The use of pilot ratings in rotorcraft flight simulation fidelity assessment

MiletoviC, I.; Pool, D. M.; Stroosma, O.; Pavel, M. D.; Wentink, M.; Mulder, M.

Publication date
2017

Document Version
Accepted author manuscript

Published in
73rd Annual AHS International Forum and Technology Display

The Use of Pilot Ratings in Rotorcraft Flight Simulation Fidelity Assessment

I. Miletović
I.Miletovic@tudelft.nl
PhD Student
Delft University of Technology
Delft, The Netherlands

D. M. Pool
D.M.Pool@tudelft.nl
Assistant Professor
Delft University of Technology
Delft, The Netherlands

O. Stroosma
O.Stroosma@tudelft.nl
Research Engineer
Delft University of Technology
Delft, The Netherlands

M. D. Pavel
M.D.Pavel@tudelft.nl
Assistant Professor
Delft University of Technology
Delft, The Netherlands

M. Wentink
mark.wentink@desdemona.eu
Technical Director
Desdemona B.V.
Soesterberg, The Netherlands

M. Mulder
M.Mulder@tudelft.nl
Full Professor
Delft University of Technology
Delft, The Netherlands

ABSTRACT

The fidelity of a rotorcraft flight simulator is influenced by many factors, such as the vehicle dynamic model and the motion cueing algorithm (MCA). To quantify the fidelity of a simulator objectively requires detailed knowledge of human pilot perception and control behaviour that is not yet available. As a consequence, subjective assessments made by qualified pilots remain the most important way to assess flight simulation fidelity. The use of standardized rating scales during such assessments can increase the level of objectivity above that provided by less structured evaluations. The current paper describes the result of an experiment performed on the Desdemona simulator to evaluate two rating scales, namely the Simulator Fidelity Rating (SFR) scale and the Motion Fidelity Rating (MFR) scale, as suitable indicators of flight simulation fidelity. In this experiment, two characteristics of the simulated environment were varied, namely rotorcraft dynamics and MCA configuration, and the type of rating scale used was treated as an additional independent variable. The primary results of the experiments suggest that pilots are able to recognize a strong decline in flight simulation fidelity when both rotorcraft dynamics and motion are degraded simultaneously. However, when either one of these characteristics are varied independently of the other, the results are inconclusive. The paper presents a more detailed review of the various results gathered during the experiment and formulates recommendations for future experiments in rotorcraft flight simulation fidelity assessment that involve the use of pilot ratings.

INTRODUCTION

Over the last decade, a strong incentive to develop objective metrics for the assessment of flight simulation fidelity has emerged, in particular for simulator motion (Refs. 1, 2). At the same time, however, it is recognized that the current knowledge on human self-motion perception and manual control is too limited to warrant a full understanding of the effect of simulated environment characteristics on perceived fidelity (Ref. 3). Here, perceived fidelity refers to the degree to which a pilot's perception and action in the simulated environment matches actual flight (Ref. 4). Subjective assessments made by qualified pilots therefore remain the most important measures of flight simulation fidelity demanded by regulatory bodies (Refs. 5, 6). The fundamental problem with subjective assessment of simulation fidelity, however, is the difficulty of obtaining reproducible results and, consequently, the lack of accepted standards. Differences between pilots, such as experience, background and natural variations in human perception and performance, significantly complicate the generalization of results required for the formulation of accepted standards.

Many subsystems in a modern flight simulator interact to produce a realistic flight environment to pilots. Often these interactions, and especially their effect on perceived fidelity, are poorly understood. Perhaps the most notorious example is the motion system of a full-flight simulator. In order to constrain the motion of simulators within their available workspace, the vehicle motion as computed by the flight model is processed using so-called Motion Drive Algorithms (MDA's) (Ref. 7), also commonly referred to as Motion Cueing Algorithms (MCA's). Even though these MCA's insuperably affect the motion cues perceived by pilots, objective standards regarding acceptable levels of mismatch are yet to be defined. As a result, experienced evaluation pilots are usually involved in both the configuration of MCA's and, consequently, the acceptance of the flight simulator as a whole. While there have been studies into objective motion tuning and evaluation methodologies, e.g., (Refs. 8, 9), the general consensus seems to be that subjectively tuned motion is often preferred by pilots.

Submitted for the AHS 73d Annual Forum, Fort Worth, Texas, USA, May 9–11, 2017.

In order to leverage the potential of such subjective tuning and evaluation methods, many rating scales have been proposed in the literature over the years. A prominent example is the well-known Cooper-Harper handling qualities rating (CHR) scale (Ref. 10). Although originally developed for the evaluation of aircraft handling qualities, it is and has been used for the assessment of flight simulation fidelity, e.g., (Refs. 11, 12). Other scales, developed specifically for the evaluation of simulation fidelity, have also been proposed. Early examples include the Motion Fidelity Rating (MFR) scales used in (Refs. 11, 13) to study simulator motion fidelity. More recently, the Simulator Fidelity Rating (SFR) scale (Ref. 14) and an alternative MFR scale (Ref. 15) have also been proposed, both bearing many similarities to the CHR scale. These rating scales have been used extensively to study the perceived fidelity of simulated flight environments (Refs. 8, 15–18).

An issue that has received less attention, however, is the ability of pilots to distinguish between the effects of changes in different simulator characteristics when assessing the fidelity of the simulated environment. For example, in addition to subjective tuning of the MCA, artificial tuning of the rotorcraft simulation model is often also required to match its response to flight test data as demanded by regulations (Ref. 19). It was shown that this process can produce handling qualities of the flight model that differ significantly from those of the reference aircraft (Ref. 19). With both the mathematical vehicle model and the MCA subject to tuning and having an influence on perceived fidelity, it is of interest to study the relationship between these two subsystems. Their interaction can lead to deficiencies present in either to influence the perceived fidelity of the other (Ref. 8). It is therefore important to investigate the sensitivity and effectiveness of rating scales in the presence of such complex interactions. This paper presents the results of an experiment performed in the Desdemona simulator to assess the SFR and MFR scales as suitable indicators of rotorcraft flight simulation fidelity in a lateral reposition task.

BACKGROUND

Drawing from experience in aircraft handling qualities, the application of Handling Qualities Ratings (HQRs) for simulation fidelity assessment are actively applied and researched. For example, (Ref. 20) first applied the well-known Cooper-Harper scale to assess the effects of simulator motion system properties and time delays on perceived rotorcraft flying qualities. The Cooper-Harper scale was also applied more recently in (Refs. 12, 17, 18) as part of studies on flight simulator optimization for the UH-60A Black Hawk helicopter. However, it is important to emphasize that the Cooper-Harper HQR scale was not originally intended to be applied for such evaluations of flight simulators. It is recognized that assessing the fidelity of flight simulators is fundamentally different from assessing the performance and Handling Qualities (HQs) of actual aircraft. This is because the purpose of flight simulators is to reproduce, as accurately as required and physically possible, the flying characteristics of the modelled aircraft, *including* any specific deficiencies in HQs. That is, while the Cooper-Harper HQR scale provides an *absolute* measure of aircraft HQs, a measure *relative* to the baseline aircraft is actually desired. Simply matching Cooper-Harper HQRs between actual and simulated flight therefore cannot guarantee that simulation fidelity is satisfactory. After all, two different aircraft with similar HQs may still possess distinct dynamic characteristics and may therefore require vastly different control strategies from pilots (Ref. 14).

To address these issues, the Simulator Fidelity Rating (SFR) scale (Refs. 14, 21) was proposed. The SFR scale aims to capture the relative difference in adopted control strategy between actual and simulated flight to achieve a prescribed level of desired or adequate task performance (see Fig. 2). The SFR scale was applied independently by (Refs. 8, 18) and although only based on evaluations collected from experiments performed by a relatively small number of test pilots, the scale appeared consistent. Though it can be argued that the SFR scale is an improvement over the Cooper-Harper scale for flight simulator fidelity assessment, it does lack its well-established level of maturity as well as its level of familiarity and acceptance within the aeronautical community. Also, it lacks a specification of accompanying objective task performance criteria, but instead relies on attained task performance *relative* to the baseline aircraft. Moreover, it remains questionable to what extent pilots are able to recognize and quantify required adaptions in their control strategy with respect to an established baseline vehicle. Recommended means to enhance the applicability of the SFR scale in capturing simulator fidelity include minimizing the timespan between experiment trials in the actual aircraft and in the simulator as well as using highly experienced pilots for the specific type of aircraft simulated (Ref. 14).

