

## Retention of Manual Control Skills in Multi-Axis Tracking Tasks

Wijlens, R.; Zaal, Peter; Pool, Daan

**DOI**

[10.2514/6.2020-2264](https://doi.org/10.2514/6.2020-2264)

**Publication date**

2020

**Document Version**

Final published version

**Published in**

AIAA Scitech 2020 Forum

**Citation (APA)**

Wijlens, R., Zaal, P., & Pool, D. (2020). Retention of Manual Control Skills in Multi-Axis Tracking Tasks. In *AIAA Scitech 2020 Forum: 6-10 January 2020, Orlando, FL* (pp. 1-26). Article AIAA 2020-2264 (AIAA Scitech 2020 Forum; Vol. 1 PartF). American Institute of Aeronautics and Astronautics Inc. (AIAA). <https://doi.org/10.2514/6.2020-2264>

**Important note**

To cite this publication, please use the final published version (if applicable). Please check the document version above.

**Copyright**

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

**Takedown policy**

Please contact us and provide details if you believe this document breaches copyrights. We will remove access to the work immediately and investigate your claim.



# Retention of Manual Control Skills in Multi-Axis Tracking Tasks

Rowenna Wijlens\*  
*Delft University of Technology  
 Delft, The Netherlands*

Peter M. T. Zaal†  
*San José State University  
 NASA Ames Research Center  
 Moffett Field, CA, USA*

Daan M. Pool‡  
*Delft University of Technology  
 Delft, The Netherlands*

**This paper presents the results of a training and retention experiment conducted to objectively and quantitatively evaluate the acquisition, decay, and retention of skill-based manual control behavior in a compensatory dual-axis roll and pitch attitude tracking task. In this study, thirty-eight fully task-naive participants were trained in a fixed-base setting in the Human-Machine Interaction Laboratory at Delft University of Technology and subsequently divided into three matched groups based on their training performance and control behavior. Performance of the first group was re-evaluated after a period of non-practice of six months, whereas the second group was retested at both three and six months after training, and skill retention of the third group was measured after two, four, and six months. The goal of the experiment was to model the decay curve of skill-based manual control behavior and to determine the re-acquisition rate of lost skills compared to their initial acquisition rate. To explicitly quantify changes in manual control skills, learning curve models were fitted to metrics of task performance and control activity. The results suggest that control skills decay following a negatively accelerating decay curve and that lost skills are re-acquired at a higher rate than their initial development rate.**

## Nomenclature

$A_t[k]$	=	amplitude of $k^{\text{th}}$ sinusoid, deg	$s$	=	Laplace operator
$e$	=	tracking error signal, deg	$T_{A_1}, T_{A_2}$	=	amplitude filter time constants, s
$F$	=	learning curve learning rate	$T_m$	=	tracking run measurement time, s
$f_t$	=	target forcing function, deg	$t$	=	time, s
$H(j\omega)$	=	frequency response function	$u$	=	human operator control signal, deg
$H(s)$	=	transfer function	$y_{lc}$	=	learning curve value
$H_c$	=	controlled aircraft dynamics	$\delta_a$	=	aileron deflection, deg
$H_{pc}$	=	human operator crossfeed response	$\delta_e$	=	elevator deflection, deg
$H_{pe}$	=	human operator error response	$\theta$	=	pitch attitude, deg
$i$	=	tracking run index	$\rho$	=	Pearson's correlation coefficient
$j$	=	imaginary unit	$\sigma^2$	=	variance
$K_s$	=	stick gain	$\phi$	=	roll attitude, deg
$k$	=	forcing function sinusoid index	$\phi_t[k]$	=	phase shift of $k^{\text{th}}$ sinusoid, rad
$N_t$	=	number of sinusoids in forcing function	$\phi_{t,m}$	=	phase shift of $m^{\text{th}}$ sinusoid realization, rad
$n$	=	human operator remnant signal, deg	$\omega$	=	frequency, $\text{rad} \cdot \text{s}^{-1}$
$n_t[k]$	=	frequency integer factor of $k^{\text{th}}$ sinusoid	$\omega_m$	=	measurement time base frequency, $\text{rad} \cdot \text{s}^{-1}$
$p_a$	=	learning curve asymptotic value	$\omega_t[k]$	=	frequency of $k^{\text{th}}$ sinusoid, $\text{rad} \cdot \text{s}^{-1}$
$p_0$	=	learning curve initial value			
$S_{uu}$	=	control signal spectrum, $\text{deg}^2/(\text{rad} \cdot \text{s}^{-1})$			

\*M.Sc. student, Control & Simulation Section, Faculty of Aerospace Engineering, P.O. Box 5058, 2600GB Delft, The Netherlands; rowenna.wijlens@gmail.com. Student member AIAA.

