Just Noticeable Differences for Variations in Quasi-Steady Stall Buffet Model Parameters

Imbrechts, A.; de Visser, C.C.; Pool, D.M.

DOI
10.2514/6.2022-0510

Publication date
2022

Document Version
Final published version

Published in
AIAA SCITECH 2022 Forum

AIAA SciTech Forum
January 3-7, 2022, San Diego, CA & Virtual
AIAA SCITECH 2022 Forum

10.2514/6.2022-0510

Just Noticeable Differences for Variations in Quasi-Steady Stall Buffet Model Parameters

A. Imbrechts*, C.C. de Visser† and D.M. Pool‡
Delft University of Technology, Delft, The Netherlands

To gain more insight into human sensitivity to variations in simulated stall buffets, Just Noticeable Difference (JND) thresholds were estimated using a passive human-in-the-loop flight simulator experiment. Using an in-house developed flow separation-based stall and buffet model of the Cessna Citation II, JND thresholds were determined for the model's buffet characteristic frequency parameter ω_0 and the buffet onset threshold parameter X_{thres} for the vertical stall buffet only. With a subjective yes/no 1-up/1-down staircase procedure that uses repeated pairwise comparisons of quasi-steady symmetric stall simulations (where one is a stall with the baseline buffet model and the other one has an offset buffet parameter), upper and lower JND thresholds were measured from 21 pilots. The experiment results show that the pilots noticed the differences in simulated buffet dynamics at comparably similar percentage-wise offsets for X_{thres} and ω_0 with respect to the baseline parameter values. The maximum observed JND thresholds did not exceed 30-35% across all experiment conditions, indicating that pilots are fairly sensitive to even small offsets in the key stall buffet model parameters. Moreover, the estimated JND thresholds for ω_0 are in agreement with the ±2 Hz tolerance currently used in stall buffet simulation qualification standards. However, for X_{thres}, the results show that human pilots already notice differences in stall buffet onset characteristics well before the maximum allowed tolerance (±2.0° angle of attack) is reached, which suggests that stricter tolerances on simulated buffet onsets for quasi-steady symmetric stalls may help to further enhance stall training in simulators.

Nomenclature

Roman Symbols

a_1	Stall abruptness parameter, -
a_y	Lateral stall buffet acceleration, m/s^2
a_z	Vertical stall buffet acceleration, m/s^2
C_L	Lift coefficient, -
H_0	Buffet shaping filter gain, -
$K_{\{q,x,z\}}$	Motion filter gain in pitch, surge and heave, -
Q_0	Buffet shaping filter quality factor, -
X	Flow separation point, -
X_{thres}	Buffet onset threshold on X, -

Greek Symbols

α	Angle of attack, rad or deg
α^*	Angle of attack for which $X = 0.5$, rad
$\zeta_{\{q,x,z\}}$	Motion filter damping coefficient in pitch, surge and heave, -
τ_1	Lag time constant of flow separation point X, s
τ_2	Hysteresis time constant of flow separation point X, s
ω_0	Buffet shaping filter natural frequency, rad/s
$\omega_{b_{\{q,x,z\}}}$	Motion filter break frequency in pitch, surge and heave, rad/s
$\omega_{n_{\{q,x,z\}}}$	Motion filter natural frequency in pitch, surge and heave, rad/s

*M.Sc. student, Control and Simulation Division, Faculty of Aerospace Engineering, P.O. Box 5058, 2600GB Delft, The Netherlands; a.imbrechts@student.tudelft.nl.
†Assistant Professor, Control and Simulation Division, Faculty of Aerospace Engineering, P.O. Box 5058, 2600GB Delft, The Netherlands; c.c.devisser@tudelft.nl. Member AIAA.
‡Assistant Professor, Control and Simulation Division, Faculty of Aerospace Engineering, P.O. Box 5058, 2600GB Delft, The Netherlands; d.m.pool@tudelft.nl. Senior Member AIAA.

I. Introduction

Aerodynamic stall is an intense dynamic, nonlinear and unsteady condition that may lead to unrecoverable airplane upset conditions if not corrected in time. Stalls are an important contributor to fatal accidents in civil aviation and are the primary cause of fatal accidents in general aviation [1–3]. Until recently, training simulators were not required to provide high-accuracy stall simulation [4–6]. However, this has changed with the mandatory requirement for all airline crew to receive flight simulator-based stall prevention and recovery training that is effective since 2019 [7–9]. As a result, there is a strong need for accurate and cost-effective stall and post-stall dynamic models for use in flight simulators.

A key characteristic of a stall is the *stall buffet*, as buffeting is an initial cue for pilots that indicates entering of the unsafe part of the flight envelope. The stall buffet, which occurs at high angles of attack, is the aerodynamic excitation due to flow separation causing pressure fluctuations over the wing [10–12]. A common remaining deficiency in current Flight Simulating Training Devices (FSTDs) is the insufficient haptic and physical vibratory feedback of buffeting felt by pilots in simulated stalled conditions [6, 13]. While this can partly be attributed to practical considerations, as limiting buffeting vibrations positively benefits the required FSTD maintenance and downtime, a major second reason is that it is, in fact, unknown what level of stall buffet accuracy or fidelity is required for realistic stall simulations and effective stall training. Available regulatory standards for stall buffet simulation [7] reflect this persisting uncertainty with seemingly lenient tolerances on buffet responses, e.g., "*the flight simulator results should exhibit the overall appearance and trends of the airplane plots, with at least some of the frequency 'spikes' being present within 2 Hz of the airplane data*" [7].

Therefore, the goal of this paper is to provide additional quantitative guidance on the required accuracy for replicating stall buffets in flight simulators. This paper presents the results of a human-in-the-loop experiment in the SIMONA Research Simulator at TU Delft that was performed to measure Just Noticeable Difference (JND) thresholds for key parameters that characterize the frequency content and buffet onset threshold of simulated stall buffet vibrations. The experiment was performed by 6 type-licensed Citation II pilots and 15 other (commercial, private, or glider) pilots. We utilized a Cessna Citation II stall model identified in our earlier research [14, 15]. For measuring the JND thresholds, use was made of the same experimental paradigm as described in Ref. 16, where participating pilots experienced simulated symmetrical quasi-steady stall maneuvers at an altitude of 5,500 m (18,000 ft), induced with a 1 kt/s deceleration into the stall, as an observer. Through a subjective staircase procedure, consisting of repeated pairwise comparisons of the stall with our baseline buffet model and a stall with adjustments to the buffet model parameters, the JND thresholds for these parameter variations were determined. Furthermore, the measured JND thresholds were objectively compared to tolerances that apply to simulated stall buffet characteristics from current regulatory standards [7, 17].

This paper is structured as follows. Section II provides a background on stall buffet model requirements for FSTDs. Furthermore, the Citation II buffet model used in this research is presented in the same section, as well as an offline sensitivity analysis of the buffet's responses to changes in the model parameters. Section III describes the human-in-the-loop experiment and methods used to estimate the JND thresholds of key stall buffet model parameters. The results of the experiment are presented in Section IV, which is followed by a discussion and conclusions towards the end of the paper.

II. Stall Buffet Modeling

A. FAA Requirements for Stall Buffet Models in FSTDs

In an effort to reduce the occurrences of loss of control in-flight related incidents, the Federal Aviation Administration (FAA) introduced the binding requirement for airline flight crew to receive simulator-based stall prevention and recovery training. This mandatory stall training is effective since 2019 and reflects the need for accurate stall simulation models for flight simulators [9, 17, 18]. A key characteristic of stalled flight is the *stall buffet*. The stall buffet serves as an initial cue to pilots upon entering of the unsafe part of the flight envelope. A remaining deficiency in FSTDs is the insufficient modeling of such buffet responses, as it is largely uncertain what level of stall buffet model fidelity is required for accurate stall simulations [6, 13]. Current regulatory standards describe the tolerances on simulated buffet responses as:

- "*Buffet onset threshold of perception should be based on 0.03g peak to peak normal acceleration above the background noise with a tolerance of* ***±2.0° angle of attack***" [7, p. 95].
- "*Correct trend of growth of buffet amplitude from initial buffet to stall speed for normal and lateral acceleration will have to be demonstrated*" [7, p. 95].
- "*FSTD manufacturers may limit maximum buffeting based on motion platform capability/limitations or other simulator system limitations*" [7, p. 95].
- "*The overall trend of the PSD plot should be considered while focusing on the dominant frequencies*" [7, p. 131].

- "*The appearance and trend of the buffet's power spectra should match flight data with at least three of the predominant frequency spikes being within* ***±2 Hz*** *of the flight data frequency spikes*" [7, p. 105].
- "*Conduct an approach-to-stall with engines at idle and a deceleration of 1 knot/second. Check that the motion cues of the simulated buffet, including the level of buffet increase with decreasing speed, are representative of the actual airplane*" [7, p. 182/183].
- "*Tolerances on stall buffet are not applicable in case the first indication of the stall is the activation of the stall warning system (i.e. stick shaker/pusher, stall horn, ...)*" [7, p. 96].