A rating scale that is closely related to the SFR scale is the Motion Fidelity Rating (MFR) scale, shown in Fig. 3. As such, it has similar characteristics. Like the SFR scale, the MFR scale asks pilots to quantify fidelity, in this case *specifically* motion fidelity, with respect to the actual aircraft. Pilots are furthermore expected to express to what extent the available motion cues contribute towards enhanced task performance and can highlight specific deficiencies using the letter abbreviations as shown in the lower part of Fig. 3. The MFR scale was successfully applied in conjunction with the Cooper-Harper HQR scale to quantify pilot opinion in recent studies (Refs. 15, 22). In these studies, the performance of two different Motion Cueing Algorithms (MCAs) for application to a small motion-base flight simulator in a low-speed helicopter flying task were investigated. Two test pilots provided MFR scale ratings that appeared to be in good agreement.

In addition, another important question of interest is the ability of regular operational pilots to differentiate between the influence on perceived fidelity of core characteristics of the simulated environment on the basis of these two scales. After all, while the SFR and MFR scales were developed primarily for application with expert pilots, a sufficiently large sample size of such pilots may not always be available for simulation fidelity assessment studies. Also, while recent studies have shown that various crucial simulator characteristics have significant effects on task performance, control activity and perceived fidelity in terms of awarded fidelity ratings (e.g, (Refs. 8, 16, 22, 23)), a poorly addressed issue is the ability of pilots to distinguish between the effects of each of these subsystems. For example, (Ref. 8) provides some anecdotal evidence that pilots may confuse deficiencies in motion cues with degraded rotorcraft HQs. Fig. 1 depicts this problem from the point of the pilot. In this figure, the pilot is assumed to perform an arbitrary manual control task in a simulator. It becomes evident that the *equivalent system* perceived by the pilot is the aggregate of the helicopter model *and* the MCA. After all, motion cues from the helicopter model cannot be directly represented by the simulator due to physical constraints. Consequently, pilots must rely on the *actual motion feedback* provided by the simulator.

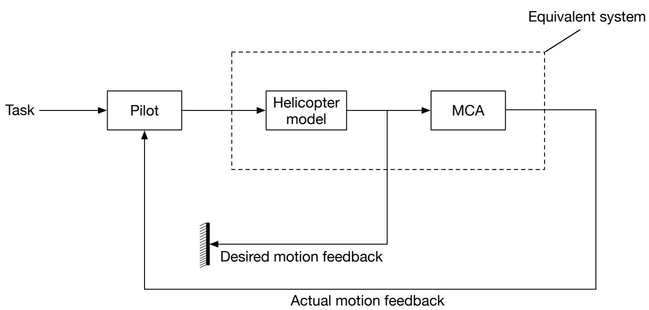

Fig. 1. Schematic of a manual control task in a simulator.

In subjective tuning and assessment of simulator motion, situations could therefore arise where perceived issues in either of the two subsystems (i.e., rotorcraft dynamics or the MCA) is addressed in the other and vice-versa. This, in turn, could cause actual deficiencies in one subsystem to be masked, or worse, aggravated by the changes inadvertently applied in another subsystem.

Therefore, the primary purpose of the experiment proposed in the current work is the assessment of two subjective metrics, namely the SFR and MFR rating scales as suitable indicators of simulation fidelity. To this end, two core characteristics of the simulated environment are varied in the proposed experiment, namely the *rotorcraft model* and *motion*.

EXPERIMENT DESIGN

To gain more insight in the use of rating scales to assess flight simulation fidelity assessment in the presence of varying combinations of rotorcraft dynamics and motion, an experiment with a design as outlined in Tab. 1 was conducted. In the experiment, operational helicopter pilots were asked to perform a helicopter flight task in a simulated flight environment with physical motion. The task, the simulator platform and the pilot population were kept constant throughout the experiment.

Three different experimental variables were varied in the experiment: the rotorcraft dynamics, the MCA configuration and, finally, the rating scale used. This design differs from that of previous experiments, where the *type* of rating scale(s) used is most commonly kept constant. In the proposed experiment, the rating scale used is varied in a within-subject fashion. In the following, the experiment design details are further elaborated upon.

Table 1. Independent, dependent and controlled experimental variables.

Independent variables	*Dependent measures*	*Controlled variables*
Rotorcraft dynamics	Pilot rating	Simulator platform
MCA configuration	Task performance	Flight task
Rating scale		Pilot population

Apparatus

The experiment was conducted on the Desorientation Demonstrator Amst (Desdemona) simulator (see Fig. 4) located in Soesterberg, the Netherlands (Ref. 25). The simulator comprises a cabin that is suspended in a three degree-of-freedom (DoF) gimbal with a continuous range of motion around all axes. The gimballed cabin, in turn, is mounted on a short vertical track to provide a translational degree of freedom, 1.5 meters in length. A second translational degree of freedom is provided by a longer horizontal track, 8 meters in length. Finally the horizontal track is connected to a large central pivot in order to provide an additional centrifugal DoF with a continuous range of motion. This yields a total of six DoF of motion. This innovative configuration has a considerably larger workspace than conventional Stewart platform motion bases and can reproduce sustained and large magnitude accelerations of up to three times the gravitational acceleration in centrifuge mode.

Pilot population

The four pilots that participated in the experiment were operational AH–64 Apache pilots from the Royal Netherlands Airforce (RNLAF) and the US Army who had received prior training in the Desdemona simulator. None of the participating pilots had any prior experience with flight testing or the use of rating scales in an experimental setting. One of the participating pilots is an AH–64 flight instructor in the RNLAF.

Flight task

The flight task considered in this experiment is a lateral reposition maneuver over a distance of 400 ft, with corresponding task performance specifications as stipulated in ADS–33E (Ref. 26):

> ”Start in a stabilized hover at 35 ft wheel height (or no greater than 35 ft external load height) with the longitudinal axis of the rotorcraft oriented 90 degrees to a reference line marked on the ground. Initiate a lateral acceleration to approximately 35 knots groundspeed followed by a deceleration to laterally reposition the rotorcraft in a stabilized hover

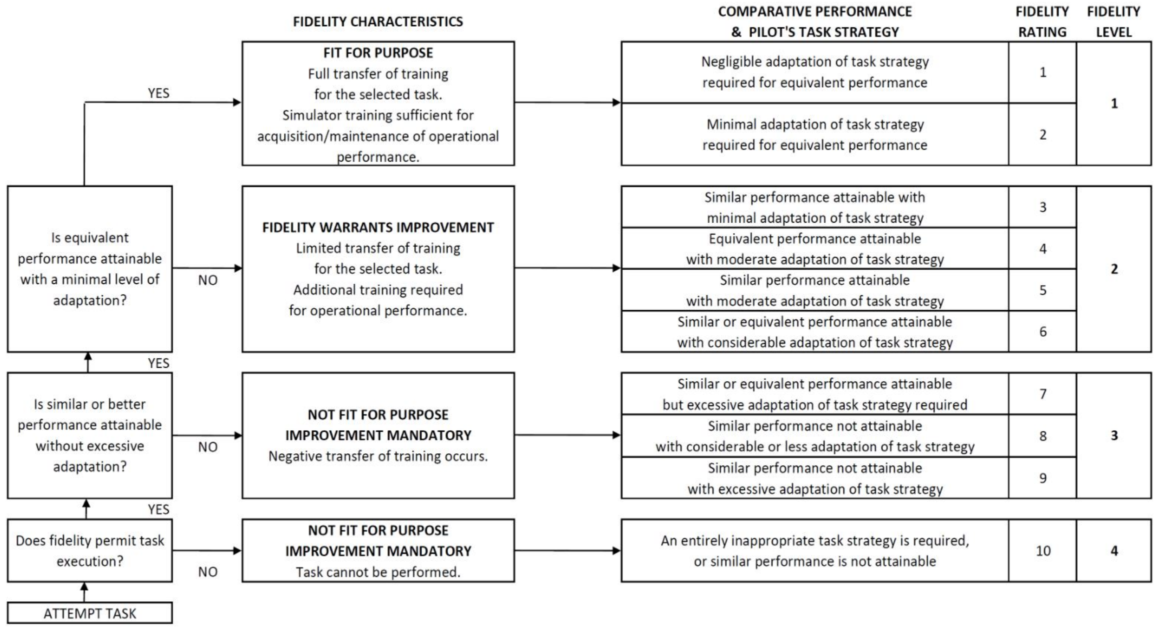

Fig. 2. The Simulator Fidelity Rating (SFR) scale (Ref. 14).