†Senior Research Engineer, Human Systems Integration Division, NASA Ames Research Center, Moffett Field, CA, 94035; peter.m.t.zaal@nasa.gov. Senior Member AIAA.

‡Assistant Professor, Control & Simulation Section, Faculty of Aerospace Engineering, P.O. Box 5058, 2600GB Delft, The Netherlands; d.m.pool@tudelft.nl. Senior Member AIAA.

### Subscripts

$c$	=	crossfeed response
$e$	=	error response
end	=	end of session
st	=	start of session
$t$	=	target forcing function

tr	=	training phase
$x$	=	signal component
$\theta$	=	pitch
$\phi$	=	roll
6M	=	6-month retention test

## I. Introduction

PILOTS' manual flying skills have degraded due to the increase in flight deck automation over the last decades [1, 2]. This has resulted in a growing concern that today's pilots perhaps lack the skills to safely and successfully prevent or recover from unexpected upset events or to take over control after a sudden transition to manual flying [3]. Although the development of additional standards and guidelines for (recurrent) training procedures is a topic of current interest [1–8], additional research is required to be able to implement scientifically substantiated standards to ensure pilots receive sufficient training opportunities to develop, maintain, and improve manual flying proficiency [2, 3].

Likewise, the retention of manual control skills, i.e., the ability to perform these skills after a period of disuse and non-practice, becomes ever more important in space flight [9, 10]. Especially for long-duration deep space missions, mission success and safety depend on the autonomy and retained skills of the astronaut crew, as traditional risk mitigation factors are inaccessible. To prevent skill decay, i.e., the loss of trained skills over a period of non-practice, during space operations, in transit to the mission destination, or even before launch – i.e., due to the lengthy process of ground training – it is necessary to investigate how (e.g., onboard refresher) training should be designed to better support long-duration deep space operations [9].

Although the first retention studies concerning the development of manual flying skills were already conducted in the 1960s and 1970s [11–18], understanding *skill decay* and *retention* has remained a challenging task up to this day. Besides the fact that many who have tried believe that a universal skill decay curve – i.e., the trend of skill decay over time – might not exist [19], two other reasons for this challenge can be put forward. Firstly, skill decay research has been limited due to its challenging nature, as generally long retention periods (i.e., periods of non-practice) are involved. Earlier research was often comprised of short retention intervals (RIs) ranging from less than an hour to a few days, or a few weeks at most [20, 21], evaluating only a small part of the skill decay curve. Secondly, earlier research on skill decay cannot be compared in a fair manner due to their use of different performance measures, which could influence the shape of the measured skill-decay trend [19]. The main focus of this paper will be on the former concern, whereas the latter will be focused upon in our future work.

The goal of this paper is to explicitly analyze the retention of skill-based manual control behavior in a multi-axis tracking task. To accomplish this, a human-in-the-loop experiment was performed in the fixed-base simulator in the Human-Machine Interaction Laboratory (HMILab) at Delft University of Technology (TU Delft) with 38 fully task-naive participants. In the first phase of the experiment, all participants were trained (100 tracking runs) under the same conditions in a challenging compensatory dual-axis roll and pitch attitude tracking task. After the training phase, the participants were divided into three matched groups based on their task performance and control behavior during training. In the subsequent retention phase, participants performed the same tracking task as during the training phase to be able to re-evaluate their performance after a period of non-practice. The three groups differed from one another in the length of the RI and the number of retention tests (RTs) they performed, with intervals ranging from 2 to 6 months.

This experiment setup of three experiment groups with different RIs enabled the current study to address three questions. First, “what trend does the decay curve of manual control skills of novices follow?” Second, “what is the optimal RI to ensure that manual control skills of novices do not decay significantly, while at the same time minimizing the amount of refresher training?” The last question is “how does the re-acquisition rate of manual control skills of novices during retention testing compare to their initial acquisition rate?” This research is performed with novices, as it is impossible for general aviation pilots to refrain from flying for the duration of the research, a stringent and important requirement to obtain reliable results.

Changes in human operator control behavior over the course of the experiment were explicitly analyzed using metrics of task performance and control activity. In addition, learning trends in the considered control behavior metrics over the course of the experiment were quantified using fitted exponential learning curve models. Finally, the considered metrics were decomposed into individual contributions from distinct sources to gain a deeper insight into the characteristics and retention effects of human operator control behavior in dual-axis tracking tasks.