B. Stall Buffet Model

In this paper, a limited-envelope aerodynamic stall model of the Cessna Citation II developed at our research group was used [14, 15, 19]. This model was identified from flight test data collected with TU Delft's Cessna Citation II research aircraft and includes the quasi-steady stall aerodynamics and buffet dynamics based on Kirchhoff's theory of flow separation [20–22]. The model explicitly accounts for the flow separation point X, which ranges from 0 to 1, where 1 represents a fully attached flow and 0 means fully separated flow. The dynamics of X are modeled by a first-order differential equation:

$$\tau_1 \frac{\mathrm{d}X}{\mathrm{d}t} + X = \frac{1}{2}\left(1 - \tanh\left[a_1\left(\alpha - \tau_2\dot{\alpha} - \alpha^*\right)\right]\right) \tag{1}$$

Eq. (1) shows that the Kirchhoff model has four parameters for characterizing the stall dynamics, i.e., τ_1, τ_2, a_1 and α^*. The time constant τ_1 characterizes the effects of flow inertia, i.e., the time the air flow needs to readjust to a new condition. τ_2 models the effects of hysteresis during flow re-attachment. a_1 is a stall abruptness parameter and α^* equals the stall angle of attack where the flow separation point X is equal to 0.5. The effects of varying the Kirchhoff model parameters on C_L and X have been investigated and are presented in past research [14, 16, 22]. In our stall buffet model, as described by Van Horssen et al. [14], also the temporal variations in stall buffet intensity are directly linked to the flow separation state X.

The stall buffet model proposed in Ref. 14 was derived from buffet vibrations measured during flight tests with TU Delft's Cessna Citation II laboratory aircraft. The Power Spectral Densities (PSDs) of measured vertical and lateral accelerations during quasi-steady stall maneuvers were used to model the stall buffet frequency spectrum, see Fig. 1. Fig. 1 also shows the dominant frequencies of the buffet vibrations. In the vertical direction there is only one dominant frequency peak at around 12 Hz, while in lateral direction two main peaks are observed, at 6 Hz and 10 Hz. The longitudinal accelerations are not shown as stall buffet accelerations were found to be negligible in the surge direction [14]. Fig. 1 also shows that, as expected, the PSD of the vertical buffets is approximately 10 times larger in magnitude than for a_y. Hence, in this paper we focus on measuring how noticeable variations in the vertical stall buffet are for human pilots.

The vertical stall buffet model proposed by Van Horssen et al. [14], see Fig. 2, passes unity-variance white noise through a second-order shaping filter $H(s)$, given by Eq. (2), that accounts for the average frequency spectrum of the buffet vibrations (see Fig. 1). The resonance frequency ω_0 of the second-order filter is used to create a band-pass filter focused on the 12 Hz (75 rad/s) frequency spike that dominates the vertical buffet characteristics of the Citation II.

$$H(s) = \frac{H_0\omega_0^2}{s^2 + \frac{\omega_0}{Q_0}s + \omega_0^2} \tag{2}$$

The parameters in the buffet model of Eq. (2) were estimated by fitting the PSD of the filter output to the PSD of the raw buffet flight data, see Fig. 1. The baseline parameter values of $H(s)$ reported in Ref. 14 were a gain H_0 of 0.125, a resonance frequency ω_0 of 75.92 rad/s (12.08 Hz), and a quality factor Q_0 of 8.28. As shown in Fig. 2, the shaping filter output is further multiplied with a factor $1 - X$ to account for buffet intensity variations with the level of separated flow along the wing. Finally, the stall buffet model uses a threshold on X, i.e., X_{thres}, to trigger the buffet model: only when $X < X_{thres}$ (here equal to 0.89) is the stall buffet model active and adding vibrations to the aircraft's simulated vertical acceleration.

C. Quasi-Steady Stall Buffet Simulations

For the buffet model simulations, the Cessna Citation II stall dynamics model from Ref. 15 that includes the buffet model from Ref. 14 has been implemented in a Matlab/Simulink framework. A simulated "stall autopilot" developed for

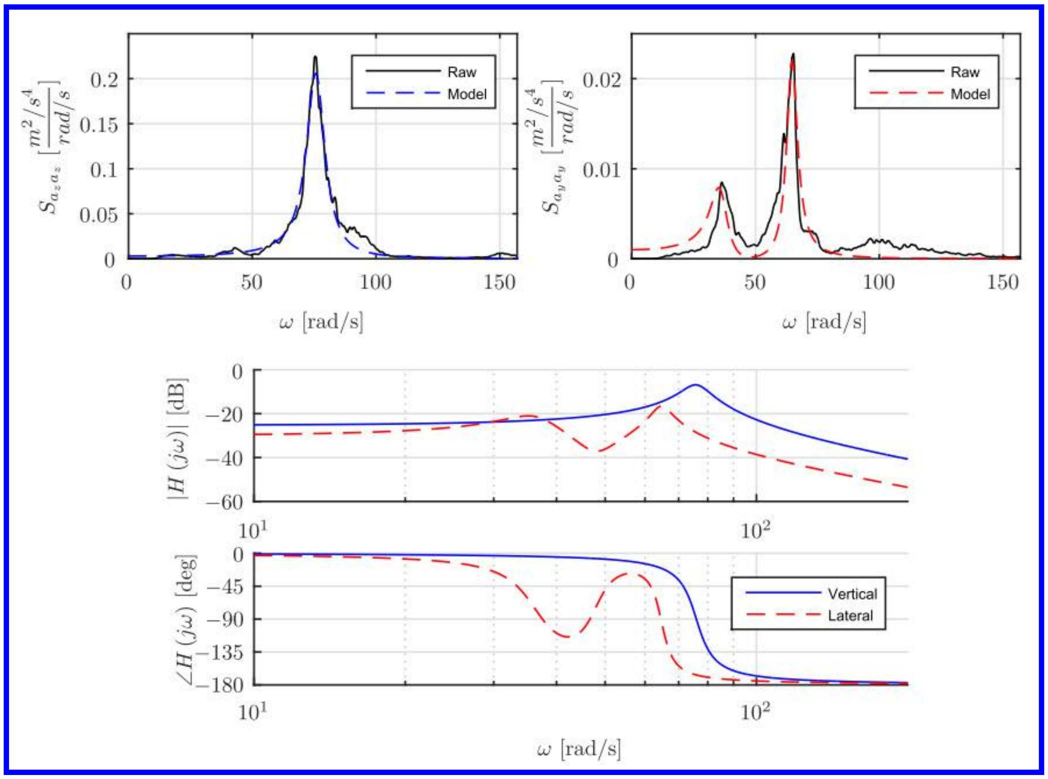


Fig. 1 Measured and modeled PSDs and buffet model Bode diagrams for buffet accelerations in vertical and lateral directions from Ref. 14.