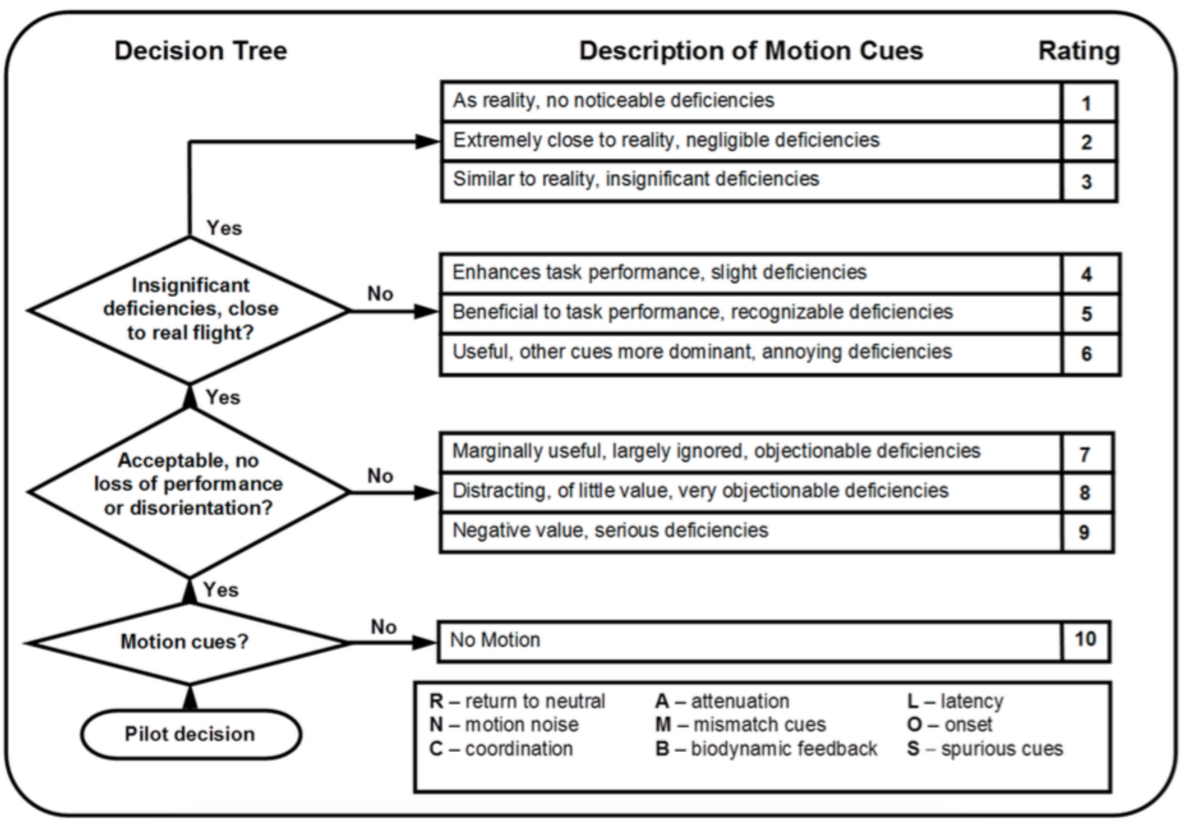

Fig. 3. The Motion Fidelity Rating (MFR) scale (Ref. 24).

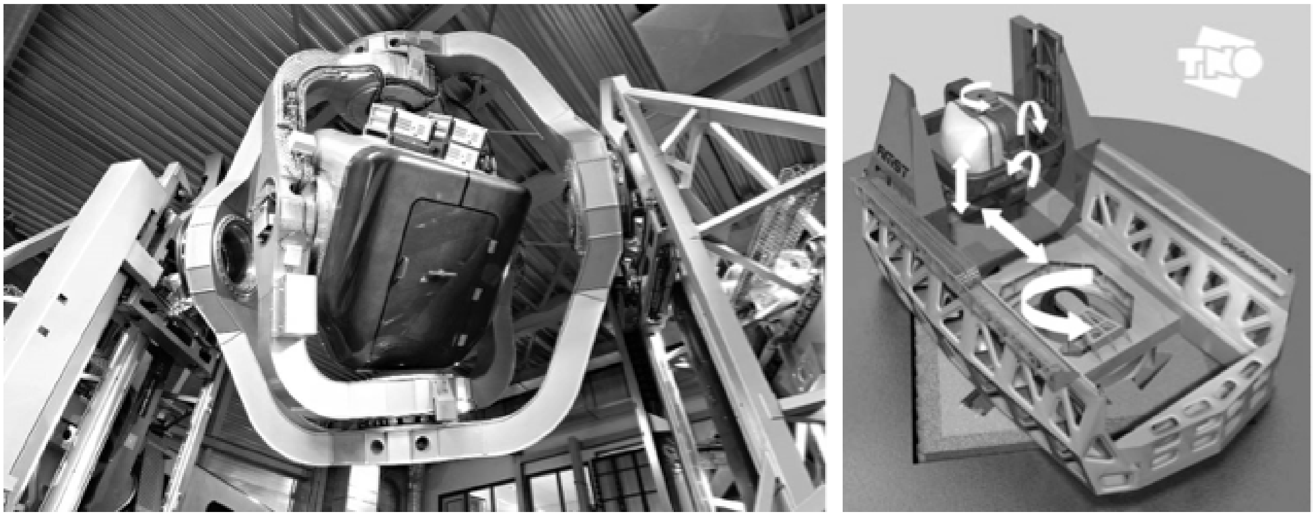

Fig. 4. The Desdemona simulator at TNO/Desdemona B.V. in Soesterberg, The Netherlands (Ref. 25).

> 400 ft down the course within a specified time [*18s desired, 22s adequate*]. The acceleration and deceleration phases shall be accomplished as single smooth maneuvers. The rotorcraft must be brought to within 10 ft of the endpoint during the deceleration, terminating in a stable hover within this band. Overshooting is permitted during the deceleration, but will show up as a time penalty when the pilot moves back within 10 ft of the endpoint. The maneuver is complete when a stabilized hover is achieved."

The selected task is also similar to those used in recent studies (Refs. 15, 16, 23). In the current work, however, the lateral reposition was preferred over the more aggressive sidestep maneuver. This is because it was expected that the participating pilots would have more exposure to basic maneuvers performed with moderate levels of agility. Moreover, it was anticipated that a maneuver with a less demanding level of agility would be more suitable in combination with the *linear* rotorcraft model used. Fig. 5 shows an impression of the task setup in the visual environment of Desdemona.

Rotorcraft dynamics

In the current experiment, a simplified lateral position control task is performed. For the controlled dynamics, the altitude, pitch and heading were fixed, so the only DoFs controlled by pilots were roll and, consequently, sway. The roll-sway dynamics of rotorcraft in this case can be approximated by a linear function of roll damping and control power (Refs. 23, 27):

$$\begin{aligned} \ddot{\phi} &= L_p\, p + L_\delta\, \delta \\ \ddot{y} &= g\,\phi + Y_v\, v \end{aligned} \tag{1}$$

The roll damping L_p is a measure of the tendency of the rotorcraft to counteract a rolling moment, while the control power L_δ determines the instantaneous angular acceleration for a given lateral cyclic input. The smaller L_p and the higher L_δ, the more agile the rotorcraft's roll response for a given control input. Consequently, these parameters fully characterize the roll dynamics of the rotorcraft and their direct manipulation can therefore be used to artificially degrade HQs. A similar model was also used in recent experiments, e.g., (Refs. 15, 16, 23). In contrast to these previous experiments, however, the effect of lateral drag (through Y_v) is included here. This was done after preliminary evaluations with pilots, in which it became apparent that the task could otherwise not be completed with adequate performance.

The primary reason for selecting this simple model is that its fidelity is considered sufficient for the purposes of the current experiment, where only *relative* changes in the fidelity of the overall simulated flight environment are of interest. The *absolute* fidelity of the baseline model is of lesser importance in this case, as long as it is representative for both the task and the pilot population of interest. The stability derivatives L_p and Y_v that appear in Equation 1 were chosen such to be representative of the AH–64 and originate from flight test data as documented in (Ref. 28). The values for these parameters are -1.828 s^{-1} and -0.2788 s^{-1}, respectively. The control power L_δ was also determined from preliminary pilot evaluations and was assigned a value of 1.5 s^{-2}. Rotorcraft dynamics were varied by diminishing the roll subsidence L_p, thereby enabling greater agility in roll at the cost of degraded handling qualities. Two conditions were selected, one with the original value for L_p (-1.828 s^{-1}) and one in which the magnitude of L_p was reduced by half to a value of –0.914 s^{-1}.