This paper is structured as follows. The methods, experimental setup, and hypotheses are described in Sec. II. Section III presents the results of the experiment. The paper ends with a discussion and conclusions.

## II. Method

### A. Control Task

A schematic representation of the compensatory dual-axis roll and pitch attitude tracking task performed to assess the retention of manual control skills is shown in Fig. 1. Participants were required to follow the desired roll and pitch attitudes, specified by the target forcing functions  $f_{t\phi}$  and  $f_{t\theta}$ , as accurately as possible by simultaneously minimizing the roll and pitch errors,  $e_\phi$  and  $e_\theta$ , respectively. Participants controlled the roll and pitch attitudes,  $\phi$  and  $\theta$ , which are the outputs of the aircraft roll and pitch dynamics,  $H_{c\phi}$  and  $H_{c\theta}$ , respectively, using a sidestick with roll and pitch gains  $K_{s\phi}$  and  $K_{s\theta}$ . The roll and pitch errors were presented on a dual-axis compensatory display, similar to an attitude indicator, as, respectively, the angle and vertical distance between a reference line, representing the artificial horizon, and a static aircraft symbol, as also depicted in Fig. 1.

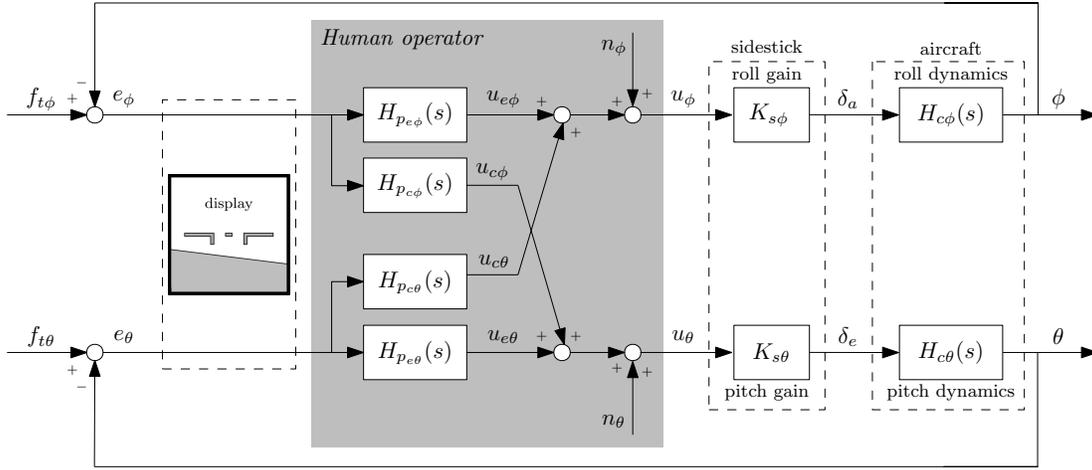


Fig. 1 Schematic representation of the compensatory dual-axis roll and pitch attitude tracking task with crossfeed present.

Whereas, following the current state-of-the-art in manual control, multi-axis control is often modeled as if multiple fully independent single axes are controlled [22, 23], earlier research has shown that manual control in multi-axis tasks is actually markedly different from single-axis control [24–28]. One of the observed phenomena in multi-axis human manual control is the presence of crossfeed [24, 25, 27–29]. Crossfeed can be described as a form of task interference in which the human operator is not able to completely decouple two tasks [25].

In the compensatory dual-axis roll and pitch attitude tracking task, as shown in Fig. 1 with crossfeed present, the human operator can be modeled through four different linear operator response functions: the error responses  $H_{pe\phi}$  and  $H_{pe\theta}$  respond to the principal-axis error signals in roll and pitch, respectively, and the crossfeed responses  $H_{pc\phi}$  and  $H_{pc\theta}$  react to the off-axis error signals. The operator’s roll and pitch control inputs,  $u_\phi$  and  $u_\theta$ , respectively, both consist of a linear error and a linear crossfeed response,  $u_e$  and  $u_c$ , respectively, and an additional remnant  $n$ , accounting for nonlinear behavior and measurement noise [30]. This control structure has successfully been applied in earlier multi-axis manual control research in which the presence of crossfeed was investigated [28].