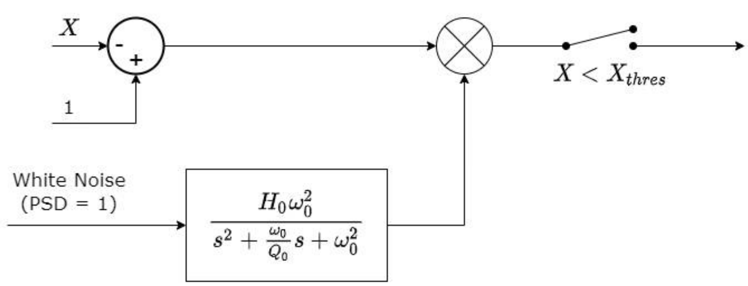
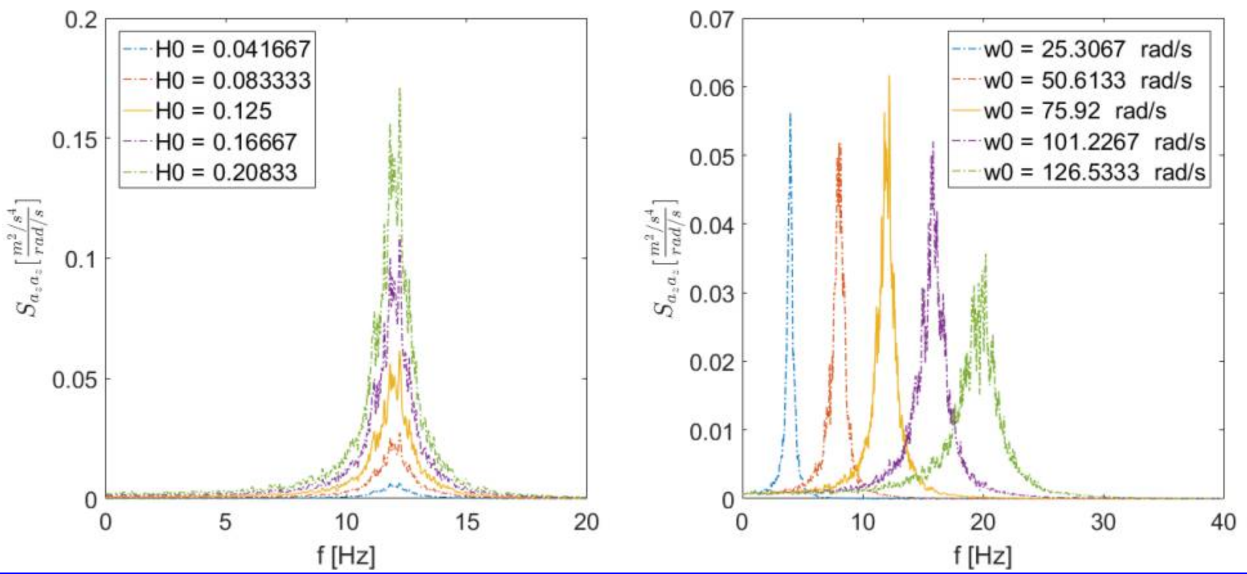
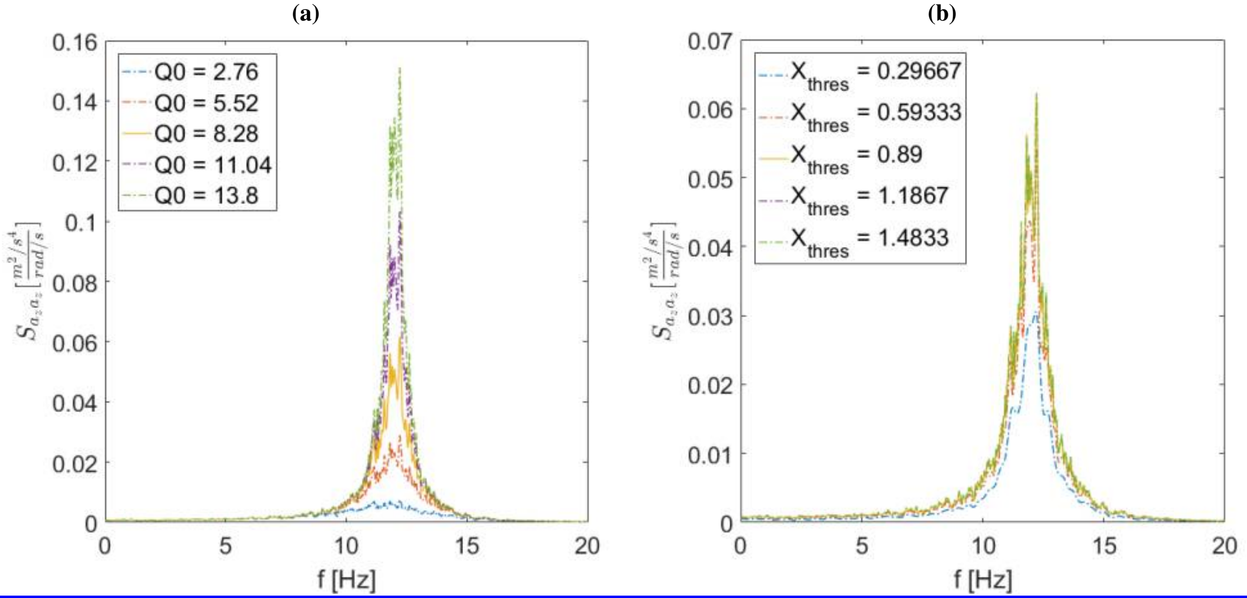


Fig. 2 Schematic overview of the vertical stall buffet model proposed in Ref. 14.

the earlier experiment of Ref. 16 was implemented to consistently perform 1 kt/s deceleration and the quasi-steady stall maneuver. Using this model, the effects of varying the buffet model parameters (H_0, ω_0, Q_0 and X_{thres}) was evaluated. Fig. 3 shows example symmetric quasi-steady vertical stall buffet simulation results in the frequency domain (PSD of a_z) for variations of -66%, -33%, 0%, +33% and +66% in all buffet model parameters individually with respect to their baseline values. The PSD plots are calculated experimentally using the time simulation results from the buffet model and the Fast Fourier Transform (FFT) algorithm in Matlab. To get a smoother estimate of the PSD, the results from 50 different buffet simulations (with different driving noise realizations) have been averaged.



(a) **(b)**



(c) **(d)**

Fig. 3 Example symmetrical quasi-steady vertical stall buffet model simulation results showing the PSD of a_z under variation of the buffet model parameters.

In addition to the PSD sensitivities shown in Fig. 3, the sensitivity of the buffet model in the time domain was analysed, see Fig. 4. This was done by calculating the Variance Accounted For (VAF) values between the baseline buffet model a_z time response and the buffet model output with offsets in the model parameters (i.e., $\hat{a_z}$) according to Eq. (3) and averaged over a similar set of 50 different noise realizations applied at the buffet model input. The VAF

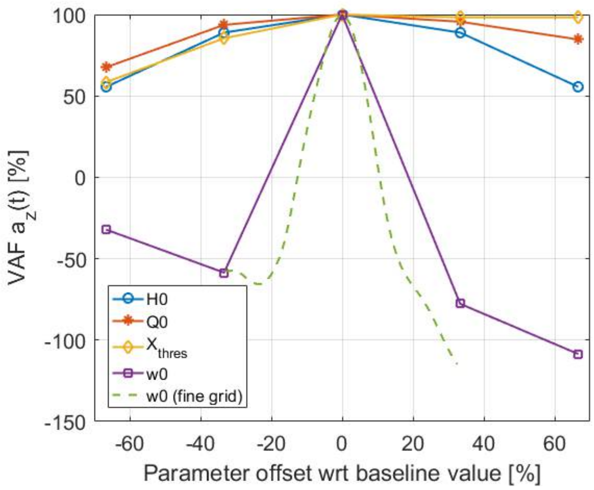


Fig. 4 VAF of the vertical buffet model output (a_z) presented as a function of model parameter offsets.

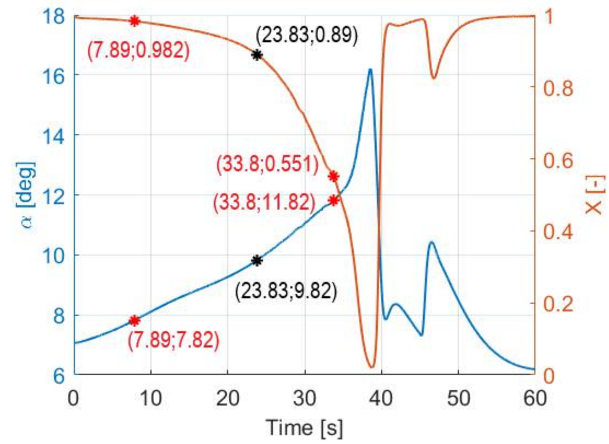


Fig. 5 Quasi-steady stall simulation results of the Cessna Citation II stall model with baseline settings (Angle of Attack α and Flow Separation Point X can be seen) showing the current buffet onset point (X = 0.89) in black as well as the limits for simulator qualification set by the FAA on buffet onset threshold in red.

quantifies the variance that is common between two time signals; the closer it is to 100 percent, the more identical the two signals are.

$$\text{VAF} = \left(1 - \frac{\sum_{i=1}^{N}(a_z(t_i) - \hat{a_z}(t_i))^2}{\sum_{i=1}^{N} a_z(t_i)^2}\right) \cdot 100\% \tag{3}$$

As shown in Fig. 3, the characteristic frequency parameter ω_0 has the biggest influence on the frequency components that are present in the simulated buffet. It directly defines the frequency at which the dominant spike in the buffet spectrum occurs. According to regulatory standards from the FAA for qualification of stall dynamic models in FSTDs, the dominant frequency spikes of a simulated buffet should be within a tolerance of ±2 Hz compared to reference flight data [7]. For our current model, this FAA tolerance implies a maximum 17% offset with respect to the baseline ω_0 (12 Hz). Fig. 4 shows that the VAF value for such an offset in ω_0 already drops below the 0% mark, which indicates significant differences in buffet time response exist at this tolerance limit.

Another parameter closely related to FSTD stall buffet requirements is X_{thres}, which controls the buffet onset point in our stall buffet model. Lowering the threshold value would delay buffet onset in the model and increase the initial buffet amplitude. According to FAA qualification standards, the buffet threshold of perception should be based on a 0.03g peak-to-peak initial normal acceleration with a tolerance of ±2.0 degrees in angle of attack [7]. In the specific case of quasi-steady stall simulations of the Cessna Citation II model, this translates to the limits shown in Fig. 5. Fig. 5 shows a considerable time frame of around 26 seconds that falls within the ±2.0 degrees angle of attack range around the nominal buffet onset threshold point (X = 0.89). When increasing the X_{thres} parameter, almost no change in VAF is observed with respect to the baseline settings, i.e., the VAF remains equal to around 100%, see Fig. 4. As this invariance indicates that increased X_thres values compared to the baseline would likely not be noticeable for pilots, an upper JND threshold for X_thres parameter was not measured during the experiment.