MCA configuration

The motion system kinematics of the Desdemona simulator allow for various potential motion cueing strategies in support of the lateral reposition maneuvre considered in the current experiment (Ref. 29). However, it was chosen to simplify

Fig. 5. Overview of lateral reposition task in Desdemona.

the motion cueing strategy applied significantly by only presenting motion in roll. The reason for this is twofold. First, for the purpose of the current experiment, it is sufficient to offer two conditions of motion that, *objectively*, are significantly different in terms of their fidelity level. By presenting a motion condition with relatively good fidelity roll motion and one in which the roll motion fidelity is significantly degraded, this premise is already satisfied. Even though the simulator allows for both motion conditions to be extended with sway, this would introduce complexities that are deemed unnecessary for the objective of the current experiment. Moreover, recent research has demonstrated a strong tendency for roll to dominate perceived fidelity in lateral helicopter maneuvering (Refs. 16, 22). This is true especially for large phase distortions induced in the roll channel.

For the current experiment, two motion conditions similar to those evaluated in (Ref. 16) will be considered. The roll angle, for both cases, will therefore be filtered using a washout filter of second-order:

$$H_{\phi}(s) = K_{\phi} \frac{s^2}{s^2 + 2\zeta\omega_{\phi} + \omega_{\phi}^2} \quad (2)$$

The two MCA configurations used differ mainly in the amount of phase distortion incurred by the washout filter. The break-frequency ω_{ϕ} of the baseline configuration was chosen such to incur a phase lead of 30 degrees ($\omega_{\phi} = 0.36$ rad/s) at 1 rad/s, while the break-frequency for the second configuration was increased such to incur 60 degrees ($\omega_{\phi} = 0.67$ rad/s) of phase lead. This is similar to the conditions investigated in (Ref. 16). The scaling gain K_{ϕ} of both filters was chosen sub-unity, with a value of 0.6 for the baseline configuration and a value of 0.9 for the degraded configuration. The higher gain for the degraded configuration was chosen such to compensate for the larger attenuation at lower frequencies.

Finally, note that the degraded configuration, with a stronger washout effect (i.e., larger phase lead at 1 rad/s), also has an advantage over the baseline configuration used. Namely, the gravity-induced false specific force cues, due to cabin roll in the absence of lateral cabin motion (i.e., sway), are *lower* in the degraded configuration. However, as stated in the previous paragraphs, it is assumed that the degree of phase distortion in roll dominates the perceived motion fidelity.

Dependent measures

As mentioned previously, the type of rating scale used is treated as an independent experimental variable. This means that any variation in the type of rating scale used constitutes a separate experimental condition. The foreseen advantage of treating the rating scale as such is that this prevents pilots from consciously, or subconsciously, striving for consistency in awarded ratings from the two scales for each presented combination of rotorcraft dynamics and MCA configuration. In this experiment, two rating scales were used: the SFR scale (see Fig. 2) and the MFR scale (see Fig. 3). The ratings awarded by the participating pilots using these scales constitute the primary dependent measure collected during the experiment. Given that task performance, here in the form of maneuver execution time, is also a determining factor in both of the rating scales used, measured task performance was also recorded as a dependent measure.

Execution

With two configurations of rotorcraft dynamics, two MCA configurations and two rating scales, the total number of experimental conditions equals eight. The eight conditions were distributed over the four pilots according to a latin square design (see Tab. 2). In the table, the conditions are designated by a letter ('b' for baseline or 'd' for degraded) signifying the rotorcraft dynamics, followed by a number (30 or 60) signifying the motion filter used and, finally, another letter ('s' for SFR or 'm' for MFR) signifying the rating scale used.

Table 2. Division of experimental conditions over the four subjects.

Pilot	Conditions							
1	b30s	d60m	d30s	d30m	b30m	d60s	b60s	b60m
2	b60m	b60s	d60m	b30s	d30s	b30m	d30m	d60s
3	d60s	d30m	b60s	b60m	d60m	d30s	b30s	b30m
4	d30m	d30s	b30m	b60s	d60s	b60m	d60m	b30s

Prior to the start of the experiment, the pilots were briefed and were given the opportunity to familiarize themselves with the rotorcraft dynamics and task. In the briefing, pilots were introduced to the two rating scales used in the experiment as follows:

> "For the SFR scale, the possible ratings pertain to the attained task performance and the control strategy that was applied. The less these differ from the attainable performance and control strategy in the actual aircraft, the better (that is, *lower*) should be the awarded rating. For the MFR scale the same applies, however, here the possible ratings pertain to the extent to which motion has contributed to attaining the specified performance requirements. The MFR scale also includes indicators that allow you to communicate several specific motion cueing deficiencies you may have perceived. You are encouraged to make use of these indicators, although it is not explicitly required from you in the experiment."

In addition to the briefing, the pilots were also given the opportunity to familiarize themselves with the specific terminology used in the rating scales prior to the start of the experiment. Other than these minor guidelines, pilots were granted the freedom to interpret the rating scales in any way they deemed appropriate. With regard to task execution, pilots were instructed to perform the maneuver in one fluent motion and to strive for optimal task performance in every trial.

During familiarization, pilots were allowed to perform trial runs of the task only with the *baseline* configuration. This was done so as to minimize the risk that pilots would learn about the specific changes to be introduced in the simulated environment prior to the actual experiment (e.g., by means of "probing"). This could inadvertently influence the awarded pilot ratings. The trial runs were repeated until the subject attained at least adequate task performance, for at least two or three times, and the subject felt confident that adequate performance could be maintained.

In the actual experiment, pilots were given the opportunity to repeat the maneuver three times for each experimental condition before being asked to award a rating according to the specified rating scale. This was done so as to limit within-subject variability. The pilots were told which rating scale to use *after* completion of the task. In case of doubt, inability, or other extraordinary discrepancies noticed during the first three runs, pilots were allowed to repeat the maneuver more often. On average, pilots required four repetitions of the task before a rating could be formulated. Task performance (i.e., maneuver execution time) was measured manually using a stopwatch. Any further comments provided by pilots during the experiment were recorded.

Hypotheses

The current experiment was performed so that a better judgment can be formed regarding the use of subjective fidelity metrics as part of flight simulation fidelity assessment studies. Specifically, the central premise that was tested is the ability of pilots to distinguish degradation of motion fidelity from degradation of rotorcraft HQs. Therefore, it is of interest to establish the extent to which SFR and MFR ratings vary with different combinations of rotorcraft dynamics and the MCA. Ideally, pilots are able to distinguish between the influence of both subsystems based on ratings awarded from the SFR and MFR scales. Therefore, the following hypotheses can be formulated:

1. **SFR and MFR ratings will degrade as *both* motion cueing fidelity and rotorcraft HQs are degraded from the baseline condition.** Degraded motion cueing fidelity and rotorcraft HQs are assumed to have a degrading effect on perceived fidelity when compared to the baseline condition. SFR scale ratings are expected to diminish as a result of increased pilot workload, following a degradation of task performance and an expected adaptation of the task strategy. MFR ratings will diminish because the degraded motion cues are expected to no longer contribute to enhanced task performance.

2. **Degraded rotorcraft HQs combined with baseline motion cueing fidelity will *only* result in improved MFR ratings as compared to a condition with *both* degraded rotorcraft HQs and motion cueing fidelity.** Degrading only rotorcraft HQs, while not degrading motion cueing fidelity, is expected to result in a degradation of SFR scale ratings. Degraded rotorcraft HQs are expected to result in a significant control strategy adaptation when compared to the baseline condition. The baseline MCA configuration may still contribute to enhanced task performance. In this case, motion may still help the pilot to generate sufficient lead (Ref. 16). Hence, this hypothesis effectively stipulates that the MFR scale is not sensitive to changes in rotorcraft model fidelity and therefore enables pilots to reliably express problems with motion cueing fidelity.

3. **Degraded motion cueing fidelity combined with baseline rotorcraft dynamics will *only* result in improved SFR ratings as compared to a condition with *both* degraded rotorcraft HQs and motion cueing fidelity.** Similar to the condition with both degraded rotorcraft HQs and motion, degrading only motion cueing fidelity is expected to result in diminished MFR scale ratings. In both conditions, motion is not expected to contribute to enhanced task performance.