### B. Controlled Aircraft Dynamics

To make the tracking task feel as realistic as possible different aircraft roll and pitch dynamics were used. The linearized roll and pitch dynamics are defined by Eqs. (1) and (2), respectively. These are the controlled aircraft dynamics of a medium-sized twin-engine transport aircraft, similar in size to a Boeing 757. The gross weight of the aircraft was set to 185,800 lbs. The aircraft dynamics were linearized at a flight condition close to the stall point, at an airspeed of 150 kts and an altitude of 41,000 ft. These aircraft dynamics have successfully been applied in earlier research into the training of multi-axis manual control tasks [31].

$$H_{c\phi}(s) = \frac{\phi}{\delta_a} = \frac{0.76773(s^2 + 0.2195s + 0.5931)}{(s + 0.7363)(s - 0.01984)(s^2 + 0.1455s + 0.6602)} \quad (1)$$

$$H_{c\theta}(s) = \frac{\theta}{\delta_e} = \frac{0.33282(s^2 + 0.09244s + 0.002886)}{(s^2 - 0.01388s + 0.004072)(s^2 + 0.446s + 0.4751)} \quad (2)$$

The linearized roll dynamics of Eq. (1) have a mildly unstable pole (spiral) in this flight condition. The roll dynamics approximate a single integrator ( $\frac{1}{s}$ ) at low frequencies up to 0.8 rad/s and a double integrator ( $\frac{1}{s^2}$ ) at higher frequencies, as shown in Fig. 2a. The linearized pitch dynamics of Eq. (2) have an unstable phugoid. The pitch dynamics approximate a double integrator ( $\frac{1}{s^2}$ ) at frequencies higher than 0.6 rad/s, as shown in Fig. 2b. Both the roll and pitch dynamics required the operator to adopt lead equalization, making the task rather challenging to perform [30, 32].

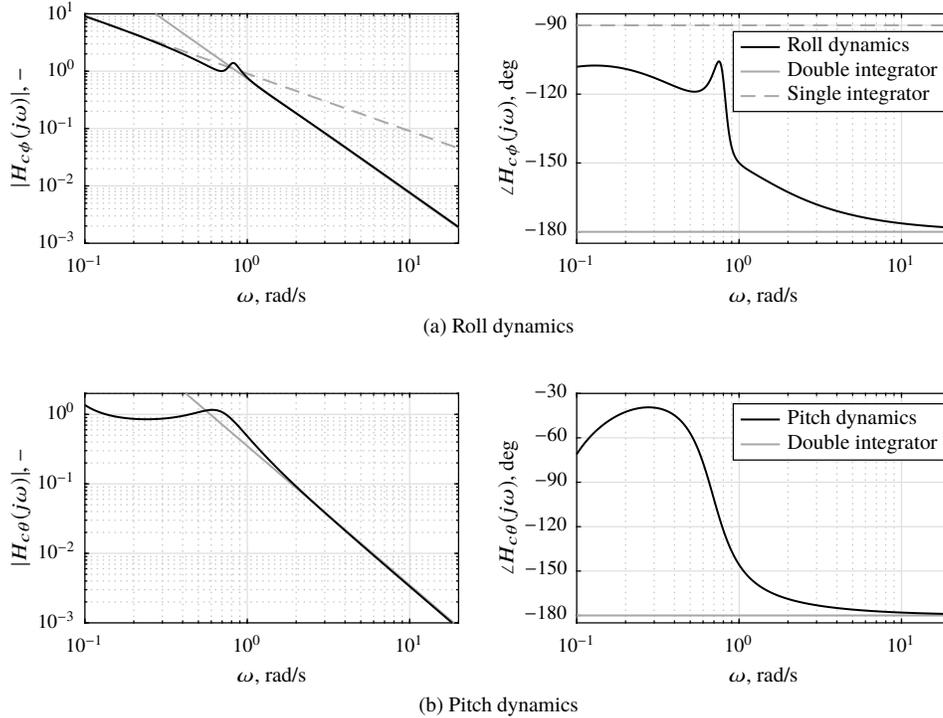


Fig. 2 Frequency response of the linearized aircraft dynamics.

### C. Forcing Functions

To facilitate reliable identification of human operator control dynamics, the roll and pitch target forcing functions,  $f_{t\phi}$  and  $f_{t\theta}$ , respectively, were constructed as two independent sum-of-sines signals [32, 33]:

$$f_{t\phi,\theta}(t) = \sum_{k=1}^{N_{t\phi,\theta}} A_{t\phi,\theta}[k] \sin(\omega_{t\phi,\theta}[k]t + \phi_{t\phi,\theta}[k]) \quad (3)$$

where  $N_{t\phi,\theta}$  is the number of sines used and  $A_{t\phi,\theta}[k]$ ,  $\omega_{t\phi,\theta}[k]$  and  $\phi_{t\phi,\theta}[k]$  represent the amplitude, frequency, and phase of the  $k^{\text{th}}$  sine in  $f_{t\phi}$  or  $f_{t\theta}$ , respectively.