As is clear from Fig. 3 and Fig. 4, the H_0 and Q_0 parameters have similar effects on the simulated buffet model output. They both lift the buffet power spectra up and down around the peak, changing the buffet intensity. No clear requirement is set on this stall buffet characteristic from regulatory standards. Ideally this peak should match the power spectra of the measured flight data as closely as possible and these model parameters can thus be identified accordingly by matching the peak heights. Hence, the parameters H_0 and Q_0 were not further investigated in this research.

In conclusion, based on analysis of the sensitivity of the buffet model responses to variations in H_0, ω_0, Q_0 and X_{thres} as shown in Fig. 3 and Fig. 4, the experiment only measured the JND thresholds for two key parameters that characterize the frequency content and temporal amplitude variations in the stall buffet vibrations: ω_0 and X_{thres}, respectively. The other parameters remained fixed at their baseline values as determined in Ref. 14 throughout the experiment.

D. JND Threshold Hypotheses

Two hypotheses regarding the expected JND thresholds for ω_0 and X_{thres} were formulated for the human-in-the-loop experiment. The hypotheses are both based on the results of the buffet model sensitivity analysis (see Section II.C) and enable verification of the tolerances set by the FAA on simulated buffet responses, see Section II.A.

H1 *The lower JND threshold for X_{thres} is percentage-wise larger than the upper and lower JND thresholds for ω_0.* Based on the sensitivity analysis in the time domain (see Fig. 4), the same percentage-wise offset clearly results in stronger VAF changes for ω_0. As this implies that changes in ω_0 are likely more noticeable, it is expected that the JND threshold for changes in ω_0 is lower than for X_{thres}.

H2 *The upper and lower JND thresholds for ω_0 are symmetric with respect to the baseline value.* The green dashed line in Fig. 4 shows a symmetric variation of the VAF around the baseline value for ω_0 until an offset of about 20%. The JND thresholds are expected to be within this symmetric portion of the figure because the FAA tolerance of ±2 Hz on buffet characteristic frequency suggest a maximum allowable variation of about 17% offset with respect to the baseline value of ω_0 (12 Hz). Hence, the JND threshold results for ω_0 are expected to be symmetric around its baseline value.

III. Methods and Experiment Setup

This paper describes the results of a human-in-the-loop experiment that was performed to measure the sensitivity of human pilots to variations in the two most relevant parameters of the stall buffet model, i.e., ω_0 and X_{thres} (see Section II.B). The goal of the experiment was to determine threshold values on the allowable variation in ω_0 and X_{thres} parameters before the parameter changes become noticeable. To measure these JNDs from different pilots, the same experimental paradigm also used by Smets et al. [16] was used. Here participants experience (as observers, i.e., a stall autopilot flew the maneuver) different sets of two sequential quasi-steady stalls, of which one represented the baseline parameter settings and the other a modified buffet parameter (ω_0 or X_{thres}) setting. Using a subjective staircase measurement procedure that also includes "null measurements" (i.e., baseline-to-baseline comparisons) to estimate the reliability of pilots' responses [16], thresholds were estimated for 21 active pilots.

A. Apparatus

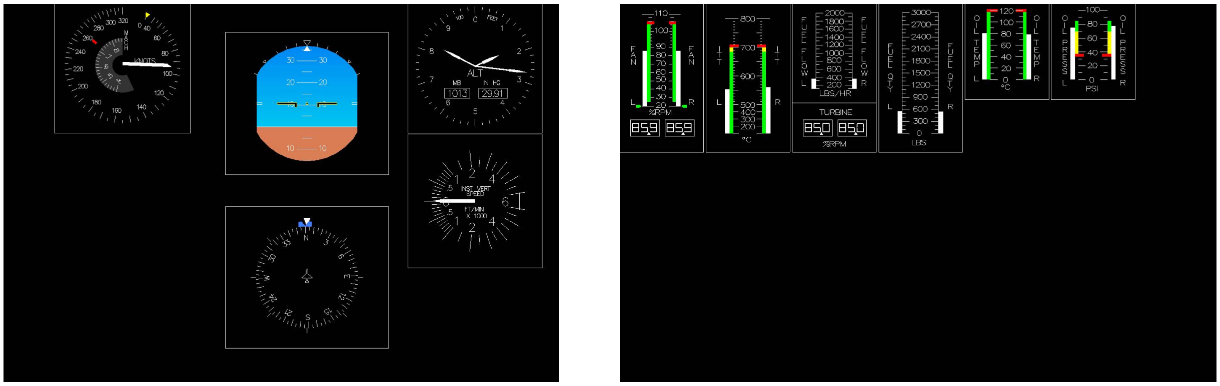
The SIMONA Research Simulator (SRS) at TU Delft was used to perform the experiment (see Fig. 6). The existing Cessna Citation II simulation environment available in the SRS was used, with our custom stall dynamics model implemented [14, 15, 19]. Participants were seated in the left pilot seat in the SRS cockpit (see Fig. 6b) and wore a noise-canceling headset to mask any noise coming from the simulator motion system.

(a) Outside

(b) Inside

Fig. 6 The experiment setup in the SIMONA simulator at TU Delft.

During the experiment, outside visual cues were generated using a FlightGear database and projected onto the 180x40 deg collimated screen of the SRS with an update rate of 60 Hz [23]. Next to the outside visual, also head-down visual cues were provided using the primary and secondary flight displays available in the SRS. Those displays showed in-house developed primary flight instruments and engine parameters, see Fig. 7.



(a) Primary flight display **(b) Secondary flight display**

Fig. 7 The primary and secondary flight displays, showing conventional flight instruments and engine information, used during the experiment.

The SRS's motion system has a six degree-of-freedom (DOF) hydraulic hexapod configuration, which can provide motion feedback at low latency and high accuracy [23, 24]. The experiment only focuses on symmetrical quasi-steady stall simulations. Hence, the asymmetric DOFs (i.e., roll, yaw and sway) were not used. The symmetric DOFs (i.e., pitch, surge and heave) were cued using a classical washout filter algorithm [25]. The used motion filter settings that remained constant during the experiment are listed in Table 1. Pitch (q) and surge (x) settings are set to typical reference values, identical to the ones used in Ref. 16. The heave settings, i.e. K_z and ω_{n_z}, were optimized using a "Gouverneur" analysis [26], as can be seen in Fig. 8. In this figure, each dot represents a filter configuration with K_z varying from 0.1 to 1 horizontally and ω_{n_z} from 0.1 to 4 rad/s vertically. The colored lines separate the feasible configurations (left and above the boundaries) from the ones where the simulator would hit its limits. The different colored boundaries represent stall simulations with different stall buffet model parameter settings, i.e. the most extreme experiment conditions possible and the baseline settings.

Ideally, a heave motion setting is chosen above and to the left of the colored lines, but as close to the right bottom corner as possible. That would result in the highest possible fidelity of the motion filter according to the available fidelity criteria [27, 28]. Three different settings were analysed during testing of the experiment, shown with blue or red circular markers in Fig. 8. They correspond to a heave gain K_z of 0.5 and break frequency ω_{n_z} equal to 3.0, 2.0 or 1.2 rad/s, respectively. The 3.0 rad/s represents the original (safe) setting used in Ref. 16. Here it was investigated if a more optimal setting closer to the boundaries was possible for the current experiment. A setting of 1.2 rad/s was considered, as this would ensure matched heave and surge motion filter settings. However, a break frequency of 2.0 rad/s was selected as the final heave motion setting (red marker in Fig. 8), because with ω_{n_z} = 1.2 rad/s the simulator would hit its upper motion space limit after the simulated stalls were stopped, due to the considerable vertical accelerations during the simulated stall recovery. Note that this post-trial simulator movement was not included in the analysis of Fig. 8.

Table 1 Motion filter parameters used for the experiment.

High-pass filters						Low-pass filters	
ω_{n_q}	1.0 rad/s	ω_{n_z}	2.0 rad/s	ω_{n_x}	1.2 rad/s	ω_{n_x}	2.4 rad/s
ζ_q	0.5	ζ_z	0.7	ζ_x	0.7	ζ_x	0.7
ω_{b_q}	0.0 rad/s	ω_{b_z}	0.3 rad/s	ω_{b_x}	0.0 rad/s		
K_q	0.5	K_z	0.5	K_x	0.5		

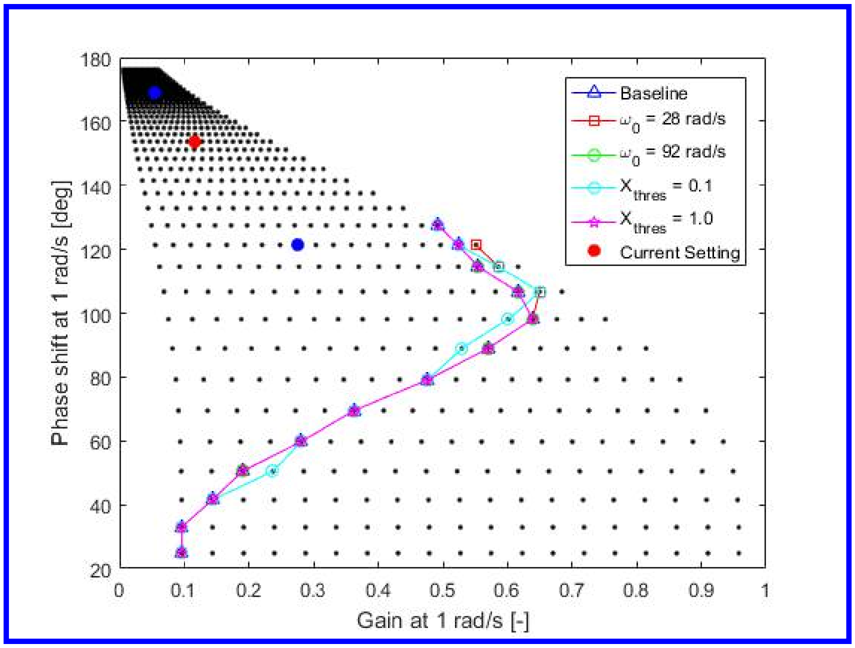


Fig. 8 Heave Gouverneur analysis [26] results for the baseline simulation settings and four extreme experiment condition settings. Configurations above and to the left of the colored lines ensure that the simulator remains within its motion limits.