In this condition, it is hypothesized that pilots will attempt to actively ignore motion cues and rely mostly on the available visual cues to optimize task performance. This strategy results in a larger effective time delay of the pilot and therefore in a lower task performance and/or adaptation of the task strategy. Consequently, SFR scale ratings are also expected to diminish but, since motion is the dominant deficiency in this condition, not as strongly as for the condition with both degraded motion cueing fidelity *and* degraded rotorcraft HQs.

RESULTS

Maneuver phase portraits

The primary experiment results collected are pilot ratings and measured task performance. However, during the experiment, pilot control inputs and rotorcraft states were recorded as well. Based on the latter, so-called phase portraits (Ref. 16) of the task executed by the pilots were constructed. Phase portraits depict the rotorcraft speed as a function of travelled distance and provide an indication of the amount of variability in applied task strategy between subjects and experimental conditions. Fig. 6 shows such phase portraits for the trial numbers associated with the *best* attained task performances per subject and condition. The trial numbers and the associated task performances are selected as a reference because they are assumed to be the most dominant criterium upon which pilots base the awarded ratings for a given condition. The vertical lines shown in the figure designate the target region.

Several observations can be made from Fig. 6. It can be seen that most subjects adopted notably different task strategies across the experimental conditions presented. Subject 4 seemed to adopt the least aggressive strategy of all participants, while the phase portraits of this subject also show the least variation over experimental conditions. Conversely, the other subjects appear to adopt a more aggressive strategy and also exhibit more variation in the resulting phase portraits. More variability in the task trajectories between experimental conditions and subjects also means that excitation levels of the rotorcraft dynamics and, consequently, motion will strongly vary between subjects and conditions. These differences can have a strong effect on awarded pilot ratings. In the following sections, the experimental results in terms of pilot ratings and task performance are presented. In addition, a detailed review of the most important pilot comments will be given.

Ratings and task performance

The awarded pilot ratings, together with the corresponding task performances attained, are shown in Fig. 7 and Fig. 8 for the SFR and MFR scales, respectively. Note that, as explained previously, only *best* attained task performances are considered. Both individual results for each subject as well as the overall trends (using the *median* ratings) and spread calculated from the results corresponding to the first three subjects are shown. The results corresponding to Subject 4 were not taken into account in the overall trends and spread of the results because of the strongly differing task strategy applied by this subject. Several interesting remarks about the results can be made.

First, it can be seen that the overall SFR and MFR ratings awarded for the condition with baseline dynamics and with a motion filter that incurred the least amount of phase distortion (b30) are very different. The same is true for the condition with *only* degraded rotorcraft dynamics (d30). Also, while subjects evidently disliked the condition with *both* degraded dynamics and motion most (d60), degrading *either* dynamics (b→d) or motion (30→60) produced less consistent results. Subjects favoured the condition with only degraded dynamics according to the SFR scale, while according to the MFR scale, the baseline condition (b30) is preferred. This shows the apparent difficulty in separating the influence of rotorcraft dynamics and MCA configuration on perceived motion fidelity.

Another interesting observation can be made regarding the relation between awarded ratings and adopted task strategy. Fig. 6 showed that Subject 4 adopted a relatively low-gain task strategy as compared to the other subjects. At the same time, however, the ratings awarded by this subject are in general more favourable for each experimental condition than those awarded by the other three subjects. This observation signifies that variability in the adopted task strategy, in terms of the excitation level of rotorcraft dynamics and motion, may indeed lead to large differences in perceived simulation fidelity.

In terms of the stipulated hypotheses, the following concluding remarks about the overall results can be made:

1. For the case of the SFR scale ratings, the baseline condition (b30s) was the second-worst rated, where degrading either motion (b60s) or rotorcraft dynamics (d30s) resulted in a better rating. Only both degraded motion and degraded handling qualities (d60s) result in a worst rating when compared to baseline. For the MFR scale ratings, the baseline was the best rated condition. Degrading either motion (b60m) or handling qualities (d30m) resulted in worst ratings, while degrading both motion and handling qualities (d60m) resulted in the worst rating. As such, the first hypothesis can be confirmed, in that both SFR and MFR ratings degrade when both motion and handling qualities are degraded.

2. The results indicate that the condition with only degraded rotorcraft dynamics received both improved SFR and MFR ratings, the improvement in the SFR ratings appearing stronger than the improvement in MFR ratings. Thus, subjects actually *preferred* the condition with degraded rotorcraft dynamics and therefore the second hypothesis must be rejected.

3. As compared to a condition with both degraded motion fidelity and rotorcraft dynamics, a slight improvement in both SFR and MFR ratings was observed for the condition with only degraded motion. In addition, it can be

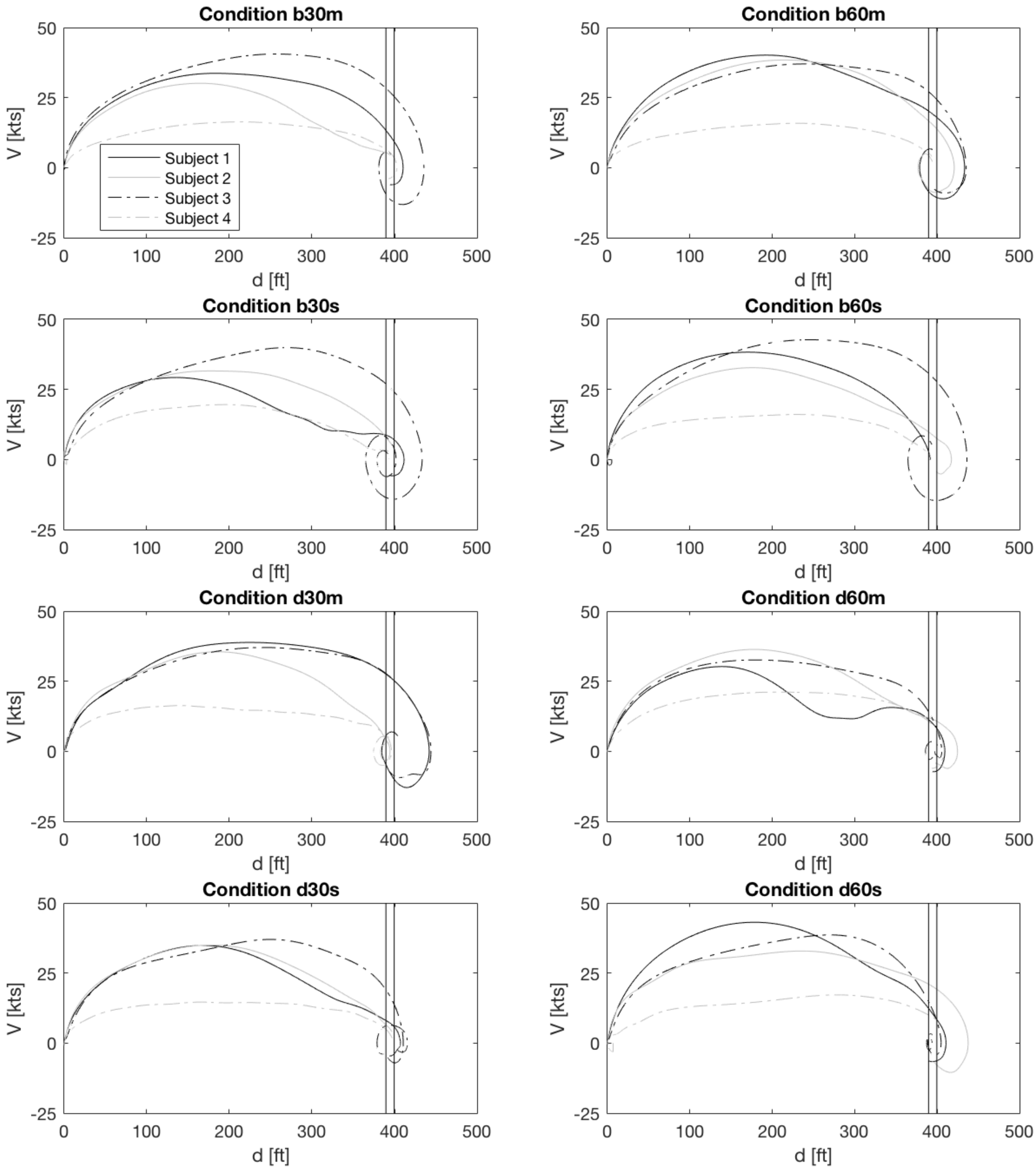

Fig. 6. Phase portraits for each experimental condition and subject.

