 Hz). The close relation is readily apparent. Similarly close relations exist for all measures with a validity index of 1 which were able to identifiy all three outliers. The signal power ratios showed a close relation with the rms stick speed for all data points but the high-gain data point of Pilot A, which they previously failed to identified as an outlier. The weakest relation exists between the rms stick speed and the number of control reversals with a nonlinear coefficient of determination $R^2$ of 0.78.

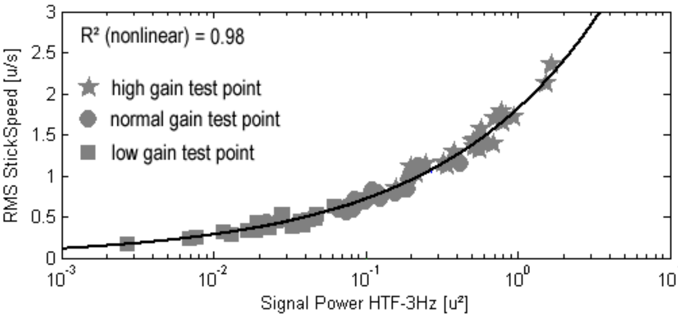

**Fig. 10** Scatter Plot for the RMS Stick Speed vs. Signal Power (HTF – 3 Hz)

## 6 Summary and Conclusions

This paper introduces a new approach for the validation of potential pilot gain measures and summarizes the results of various different measures in the time domain, the frequency domain and based on pilot models. The validation is based on data from a simulator study which was performed with 12 test pilots and 12 operational pilots who varied their pilot gain on command during a closed-loop sum of sines tracking task. The validity of a potential pilot gain measure is based on its ability to reflect the pilot gain the pilot intended to apply during the test quantified by a validity index. In addition, three significant outliers were identified during the tests which had to be identified by valid pilot gain measures.

In the end, a large number of the time domain-based and PSD-based measures achieved a reasonable validity index and 7 measures achieved a perfect result, i.e. they achieved a validity index of 1 (or -1) and allowed a clear detection of all three outliers. Measures like the rms or mean stick speed are easy to calculate and do not require assumptions or special test setups. PSD-based measures are more abstract, but PSD plots allow an in-depth evaluation of the pilot's interpretation of pilot gain which can be a valuable addition to a one-dimensional measure.

All validated parameters are physically related and additionally tied together by the sum of sines task. The results are thus limited to sum of sines tasks and have to be confirmed for other types of tracking tasks (e.g. step and ramp tasks). The recommended pilot gain measures are:

- stick speed: mean, rms and percentage of high stick speeds
- stick acceleration: mean, rms and percentage of high stick accelerations
- signal power above the highest task frequency (HTF – 3 Hz)

The final choice of a pilot gain measure should be based on the purpose of its use. When working with operational pilots, tangible measures like the mean stick speed should be preferred. Because the percentage of high stick speeds and accelerations depends on a choice for a threshold for "high" values based on engineering judgement, these measures are inferior to the other ones which do not require similar choices. The stick speed should also be preferred over the stick acceleration as the latter usually has more artifacts in the data stream due to the double differentiation. On an engineering level, frequency domain-based measures may be more suitable than time domain-based measures. The two-dimensional representation of the PSD is a good solution to provide deeper information about pilot gain if no single value is required.

In the end, mathematically very simple parameters proved to be suitable representations of pilot gain while more complex parameters provided less suitable results. One of the reasons for this may be the underlying assumptions which are required when e.g. pilot models are used. In addition, frequency response generation which is necessary for Bode diagrams requires tasks which are specifically tailored for this purpose, while in flight test this additional burden may not be accepted if its sole purpose is pilot gain determination.

## Acknowledgments

The author would like to thank all participating pilots from WTD 61, Cassidian and the 73rd Fighter Wing in Laage for their support of this study.

## References

[1] Anon: Flying Qualities of Piloted Aircraft. MIL-HDBK-1797A, Wright-Patterson AFB (1997)
[2] Anon: FAA Flight Test Guide, FAA AC No. 25-7A (1998)
[3] Celere, A.D., Varoto, P.S., Maciel, C.D.: Verifying Pilot Gain in PIO Flight Test. AIAA Atmospheric Flight Mechanics Conference and Exhibit, Hilton Head, SC, AIAA 2007-6640 (2007)
[4] Duda, H.: Fliegbarkeitskriterien bei begrenzter Stellgeschwindigkeit, DLR Forschungsbericht 97-15, ISBN 0939-2963 (1997)
[5] Field, E.J., Giese, S.E.D.: Appraisal of Several Pilot Control Activity Measures. AIAA Atmospheric Flight Mechanics Conference and Exhibit, San Francisco, CA, AIAA 2005-6032 (2005)
[6] Gibson, J.C.: Development of a Methodology for Excellence in Handling Qualities Design for Fly By Wire Aircraft. Delft University Press, Series 03: Control and Simulation 06 (1999)
[7] Gray, W.R.: Boundary-Escape Tracking: A New Conception of Hazardous PIO. Proceedings of the 48th Symposium of the Society of Experimental Test Pilots, SETP, Lancaster, CA (2004)
[8] Gray, W.: A Boundary Avoidance Tracking Flight Test Technique for Performance and Workload Assessment. 51st Symposium of the Society of Experimental Test Pilots, San Diego, CA (2007)
[9] Hess, R.A.: Structural Model of the Adaptive Human Pilot. Journal of Guidance, Control, and Dynamics, Vol. 3 No. 5, pp. 416–423 (1980)
[10] McRuer, D.: Pilot Induced Oscillations and Human Dynamic Behavior, NASA Contractor Report 4683 (1995)
[11] Neal, T.P., Smith R.E.: A Flying Qualities Criterion for the Design of Fighter Flight Control Systems. Journal of Aircraft, Volume 8, Number 10 (1971)
[12] Niewind. I.: What the Hell is Pilot Gain? SFTE Symposium, Seattle (2011)
[13] Niewind, I.: Pilot Gain and the WLB FTT: Test Conduction and Evaluation Strategy. Institute Report at the German Aerospace Center, IB 111-2012/51 (2012)
[14] Niewind, I.: Pilot Gain and the Workload Buildup Flight Test Technique: Validity of Potential Pilot Gain Measures, Institute Report at the German Aerospace Center, IB 111-2012/61 (2012)
[15] Opel, F.: Investigations on the Workload-Buildup Flight Test Technique: Dynamics, Displays and Task Variations. Master Thesis, German Aerospace Center, Institute of Flight Systems, IB 111-2012/40 (2012)
[16] Robert K.H.: Aircraft Handling Qualities Data. Washington, D.C.: NASA (1972)
[17] Seher-Weiß, S.: User's Guide: FITLAB - Parameter Estimation Using MATLAB - Version 2.0. Institute Report at the German Aerospace Center, IB 111-2007/27 (2007)