B. Experiment Procedures and Conditions

The experiment measured an upper and a lower threshold for ω_0 and a lower threshold for X_{thres}. Upper thresholds are identified with a + superscript (ω_0^+), while lower thresholds are identified using a − superscript (ω_0^- and X_{thres}^-). This resulted in an experiment with three different conditions. The three experiment conditions (i.e., X_{thres}^-, ω_0^+ and ω_0^-) were performed by the participants in a balanced and randomized order based on a Latin square design.

Every condition was tested using the yes/no staircase procedure (see Fig. 9) also used in the similar JND threshold experiment of Ref. 16. Every trial consisted of a comparison between two sequential quasi-steady stalls of around 15 seconds each, which ran from stall onset up until the recovery was initiated. One of the two stalls represented the baseline buffet settings, while the other had an offset in either ω_0 or X_{thres}, depending on the experiment condition. The order in which the baseline and offset parameter stall were presented varied randomly across trials.

At the end of the second stall, participants were asked to verbally answer the yes/no question: "*Did you notice a difference?*". The given answer of the participants then determined the next parameter offset value. A "yes" answer would update the parameter closer to the baseline value, while a "no" answer would increase the parameter offset. From the second reversal (i.e., a "yes" answer after a "no" or the other way around) onward, the step size reduced by 50% at every reversal in the answers. If four consecutive answers were identical, the step-size doubled to converge faster to the threshold. Such an approach is referred to as an adaptive 1-up/1-down staircase method, which results in a 50% level of correctness JND threshold [29]. The researcher ended the staircase procedure, see Fig. 9, either when the participant's staircase had converged (i.e., the step size was reduced to a value lower than 1/32th of the initial step size) or when a total of 30 comparisons was performed.

The initial starting position of the staircase for each experiment condition was chosen relatively far away from the baseline, such that the initial parameter offset was obvious to all participants. The initial step size was chosen accordingly to ensure convergence within a reasonable amount of trials. The initial parameter values and step sizes for each experiment condition are listed in Table 2. To mitigate the risk of participants following identical paths through the staircase procedure and increase the variety in the collected staircase data, small random parameter variations were added at every parameter update. The amplitude of the added white noise variations was set to 90% of the smallest allowed step size, making the noise more dominant close to the threshold value.

Prior to the start of the experiment, participants were briefed on the safety and experiment procedures. They were instructed to only answer verbally on the yes/no question: "*Did you notice a difference?*" at the end of two consecutive quasi-steady stalls, and that no further explanation was required. No information about the goal of the experiment, the tested conditions or the data analysis was provided. Participants were trained for about 15 minutes, where they

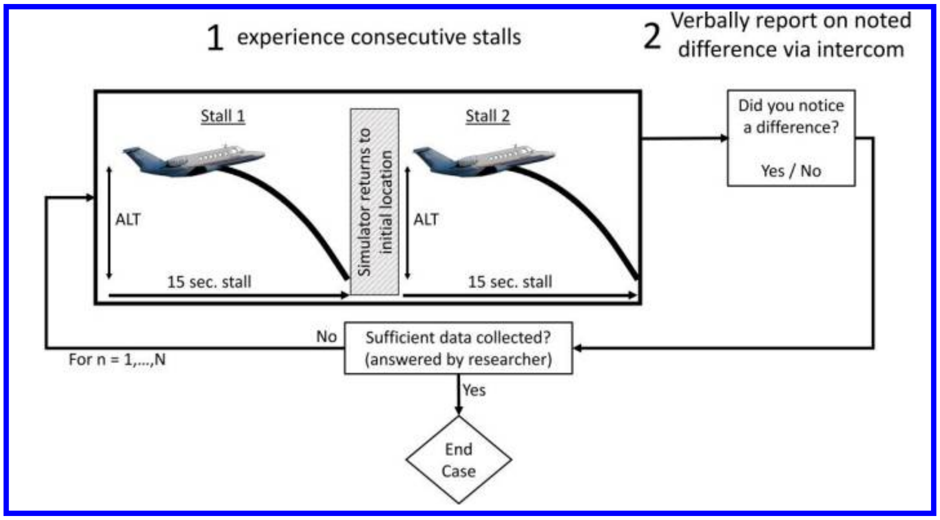


Fig. 9 Graphical representation of the experiment procedure and repeated consecutive simulated stalls, adapted from Ref. 16.

Table 2 Initial staircase values for each experiment condition.

	X^-_{thres}	ω_0^+	ω_0^-
Initial value	0.55	92 rad/s	28 rad/s
Initial step	0.1	-5 rad/s	15 rad/s

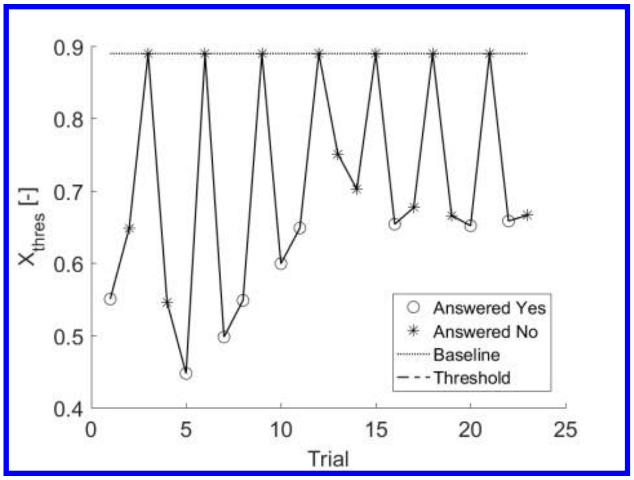
could practice detecting the differences between the two stalls using example runs of the different experiment conditions. Breaks were taken between the different test conditions to limit fatigue artifacts in the data.

C. Data Analysis

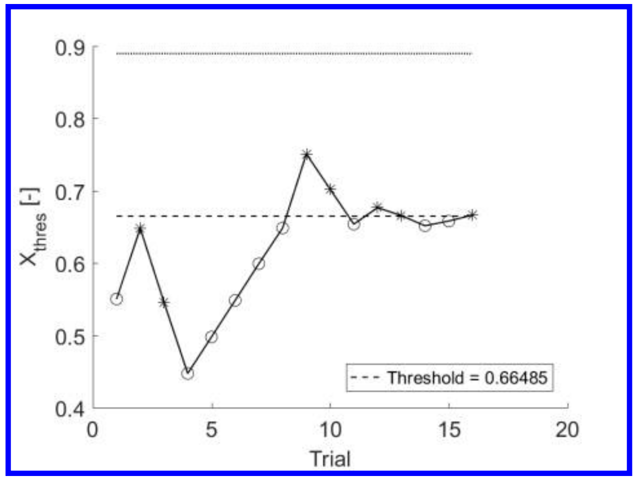
The threshold values for X^-_{thres}, ω_0^+ and ω_0^- were determined for each participant individually from the staircase data sets collected using the experiment procedure shown in Fig. 9. Such yes/no-staircase procedures are prone to biases [29, 30]. In order to detect such biases in our data, additional "null measurements" were performed at every third trial, see Fig. 10a. A null measurement was a comparison of a stall with baseline buffet model settings with itself, i.e., the two experienced stalls were identical. The answers that participants gave to the null measurements were only used to assess their consistency and reliability and had no influence on the staircase procedure itself. The final threshold participants converged to was calculated after removing the null measurements and taking the average of the last three reversals in the staircase data, see Fig. 10b for the same example X^-_{thres} data shown with the null measurements included in Fig. 10a.

D. Experiment Participants

A total of 21 active pilots (1 female and 20 males) participated in the experiment. The participants were from different age categories and they were divided into two groups to investigate potential between-group differences in the measured JND thresholds. The first group (n = 6) consisted of pilots with an active Cessna Citation II type rating, i.e., they were current on the specific aircraft simulated in the experiment. The second group (n = 15) consisted of a combination of glider, private and commercial pilots. All participants signed an informed consent form prior to starting the experiment. The study was approved by TU Delft's Human Research Ethics Committee (HREC) under application number 1741.