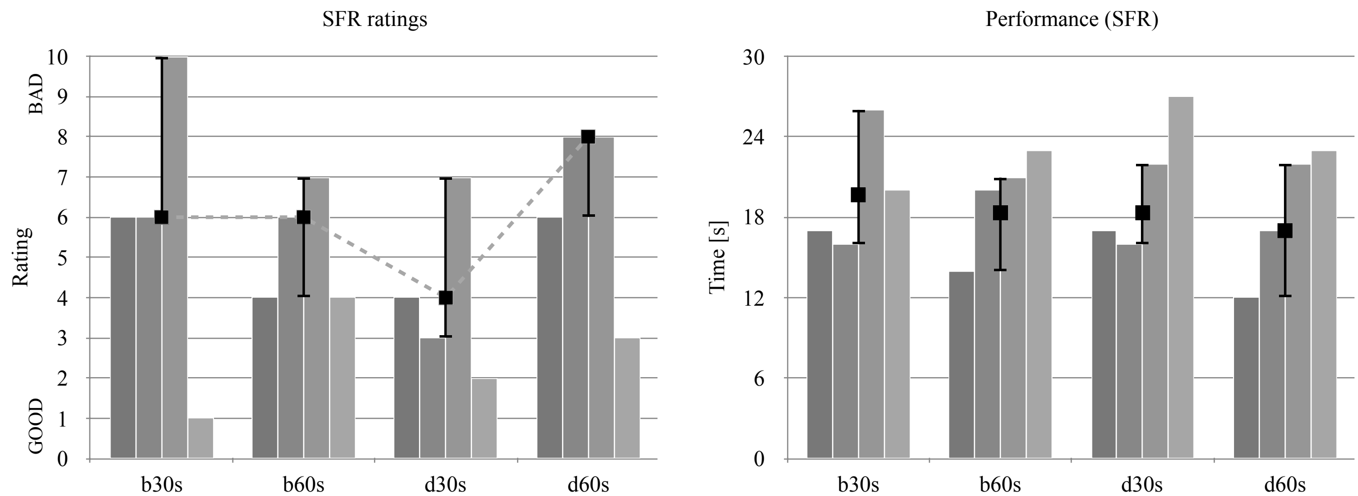

Fig. 7. SFR ratings (left) and corresponding task performances (right).

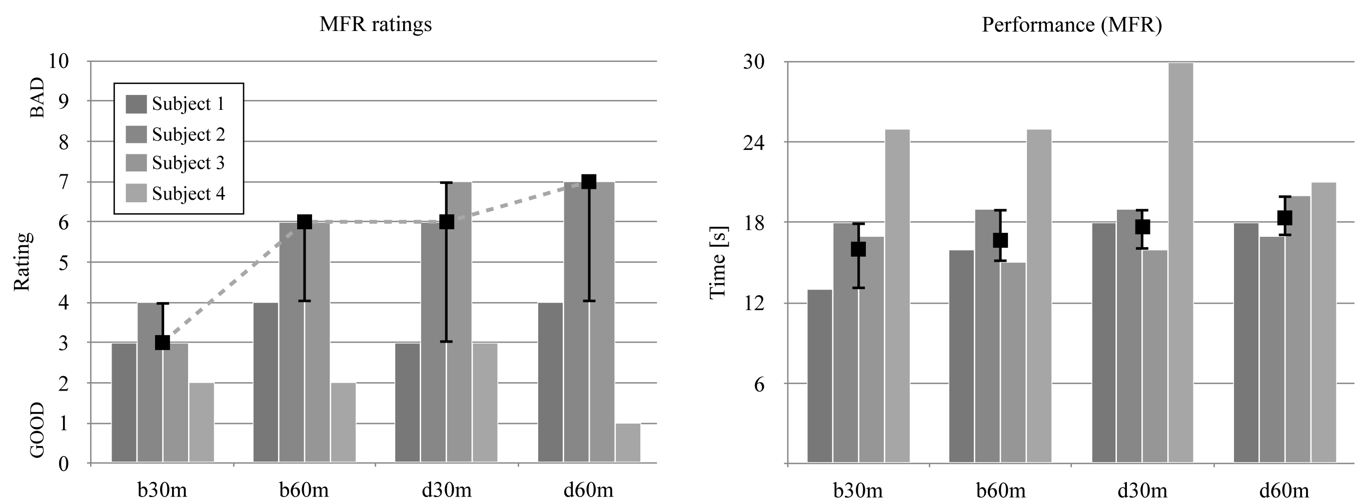

Fig. 8. MFR ratings (left) and corresponding task performances (right).

seen that both SFR and MFR ratings improved by approximately the same degree. A stronger degradation of MFR ratings as compared to SFR ratings could not be observed and therefore the third hypothesis is also rejected.

In summary, it seems pilots are indeed sensitive to changes in rotorcraft dynamics and changes in the configuration of the MCA. However, attempts to separate the contributions of rotorcraft HQs and MCA configuration on perceived fidelity, using the SFR and MFR scales, proved unsuccessful.

Pilot comments

During the experiment, the pilots were encouraged to support their awarded ratings with comments. An extensive report of the individual pilot comments can be found in the appendix to this paper. Here, the main findings are summarized.

The overall trends observed from the comments are that subjects seem to adequately perceive and communicate the deficiencies pertaining to rotorcraft dynamics (e.g., lateral damping, high inertia, sluggish, etc.) and motion (e.g., lag, latency, overshoot, etc.). At the same time, however, it is also observed that the subjects sometimes pointed out deficiencies in motion, while in fact only rotorcraft dynamics were modified. This is in agreement with the trends found in the awarded SFR and MFR ratings. Nonetheless, some pilot comments also seem to contradict the awarded ratings at times. For example, one subject noted that no change in task strategy was perceived when transitioning between two conditions, but awarded different SFR ratings for the two conditions nonetheless. Similarly, another subject assigned ratings to conditions that suggest no particular deficiencies were perceived, while the subject's comments suggest otherwise.

DISCUSSION

The results presented in preceding sections give rise to several interesting observations. From inspection of phase portraits (see Fig. 6), it seems that the subjects' task strategy varies strongly across the different experimental conditions presented. In particular, the aggressiveness and agility with which the task is completed is a measure of the observed variety in the corresponding maneuver phase portraits. The more mild the task strategy of the subject, the lesser the observed variety in the phase portraits.

Based on the awarded pilot ratings, it appears that the subjects are able to discern strong degradations in the simulated environment. In particular, conditions with *both* degraded rotorcraft dynamics *and* motion were rated worst overall with both rating scales used, as hypothesized. On the other hand, conditions in which only one of the two simulation characteristics was degraded produced inconclusive results. It seems pilots preferred the condition with only degraded rotorcraft dynamics in terms of the SFR scale, while preferring the baseline condition in terms of the MFR scale. The former result, however, can be explained based on pilot comments. These indicate that pilots generally perceived the lateral drag of the rotorcraft to be too strong. The more agile characteristics of the rotorcraft in the condition with only degraded dynamics is thought to somewhat alleviate the sluggish characteristics of the rotorcraft in the baseline condition. The fact that the condition with degraded dynamics is rated worse than baseline in terms of the MFR scale shows the difficulty in separating the influence of rotorcraft dynamics and MCA settings when assessing motion fidelity. An explanation for this difficulty can be sought in the apparently ambiguous interpretation of what constitutes "good" motion cues. On the one hand, motion cues help the pilot to recognize deficiencies in rotorcraft HQs, thereby serving as a "window" to actual flight. Therefore, poor motion cues could also mask deficiencies in rotorcraft HQs. On the other hand, even poor motion cues in the presence of deficiencies in rotorcraft HQs could provide valuable feedback to the pilot and may therefore be perceived as beneficial to task performance.

The collected pilot comments also contribute to a better understanding of the subjects' perception of the deficiencies introduced in the simulated environment. In general, the participating pilots were able to provide valuable feedback regarding the specific deficiencies introduced in the simulated environment. However, these deficiencies in some conditions were attributed to the wrong subsystem (e.g., simulator motion instead of rotorcraft dynamic model). Also, several conditions with inconsistencies between comments and ratings were identified. In the current experiment, pilots were deliberately unchallenged in formulating their ratings and comments regarding the fidelity of the simulated environment. In future experiments, one could consider challenging pilot ratings and comments in case inconsistencies are observed, thereby allowing a potentially more balanced opinion to be formulated. This should be undertaken with caution, however, in such a way as to not introduce observer bias in the results.