(a) Staircase example with null measures for X_{thres} lower.



(b) Staircase example without null measures for X_{thres} lower.

Fig. 10 Example staircase data for X^-_{thres}. Note that the performed staircase also includes "null measurements" (see (a)) that were used to verify participant's consistency.

IV. Results

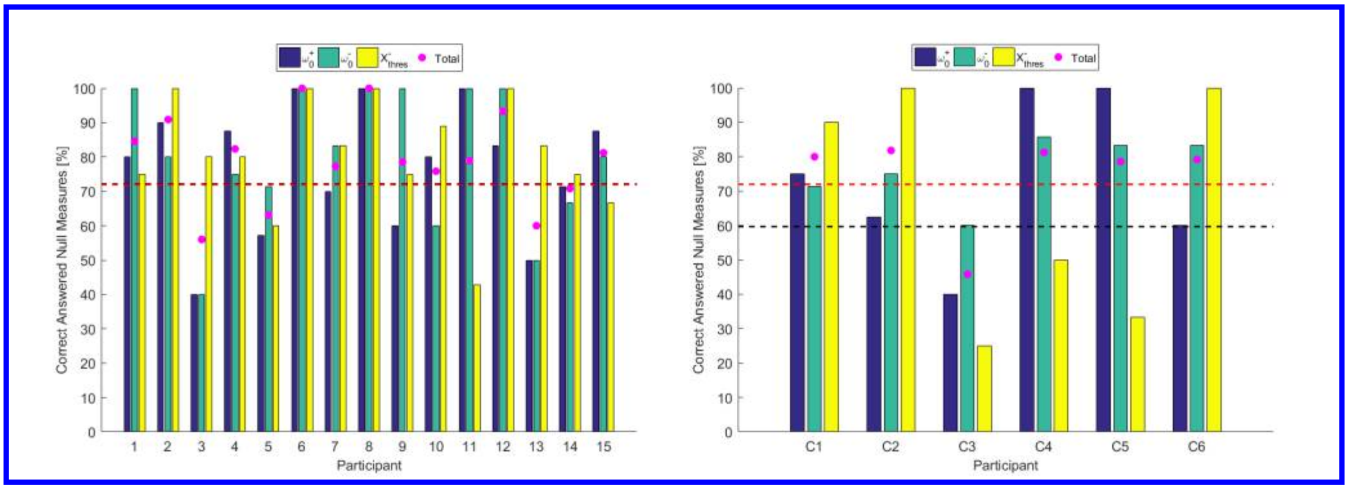
A. Participant Reliability

The consistency and reliability of participants' responses was based on their answers to the staircase's "null measurements", i.e., the comparisons of two identical stall simulations with baseline buffet model settings. Fig. 11 shows the percentage of correctly answered null measurements (i.e., when participants provided a "no" answer on the null measurements) for all 21 pilots. Separate colored bars indicate the individual results for the three different experiment conditions while the magenta circular markers show the average consistency percentage for each participant. The threshold to determine if participants' staircase data were reliable or not was based on the lower bound of the 95% confidence interval on the average total consistency. For our collected data, this resulted in a limit of 72.15% for the general pilot group (see Fig. 11a) and 59.68% for the group of Citation II pilots (see Fig. 11b). The dashed black lines in Fig. 11 indicate these boundaries. For both groups combined, the overall consistency limit was found to be 71.98%, see the red dashed lines in Fig. 11. Fig. 11 shows that a total of four general pilots (participants 3, 5, 13 and 14) and one Citation II pilot (participant C3) had a total consistency below both the group-specific and overall consistency limits. Hence, these participants were insufficiently reliable and consistent in their answers and their data was not considered for further analysis.

B. JND Thresholds for Stall Buffet Model Parameters

The experiment procedure explained in Section III.B was used to gather staircase data for each of the three experiment conditions, i.e., X^-_{thres}, ω_0^+ and ω_0^-. Average thresholds (across all consistent participants) were estimated by fitting a Gaussian Cumulative Density Function (CDF) through all staircase data [31]. Fig. 12 shows the fitted CDFs for all three experiment conditions, where a "yes" answer is represented as 1 (100%) and a "no" answer as a 0 (0%). The CDFs show 50% level of correctness thresholds that are unreliably close to the parameters' baseline values (vertical blue lines), especially for ω_0, see Fig. 12b and Fig. 12c. This is caused by some individual thresholds of participants being very close to the baseline value for ω_0 (see also Fig. 13b). With such between-participant spread and only limited data points at extreme values further away from the baseline due to our optimized staircase procedure, the CDF fits are insufficiently robust for extracting a reliable 50% correctness threshold. Therefore, instead we estimate the JND thresholds, for each participant separately, directly from the staircase data, i.e., by averaging over the last three reversals as shown in Fig. 10b. These estimated JND thresholds are shown in Fig. 13 for all three experiment conditions using a Boxplot representation. Note that the results shown in Fig. 13 only include the consistent participants, see Fig. 11.

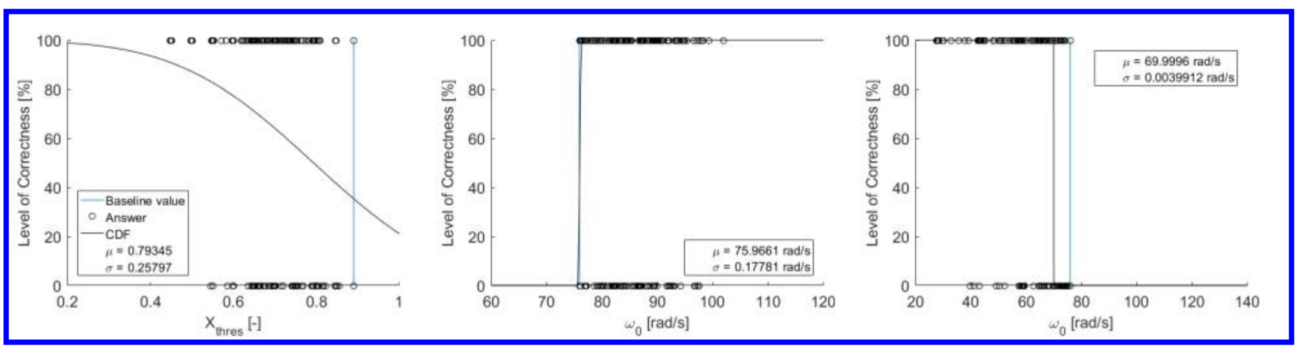
Fig. 13 shows the JND threshold results for the general pilots (green boxes) and Citation II pilots (purple boxes) separately, as well as for all data combined (yellow boxes). The left vertical axis indicates the absolute values of the JND thresholds, while the right vertical axis shows them expressed as Weber fractions, i.e., as percentage-wise difference



(a) Glider, private and commercial pilots

(b) Pilots with Cessna Citation II type rating

Fig. 11 Percentage of correctly answered null measurements (baseline comparisons) for two groups of experiment participants.



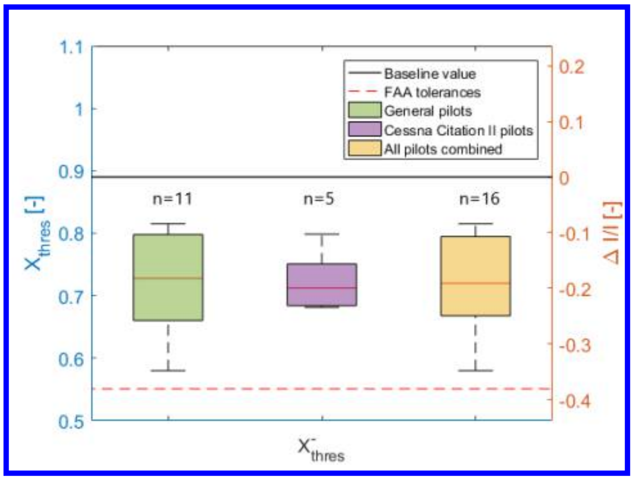
(a) Fitted CDF through X^-_{thres} answers

(b) Fitted CDF through ω^+_0 answers

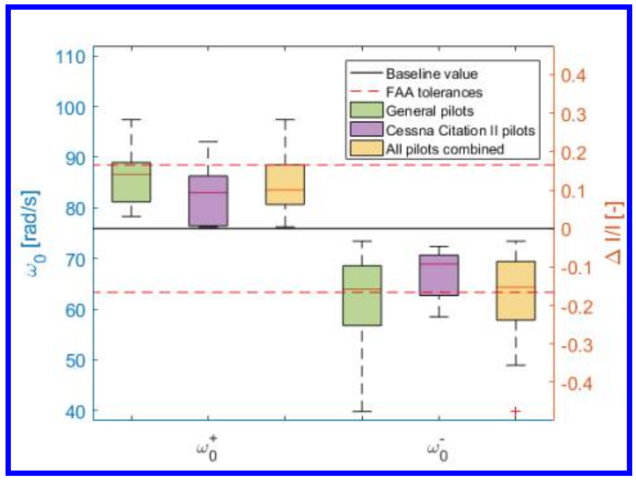
(c) Fitted CDF through ω^-_0 answers

Fig. 12 Cumulative Density Function fits through all consistent experiment data for all three estimated thresholds.

with respect to the baseline value. First, Fig. 13 shows that no differences are found between the results for both pilot groups. While the Citation pilot group showed slightly lower JNDs for both ω_0 thresholds, this effect is not statistically significant. Overall, this lack of between-group differences was expected, as the experiment focused on detecting relative differences in simulated stall buffet accelerations, not on absolute comparisons with in-flight experiences. Due to the similarity in outcomes between both groups, the final observations and conclusions are based on the combined data (yellow boxes in Fig. 13) of the general and Citation II pilots.



(a) JND thresholds for X_{thres}



(b) JND thresholds for ω_0

Fig. 13 Upper and lower JND threshold values for X_{thres} and ω_0 estimated from the staircase experiment data.

For X_{thres}, which has a baseline value of 0.89, Fig. 13a shows that the lower average JND threshold, expressed as the 95% confidence interval on the mean across all pilots, equals X^-_{thres} = 0.72 ±0.037 (-0.1896 ±0.0412 Weber fraction). For ω_0, with a baseline value equal to 75.92 rad/s, Fig. 13b shows estimated upper and lower average thresholds that are, respectively, ω_0^+ = 84.87 ±3.295 rad/s (0.1179 ±0.0434 Weber fraction) and ω_0^- = 63.04 ±4.813 rad/s (-0.1696 ±0.0634 Weber fraction). For ω_0, we thus found that an increase in the buffet peak frequency was slightly more noticeable (lower absolute Weber fraction) than a reduced ω_0. Furthermore, comparison of Figures 13a and 13b shows that the JND thresholds for X_{thres} and ω_0 are quite similar, i.e., with maximum allowed parameter offsets of between 10-20%. This indicates that differences in buffet onset, as controlled with our X_{thres} parameter, are more noticeable to pilots than would be expected based on the similarity of time-domain buffet acceleration traces (see Fig. 4).

Finally, the JND threshold results were compared to the the FAA tolerances (see Section II.A) for simulated buffet responses, see the red dashed lines in Fig. 13. For the buffet frequency ω_0, the measured JND thresholds are found to closely match with the ±2 Hz tolerance set by the FAA. On the other hand, the average JND threshold for X^-_{thres} (0.72) is well above the tolerance (±2.0° angle of attack) that the FAA requires for the buffet onset (X_{thres} = 0.551), indicating that pilots may notice differences in buffet onset even within this tolerance value.

V. Discussion

The human-in-the-loop experiment described in this paper aimed at providing additional quantitative guidance on the minimum required accuracy of simulated stall buffets. For this we measured JND thresholds for two key stall buffet model parameters, i.e., the buffet shaping filter's characteristic frequency ω_0 and the flow separation threshold parameter X_{thres} that parameterizes the buffet onset point. These two parameters are most closely related to current FSTD stall buffet requirements as described by the FAA. During the experiment, participants experienced, as passive observers, consecutive quasi-steady symmetric stalls (one with baseline settings and one with the modified buffet) simulated with the TU Delft stall dynamics model of the Cessna Citation II aircraft. Through a subjective staircase procedure, JND thresholds for ω_0 and X_{thres} variations were determined.

The sensitivity analysis results of quasi-steady stall buffet simulations showed almost no differences in the buffet model responses for an increase of the X_{thres} parameter, so it was concluded that an upper JND threshold would likely not be measurable from human pilots. Looking at the buffet model structure (see Fig. 2), the shaping filter's output is multiplied with a factor 1-X. Hence, at baseline settings (X_{thres} = 0.89) the simulated buffet starts at about 10% of its maximum amplitude, which is reached at full flow separation (X=0). Increasing the threshold value to the maximum possible value of 1 (X physically only varies between 0 and 1), would only make the buffet start with that 10% lower amplitude. During pre-testing of the experiment it was indeed confirmed that this upper threshold on X_{thres} was not sufficiently noticeable due to these subtle differences in amplitude at buffet onset compared to the baseline.

Two hypotheses were formulated based on the buffet model sensitivity results. First, it was expected that the JND thresholds for changes in ω_0 would be lower than those for X_{thres} (Hypothesis **H1**), because simulated buffet time responses showed much greater differences with the baseline model's response when ω_0 was subjected to the same percentage-wise variation. Based on the collected experiment data, however, this Hypothesis **H1** must be rejected, as the average JND threshold for X_{thres} was found to be much lower than expected and comparatively similar to the average upper and lower thresholds for ω_0. In fact, all three average JND thresholds were between a 10 and 20% maximum offset with respect to their respective baseline parameter values.

The sensitivity analysis further showed that for a similar percentage-wise increase or decrease in ω_0 up until an offset of around 20%, the VAF of the offset buffet model output compared to the baseline seemed to be symmetric. Hence, for ω_0 the upper and lower absolute JND thresholds were expected to also be symmetric with respect to the baseline (Hypothesis **H2**). However, the experiment data showed that the upper JND threshold was, on average, slightly closer to the baseline value (12% Weber fraction) than the lower threshold for ω_0 (-17% Weber fraction), indicating that an increase in ω_0 is in fact more noticeable. Based on this observation, Hypothesis **H2** was also rejected. A possible explanation for the upper threshold being closer to the baseline for ω_0 is that an increase in buffet frequency results in more intense shaking, which may be observed from increased vibrations of cockpit subsystems, such as the SRS's Mode Control Panel (MCP), while a reduction in ω_0 also reduces the perceived intensity of vibrations. While this difference in high-frequency vibration frequency is exactly what we tried to measure, a few participants indicated that they had picked up on the effects the ω_0^+ case had on cockpit structure vibrations at the end of the experiment.

Finally, the estimated JND thresholds were compared to current standards on stall buffet simulation qualification requirements. The thresholds for ω_0 measured in our experiment were found to support of the ±2 Hz tolerance defined on buffet characteristic peak frequency, as our measured ω_0^- and ω_0^+ are very close this 2 Hz limit. However, the lower JND thresholds for X_{thres} were found to be much closer to the baseline value than the maximum tolerance of ±2.0° angle of attack set for the buffet onset threshold of perception. This indicates that human pilots may already notice the difference in buffet onset characteristics well before the X_{thres} parameter offset reaches the maximum allowed tolerance, which suggests that the buffet onset requirements for quasi-steady stall training models may require stricter tolerances.

The current experiment attempted to measure the JND thresholds for variations in two key stall buffet model parameters by simulating quasi-steady symmetric stall dynamics of a Cessna Citation II aircraft. This is due to the current stall dynamics model only being validated for the quasi-steady symmetric flight envelope at an altitude of 5,500 m (18,000 ft). Ongoing research at our group currently focuses on modeling asymmetric stall dynamics, as well as investigations on altitude variations to increase the validation region of the Cessna Citation II stall model. In future experiments, we aim to collect similar stall model JND thresholds also for simulated accelerated and more sudden symmetric stall maneuvers.

A key drawback of the used 1-up/1-down staircase procedure is that it only allows for estimating a 50% level of correctness JND threshold. With the insight gained on the JND thresholds for key stall buffet model parameters from the current experiment, follow-up experiments should consider more accurate staircase procedures that would result in a higher level of correctness JND threshold (e.g., 70.7% for 1-up/2-down method). This, however, has the drawback of requiring more comparisons before staircase convergence, and hence a much longer experiment. Finally, future work should also focus on determining similar JND thresholds but in an experiment where participants actively fly themselves instead of being a passive observer, as it is uncertain if the current estimated thresholds are also applicable in case pilots are in active control of the aircraft and thus occupied. While this is a challenging experiment to design and execute, knowledge of JND thresholds in active flight is needed to further the research on required stall model fidelity for effective simulator-based stall training.

VI. Conclusions

In this paper we used subjective Just Noticeable Difference (JND) thresholds for two key stall buffet model parameters (X_{thres} and ω_0) to measure pilots' sensitivity to differences in simulated stall buffets. A buffet model sensitivity analysis and a human-in-the-loop experiment with 21 pilots in a flight simulator were used to provide additional quantitative guidance on the minimum required accuracy of simulated stall buffets. The experiment used a subjective staircase procedure where pilots experienced, as passive observers, sets of two consecutive quasi-steady symmetric stalls to compare (one of which was the baseline settings and the other one featured a modified buffet). The staircase procedure resulted in an upper and a lower JND threshold for ω_0. For X_{thres} only a lower threshold was estimated because an increase in the X_{thres} parameter (from a baseline value of 0.89 to its maximum 1.0) was found to hardly affect buffet model outputs and would thus be unnoticeable to human pilots. The current findings indicate that the estimated JND thresholds for all three conditions closely match each other (average thresholds between 10-20% offset), with the highest individual JND thresholds not exceeding 30-35% with respect to the baseline parameter values. This indicates that human pilots notice the differences in individual stall buffet model parameters already at comparatively small offsets. Furthermore, the upper and lower JND thresholds for ω_0 are found to support the ±2 Hz tolerances currently defined for simulated buffets' characteristic frequencies. However, the lower JND thresholds for X_{thres} were found to be well above the current tolerance of ±2.0° angle of attack for the buffet onset threshold of perception. This indicates that human pilots may be capable of noticing differences in buffet onset that are, in fact, within the currently accepted maximum tolerance.