In the awarded pilot ratings, a strong between-subjects variation was also observed. For example, the baseline condition was both the best and the worst appreciated condition when inspecting the ratings on a per-subject basis. Also, one subject who adopted a relatively mild task strategy awarded significantly more favourable ratings than the remaining subjects who adopted a more aggressive task strategy. This shows that possibly a relation between the rotorcraft dynamics on the one hand and task performance specifications on the other hand exists. Task performance criteria that force pilots to adopt a more aggressive task strategy, thereby "pushing" the rotorcraft model up to and beyond its region of validity are likely to result in generally poor fidelity ratings. To obtain meaningful results, it is therefore crucial to take this intricate relationship into account in the design of future experiments for simulation fidelity assessment studies. This could mean that equivalent ADS–33E task performance criteria have to be defined for simulator-based experiments, taking into account rotorcraft model and simulator centred limitations (e.g., visual and motion). On the other hand, once such criteria are defined, it is important to enforce them during experiments by motivating subjects to adopt a more aggressive, or less aggressive strategy, based on their attained task performance.

Even though pilots were given the opportunity to repeat the maneuver multiple times in each experimental condition, no repeated trials of the experimental conditions were performed. This means that it was not possible to evaluate within-subject consistency in the awarded ratings per condition. Naturally, it is also of interest to repeat the experiment outlined in the current work with a larger number of pilots and, possibly, for more comprehensive configurations of the rotorcraft dynamics and MCA.

CONCLUSION

Subjective evaluations made by qualified pilots remain the primary means for the certification of flight simulators. The aim of this study is to assess the effectiveness of the subjective metrics available as indicators of flight simulation fidelity and, consequently, the ability of operational pilots to use such metrics to distinguish between changes in the rotorcraft dynamics and motion cueing algorithm. In this study, two subjective metrics, namely the Simulator Fidelity Rating (SFR) and Motion Fidelity Rating (MFR) scales, were evaluated.

The obtained results have exposed several interesting findings. For example, there is some evidence that changes in SFR and MFR ratings are inconsistent between the experimental conditions evaluated. Also, while the participating pilots seemed able to recognize a large degradation in *both* rotorcraft dynamics and motion, degrading either *one* of these characteristics yielded less conclusive results. Pilot comments in support of the awarded ratings suggest that pilots are able to perceive and identify crucial characteristics of deficiencies in the simulated environment. However, the awarded pilot ratings and supporting comments were not always found to be in agreement with one another. A strong relation between the adopted task strategy and awarded ratings was also identified, where a less aggressive task strategy was found to result in more favourable ratings. This relation suggests that task performance criteria should be tailored more towards the limitations inherent in the rotorcraft dynamics model and other simulator properties (e.g., motion). Once defined, these performance criteria should be enforced in future experiments. Finally, pilots themselves could be challenged more strongly in case inconsistencies in ratings and comments are observed.

REFERENCES

[1] Advani, S. K. and Hosman, R. J. A. W., "Revising Civil Simulator Standards – An Opportunity for Technological Pull," Paper AIAA-2006-6248, Proceedings of the AIAA Modeling and Simulation Technologies Conference and Exhibit, Keystone, CO, 2006.

[2] Hosman, R. J. A. W., Advani, S. K., and Takats, J., "Status of the ICAO Objective Motion Cueing Test," Autumn Flight Simulation Conference: Flight Simulation Research - New Frontiers, London, UK, November 2012.

[3] Mulder, M., Zaal, P. M. T., Pool, D. M., Damveld, H. J., and van Paassen, M. M., "A Cybernetic Approach to Assess Simulator Fidelity: Looking back and Looking Forward," Paper AIAA-2013-5225, Proceedings of the AIAA Modeling and Simulation Technologies Conference, Boston (MA), August 2013.

[4] Heffley, R. K., Clement, W. F., Ringland, R. F., F., J. W., Jex, H. R., McRuer, D. T., and Carter, V. E., "Determination of Motion and Visual System Requirements for Flight Training Simulators," Technical Report Technical Report 546, U. S. Army Institute for the Behavioral and Social Sciences, August 1981.

[5] Anonymous, "JAR-FSTD H Helicopter Flight Simulation Training Devices," , 2008.

[6] Anonymous, "CS-FSTD H Helicopter Flight Simulation Training Devices," , 2012.

[7] Reid, L. D. and Nahon, M. A., "Flight Simulation Motion-Base Drive Algorithms. Part 1: Developing and Testing the Equations," Technical Report UTIAS 296, University of Toronto, Institute for Aerospace Studies, December 1985.

[8] Reardon, S. E., Beard, S. D., and Aponso, B. L., "Effects of Motion Filter Parameters on Simulation Fidelity Ratings," Proceedings of the AHS 70th Annual Forum, Montreal, Canada, 2014.

[9] Jones, M., "Optimizing the Fitness of Motion Cueing for Rotorcraft Flight Simulation," Proceedings of the AHS 72nd Annual Forum, West Palm Beach (FL), May 17-19, 2016.

[10] Cooper, G. E. and Harper, R. P., Jr., "The Use of Pilot Rating in the Evaluation of Aircraft Handling Qualities," Technical Report NASA TN D-5153, National Aeronautics and Space Administration, April 1969.

[11] Schroeder, J. A., "Helicopter Flight Simulation Motion Platform Requirements," Technical Report TP-1999-208766, NASA, Moffett Field, California, July 1999.

[12] Reardon, S. E. and Beard, S. D., "Evaluation of Motion Tuning Methods on the Vertical Motion Simulator," Proceedings of the AHS 71st Annual Forum, Virginia Beach (VA), May 2015.

[13] Sinacori, J. B., "The Determination of Some Requirements for a Helicopter Research Simulation Facility," Technical Report NASA-CR-152066, Systems Technology Inc., September 1977.

[14] Perfect, P., Timson, E., White, M. D., Padfield, G. D., Erdos, R., and Gubbels, A. W., "A Rating Scale for the Subjective Assessment of Simulation Fidelity," *The Aeronautical Journal*, 2014.

[15] Hodge, S. J., Perfect, P., Padfield, G. D., and White, M. D., "Optimising the Roll-Sway Motion Cues Available from a Short Stroke Hexapod Motion Platform," *The Aeronautical Journal*, Vol. 119, (1211), January 2015, pp. 23–44.

[16]Wiskemann, C. M., Drop, F. M., Pool, D. M., van Paassen, M. M., Mulder, M., and Bülthoff, H. H., "Subjective and Objective Metrics for the Evaluation of Motion Cueing Fidelity for a Roll-Lateral Reposition Maneuver," AHS 70th Annual Forum, Montreal, May 2014.

[17]Beard, S. D., Reardon, S. E., Tobias, E. L., and Aponso, B. L., "Simulation System Optimization for Rotorcraft Research on the Vertical Motion Simulator," Proceedings of the AIAA Modeling and Simulation Technologies Conference, Minneapolis, Minnesota, Aug. 13-16, 2012.

[18]Beard, S. D., Reardon, S. E., Tobias, E. L., and Aponso, B. L., "Simulation System Fidelity Assessment at the Vertical Motion Simulator," Proceedings of the AHS 69th Annual Forum, Phoenix, Arizona, 2013.

[19]Pavel, M. D., White, M. D., Padfield, G. D., Roth, G., Hamers, M., and Taghizad, A., "Validation of Mathematical Models for Helicopter Flight Simulators Current and Future Challenges," *The Aeronautical Journal*, Vol. 117, (1189), April 2013.

[20]Mitchell, D. G., Hoh, R. H., Atencio, A., Jr., and Key, D. L., "Ground-Based Simulation Evaluation of the Effects of Time Delays and Motion on Rotorcraft Handling Qualities," USAAVSCOM Technical Report 91-A-010, Aeroflightdynamics Directorate, U.S. Army Aviation Systems Command, Ames Research Center, Moffett Field, CA 94035-1000, January 1992.

[21]Timson, E., *Flight Simulation Fidelity for Rotorcraft Design, Certification and Pilot Training*, Ph.D. thesis, University of Liverpool, September 2013.

[22]Hodge, S. J., Manso, S., and White, M. D., "Challenges in Roll-Sway Motion Cueing Fidelity: A View from Academia," Proceedings of the Conference on Challenges in Flight Simulation, June 2015.