References

[1] Boeing, "Statistical Summary of Commercial Jet Airplane Accidents, Worldwide Operations | 1959 – 2020," , 2021.

[2] Lambregts, A. A., Nesemeier, G., Wilborn, J. E., and Newman, R. L., "Airplane Upsets: Old Problem, New Issues," *Proceedings of the AIAA Modeling and Simulation Technologies Conference and Exhibit, Honolulu (HI)*, 2008. https://doi.org/10.2514/6.2008-6867.

[3] Crider, D. A., "Accident Lessons for Stall Upset Recovery Training," *Proceedings of the AIAA Guidance, Navigation, and Control Conference, Toronto, Canada*, 2010. https://doi.org/10.2514/6.2010-8003.

[4] J. V. Foster and K. Cunningham and C. M. Fremaux and G. H. Shah and E. C. Stewart and R. A. Rivers and J. E. Wilborn and W. Gato, "Dynamics Modeling and Simulation of Large Transport Airplanes in Upset Conditions," *AIAA Guidance, Navigation, and Control Conference and Exhibit*, American Institute of Aeronautics and Astronautics (AIAA), San Francisco, CA, USA, 2005. https://doi.org/{10.2514/6.2005-5933}.

[5] Advani, S. K., Schroeder, J. A., and Burks, B., "What Really Can Be Done in Simulation to Improve Upset Training?" *Proceedings of the AIAA Modeling and Simulation Technologies Conference, 2-5 August 2010, Toronto, Canada*, 2010. https://doi.org/10.2514/6.2010-7791.

[6] Advani, S. K., and Field, J., "Upset Prevention and Recovery Training in Flight Simulators," *Proceedings of the AIAA Modeling and Simulation Technologies Conference, Portland (OR)*, 2011. https://doi.org/10.2514/6.2011-6698.

[7] Federal Aviation Administration, "Code of Federal Regulations, Title 14, Part 60: Flight Simulation Training Device Initial and Continuing Qualification and Use," , 2016.

[8] Anonymous, "Stall and Stick Pusher Training," Advisory Circular 120-109, U.S. Department of Transportation, Federal Aviation Administration, Aug. 2012.

[9] Advani, S. K., and Schroeder, J. A., "Global Implementation of Upset Prevention & Recovery Training," *Proceedings of the AIAA Modeling and Simulation Technologies Conference, San Diego (CA)*, 2016. https://doi.org/10.2514/6.2016-1430.

[10] Lemley, C. and Mullans, R., *Buffeting pressures on a swept wing in transonic flight - Comparison of model and full scale measurements*, Structures, Structural Dynamics, and Materials and Co-located Conferences, American Institute of Aeronautics and Astronautics (AIAA), 1973. https://doi.org/10.2514/6.1973-311.

[11] Caruana, D., Mignosi, A., Corrège, M., Le Pourhiet, A. & Rodde, A. M., "Buffet and buffeting control in transonic flow," *Aerospace Science and Technology*, Vol. 9, No. 7, 2005, pp. 605–616. https://doi.org/10.1016/j.ast.2004.12.005.

[12] Marschalk, S., Luteijn, P. C., van Os, D., Pool, D. M., and de Visser, C. C., "Stall Buffet Modeling using Swept Wing Flight Test Data," *AIAA Scitech 2021 Forum*, American Institute of Aeronautics and Astronautics Inc, AIAA, 2021. https://doi.org/10.2514/6.2021-0286.

[13] J. A. Schroeder and J. Bürki-Cohen and D. A. Shikany and D. R. Gingras and P. Desrochers, "An Evaluation of Several Stall Models for Commercial Transport Training," *AIAA Modeling and Simulation Technologies Conference*, American Institute of Aeronautics and Astronautics (AIAA), National Harbor, MD, USA, 2014. https://doi.org/10.2514/6.2014-1002.

[14] Van Horssen, L. J., de Visser, C. C., and Pool, D. M., "Aerodynamic Stall and Buffet Modeling for the Cessna Citation II Based on Flight Test Data," *Proceedings of the AIAA Modeling and Simulation Technologies Conference, Kissimmee (FL)*, 2018. https://doi.org/10.2514/6.2018-1167.

[15] van Ingen, J. B., de Visser, C. C., and Pool, D. M., "Stall Model Identification of a Cessna Citation II from Flight Test Data Using Orthogonal Model Structure Selection," *Proceedings of the AIAA Modeling and Simulation Technologies Conference, Nashville (TN)*, 2021. https://doi.org/10.2514/6.2021-1725.

[16] Smets, S. C., de Visser, C. C., and Pool, D. M., "Subjective Noticeability of Variations in Quasi-Steady Aerodynamic Stall Dynamics," *Proceedings of the AIAA Modeling and Simulation Technologies conference, San Diego (CA)*, 2019. https://doi.org/10.2514/6.2019-1485.

[17] Schroeder, J. A., "Research and Technology in Support of Upset Prevention and Recovery Training," *Proceedings of the AIAA Modeling and Simulation Technologies Conference 2012, Minneapolis (MN)*, 2012. https://doi.org/10.2514/6.2012-4567.

[18] Federal Aviation Administration, "Advisory Circular - Subject: Upset Prevention and Recovery Training (AC 120-111)," Tech. rep., 2017.

[19] van den Hoek, M. A., de Visser, C. C., and Pool, D. M., "Identification of a Cessna Citation II Model Based on Flight Test Data," *Proceedings of the 4th CEAS Specialist Conference on Guidance, Navigation & Control, Warsaw, Poland*, 2017.

[20] Goman, M., and Khrabrov, A., "State-Space Representation of Aerodynamic Characteristics of an Aircraft at High Angles of Attack," *Proceedings of the AIAA Astrodynamics Conference, Hilton Head (SC)*, 1992. https://doi.org/10.2514/6.1992-4651.

[21] Fischenberg, D., and Jategaonkar, R. V., "Identification of Aircraft Stall Behavior from Flight Test Data," *Proceedings of the RTO SCI Symposium on System Identification for Integrated Aircraft Development and Flight Testing, Madrid, Spain*, 1998, pp. 17–1–17–8.

[22] Dias, J. N., "Unsteady and Post-Stall Model Identification Using Dynamic Stall Maneuvers," *Proceedings of the AIAA Atmospheric Flight Mechanics Conference, Dallas (TX)*, 2015. https://doi.org/10.2514/6.2015-2705.

[23] Stroosma, O., van Paassen, M. M., and Mulder, M., "Using the SIMONA Research Simulator for Human-machine Interaction Research," *AIAA Modeling and Simulation Technologies Conference and Exhibit*, American Institute of Aeronautics and Astronautics (AIAA), 2003. https://doi.org/10.2514/6.2003-5525.

[24] Berkouwer, W. R., Stroosma, O., van Paassen, M. M., Mulder, M., and Mulder, J. A., "Measuring the Performance of the SIMONA Research Simulator's Motion System," *AIAA Modeling and Simulation Technologies Conference and Exhibit*, American Institute of Aeronautics and Astronautics, San Francisco, California, 2005. https://doi.org/10.2514/6.2005-6504.

[25] Reid, L. D., and Nahon, M. A., "Flight Simulation Motion-Base Drive Algorithms: Part 1 - Developing and Testing The Equations," Tech. rep., University of Toronto, Institute for Aerospace Studies, 1985. UTIAS 296.

[26] Gouverneur, B., van Paassen, M. M., Mulder, J. A., Stroosma, O., and Field, E. J., "Optimisation of the SIMONA Research Simulator's Motion Filter Settings for Handling Qualities Experiments," *Proceedings of the AIAA Modeling and Simulation Technologies Conference and Exhibit*, AIAA, Austin, Texas, 2003. https://doi.org/10.2514/6.2003-5679.

[27] Sinacori, J. B., "The determination of some requirements for a helicopter flight research simulation facility," Tech. rep., NASA, 1977. NASA CR 152066.

[28] Schroeder, J. A., "Helicopter Flight Simulation Motion Platform Requirements," Tech. rep., NASA, 1999. NASA/TP-1999-208766.

[29] Kingdom, F., and Prins, N., *Psychophysics: A Practical Introduction*, Academic Press, 2016.

[30] Jerald, J. J., "Scene-Motion-and Latency-Perception Thresholds for Head-Mounted Displays," Ph.D. thesis, University of North Carolina, Chapel Hill, 2010.

[31] Beckers, N. W. M., Pool, D. M., Valente Pais, A. R., van Paassen, M. M., and Mulder, M., "Perception and Behavioral Phase Coherence Zones in Passive and Active Control Tasks in Yaw," *Proceedings of the AIAA Modeling and Simulation Technologies Conference 2012, Minneapolis (MN)*, 2012. https://doi.org/10.2514/6.2012-4794.