[23]Nieuwenhuizen, F. M., Zaal, P. M. T., Teufel, H. J., Mulder, M., and Bülthoff, H. H., "The Effect of Simulator Motion on Pilot Control Behaviour for Agile and Inert Helicopter Dynamics," Proceedings of the 35th European Rotorcraft Forum, Hamburg, Germany, 2009.

[24]Hodge, S. J., Perfect, P. S., Padfield, G. D., and White, M. D., "Optimising the Vestibular Cues Available from a Short Stroke Hexapod Motion Platform," 67th Annual Forum of the American Helicotper Society, Virginia Beach, VA, USA, May 3-5, 2011.

[25]de Graaf, B., Bles, W., and Hosman, R. J. A. W., "Desdemona: Advanced Disorientation Trainer and (Sustained-G) Flight Simulator," The First Swedish-American Workshop on Modeling and Simulation, 2002.

[26]Anonymous, "Aeronautical Design Standard-33E-PRF, Performance Specification, Handling Qualities Requirements for Military Rotorcraft," Technical report, US Army AMCOM, Redstone, Alabama, USA, March 2000.

[27]Padfield, G. D., *Helicopter Flight Dynamics: The Theory and Application of Flying Qualities and Simulation Modelling*, Blackwell Publishing, second edition, 2007.

[28]Hossein Mansur, M., "Development and Validation of a Blade-Element Mathematical Model for the AH-64A Apache Helicopter," Technical Report NASA-TM-108863, NASA, 1995.

[29]Wentink, M., Bles, W., Hosman, R. J. A. W., and Mayrhofer, M., "Design & Evaluation of Spherical Washout algorithm for Desdemona Simulator," Paper AIAA-2005-6501, Proceedings of the AIAA Modeling and Simulation Technologies Conference and Exhibit, San Francisco (CA), 2005.

APPENDIX

Pilot comments

Subject 1 The subject noted during b30s that overshoot bothered him and that it felt worse in conditions d60s, noting it felt "more like sitting in a rocking chair". During d60s, pilot also indicated he noticed no adaption of control strategy between conditions, even though the ratings between the conditions differ. b60s was described as feeling a little better than d60s. During condition d30s, pilot noted that in reality, a less aggressive strategy would be adopted and that it would be easier to level off towards the end of the maneuver. This is somewhat consistent with this subject's comments during condition b30s.

In condition d30m, the subject commented that motion felt better than prior condition d30s and that the "g-force" felt better; pilot still noted motion seemed to "lag" too much. He also emphasized the importance of timing control inputs to mitigate overshoot as much as possible. During conditions b60m and d60m, pilot provided the indication "R" in conjunction to the MFR rating, indicating the motion deficiency "return to neutral". During condition b60m, pilot commented that motion seemed to level out slower than in reality.

Subject 2 During the training trials with the baseline condition, the subjected noted that motion seemed to lead visuals and that the lateral velocity of the helicopter diminished too fast after levelling out. In a later training trial, however, this pilot noted that lateral motion seemed to "continue" while being confident that the helicopter was stable. The subject attributed this peculiarity to the tail rotor. During condition b30s, the pilot remarked on the high workload required in levelling out the helicopter, noting in particular the "high inertia" and the need for a "jerky" deceleration to hover. During condition b60s, the pilot again noted that the velocity built up was lost too quickly after levelling out. During condition d30s, the pilot remarked that it felt like he was "flying a different helicopter every time". The pilot noted that d30s, however, seemed to react more promptly to cyclic inputs and was therefore easier to control and level out towards the end of the maneuver. During d60s, pilot noted that motion felt exaggerated and distracting, but otherwise “pleasant” to fly. The awarded SFR ratings in favour of degraded rotorcraft dynamics with baseline motion therefore seem consistent with the provided comments, in that the pilot preferred the more agile rotorcraft dynamics in condition d30s.

The subject furthermore perceived motion in condition b30m to be too strong, but still realistic. In addition, the subject noted fidelity diminished for a more aggressive control strategy and seemed to improve when a less aggressive strategy was adopted. During condition b60m, the pilot commented on a “noticeable Coriolis effect” and noted he was “actively ignoring motion”. In addition, the pilot noted motion seemed to “tilt” too much, noting exaggerated “leans” and provided the indicator “M” (for mismatch cues). Condition d30m was deemed “realistic” by the pilot, except toward the end of the maneuver where motion seemed to “overshoot“ in the “wrong direction”. The “flight model”, however, felt “pleasant”. Here, it clearly appears that a change in rotorcraft dynamics was perceived by the pilot as a change in motion. During condition d60m, the pilot noted that motion seemed to “overcompensate” and “overshoot”. Indicators “M” (for mismatch cues) and “O” (for onset cues) were provided as well, with the additional comment that motion seemed to “lag” and “overshoot”.

Subject 3 During training sessions, the subject noted that the helicopter lost lateral velocity too quickly as compared to the real helicopter. During condition b30s, pilot noted that adequate performance could be attained with more trials, but provided a rating of 10 because adequate performance could not be achieved on the basis of the four opportunities given. During condition d30s, the pilot noted that control inputs had to be given “more in advance” and that “anticipation for overshoot” (i.e., lead) was required. Similar comments were given during condition b60s, where the pilot also noted that the applied task strategy bared resemblance to what would be required in a Huey (Bell UH-1) helicopter. During condition d60s, the pilot remarked that the experience was “significantly” different from the real helicopter and that, in hindsight, the pilot would “maybe” improve the rating awarded in the previous condition (b60s) from a value of 7 to a value of 6.

During condition b30m, the pilot noted motion matched “reasonably well”, but during levelling out of bank the pilot commented on mismatch (“out of phase”) in motion. Condition d30m felt notably different to the pilot, again resembling the Huey helicopter and noting that more anticipation was required due to an apparent “lag” in inputs. The pilot therefore provided indicator “L” in conjunction with motion rating, indicating latency. Condition b60m was received by the pilot with the comment that it would “be a good one for PIO”, although a significant adaptation in control strategy was “not necessary”. The pilot also noticed that motion “continued somewhat longer”, but did contribute to the “overall feeling”. The indicator “L”, signifying latency, was again provided. During condition d60m, the pilot seemed to notice that motion moved in an opposite direction to control inputs. The pilot also expressed doubt in providing indicators “R” and/or “M” (for return to neutral and mismatch cues). In this condition, the pilot also noted that motion seemed “50% useful and 50% distracting”. Initial cues to control inputs seemed “OK”, but during levelling out a significant mismatch was perceivable.

Subject 4 During training sessions, the pilot remarked that the helicopter seemed “supersensitive” and seemed to act “opposite to inertia”, meaning it was much more sensitive at low movement rates than at high movement rates. Baseline condition b30s was complimented with having “by far the best correlation with actual helicopter” and that the response to controls seems “on the spot”, adding that “if training in the sim, best transfer [to real helicopter] would be obtained”. With degrading dynamics (condition d30s), the pilot noted that a minimal adaptation of control strategy was required, but that motion appeared to be “exaggerated” and that “from pressure on body” it seems like “flying faster than previous condition” (d30m) and that motion “feels like going faster than it looks”. Here, it again seems that a pure change of rotorcraft dynamics is perceived by a subject as a degradation of motion fidelity. In condition b60s, the pilot noted that “in between crosses” the experience felt “very similar” and that the motion also “feels realistic”. However, upon approaching the cross, the pilot noted that motion felt “very deceiving” and that the response to controls “seems exaggerated”. The diminishing rating was therefore attributed to the motion and it was added that an adaptation of control strategy was noticeable from the first trial to the second trial. Degradation of both rotorcraft dynamics and motion in condition d60s resulted in the pilot noticing a “big difference with this one”, elaborating that “proprioceptive cues upon bank” seemed “bigger than what appears from visuals”. Also, for the “same” bank angle, motion seemed to “feel stronger” than visuals. However, the pilot added that this observation may have been due to “pre-conditioning in earlier experimental conditions”.

During condition b30m, the pilot noted that the sensation [of motion] “matched visual very well” and that it ”overall felt very normal”. The subject also added that “others conditions felt very simulator-like, this one felt more like the real thing”. During condition d30m, the pilot only mentioned that motion “felt close to real flight”. During condition b60m, the pilot noted that “overwhelming majority felt normal”, but “close to hover at banks, motion feels exaggerated”. Finally, during condition d60m, the pilot noted that it felt “very real”, with the “caveat” being that “motion and sim” feels too responsive upon initiating maneuver, adding it “happens too fast”. However, the “visual with motion” felt “very good” in this condition.