To allow for estimating frequency domain describing functions for  $H_{pe\phi}$ ,  $H_{pe\theta}$ ,  $H_{pc\phi}$ , and  $H_{pc\theta}$  using an extended frequency-domain Fourier coefficient method [28, 33, 34], the frequencies of the individual sinusoids,  $\omega_{t\phi,\theta}[k]$ , were defined as integer multiples of the measurement time base frequency, meaning  $\omega_{t\phi,\theta}[k] = n_{t\phi,\theta}[k]\omega_m$ , where the measurement time base frequency  $\omega_m = 2\pi/T_m = 0.0767$  rad/s and the measurement time  $T_m = 2^{13} = 8192$  ms. The measurement time was taken as the last 81.92 seconds of a 90-second run, where the first 8.08 seconds were considered

the run-in time, as done in many previous tracking studies [28, 31, 35–42]. This run-in time was included in a tracking run, but discarded for data analysis to remove the initial transient response resulting from participants stabilizing the controlled aircraft dynamics and adjusting to the task. Table 1 lists all parameters of the target forcing functions used in pitch and roll.

**Table 1 Experiment forcing function data.**

Pitch target signal, $f_{t\theta}$							
$n_{t\theta}, -$	$\omega_{t\theta}, \text{rad/s}$	$A_{t\theta}, \text{deg}$	$\phi_{t\theta,1}, \text{rad}$	$\phi_{t\theta,2}, \text{rad}$	$\phi_{t\theta,3}, \text{rad}$	$\phi_{t\theta,4}, \text{rad}$	$\phi_{t\theta,5}, \text{rad}$
3	0.230	1.404	6.137	3.088	6.118	2.355	3.703
7	0.537	1.229	2.041	5.551	5.407	4.129	0.244
13	0.997	0.896	3.634	0.901	3.296	1.360	3.050
29	2.224	0.366	2.536	0.616	4.078	2.272	2.251
41	3.145	0.218	0.866	0.978	2.904	0.833	5.150
53	4.065	0.146	4.636	1.245	2.919	2.333	3.509
73	5.599	0.091	4.345	2.019	0.920	5.331	4.573
103	7.900	0.058	2.748	4.612	1.687	3.547	4.034
139	10.661	0.042	5.681	2.675	4.146	4.951	1.065
194	14.880	0.033	3.803	5.144	5.621	3.641	5.280
Roll target signal, $f_{t\phi}$							
$n_{t\phi}, -$	$\omega_{t\phi}, \text{rad/s}$	$A_{t\phi}, \text{deg}$	$\phi_{t\phi,1}, \text{rad}$	$\phi_{t\phi,2}, \text{rad}$	$\phi_{t\phi,3}, \text{rad}$	$\phi_{t\phi,4}, \text{rad}$	$\phi_{t\phi,5}, \text{rad}$
2	0.153	1.334	0.300	2.381	4.068	4.619	6.002
5	0.384	1.239	0.779	3.931	2.995	4.273	1.254
11	0.844	0.937	2.880	4.957	6.065	4.753	1.007
23	1.764	0.467	2.367	3.478	5.460	1.650	3.055
37	2.838	0.238	4.319	0.335	5.556	0.730	2.074
51	3.912	0.145	4.056	2.990	0.593	0.550	2.652
71	5.446	0.088	1.421	5.516	1.169	4.398	5.213
101	7.747	0.055	5.717	1.195	3.397	3.815	3.439
137	10.508	0.040	3.634	2.205	2.811	2.204	5.957
191	14.650	0.031	3.431	0.527	4.760	6.161	2.335

To form sufficiently unpredictable forcing functions [32, 43], as well as to capture all human operator dynamics over the frequency range of interest, while assuring a high signal-to-noise ratio to maximize identification accuracy [44], both the roll and pitch target forcing functions were the sum of  $N_{t\phi,\theta} = 10$  individual sinusoids covering the frequency range of human control at regular intervals on a logarithmic scale. Additionally, the integer multiples of the individual sinusoids were selected such that they were not multiples of one another to prevent higher harmonics and thereby ensure that the target signal was not recognizable [45].

A second-order low-pass filter was used to define the amplitudes of the individual sines in both the roll and pitch target forcing functions. This low-pass filter is described by Eq. (4) and has been considered in many previous tracking studies [28, 35, 36, 39–41, 45, 46] to reduce amplitudes at higher frequencies, yielding a more feasible control task and minimizing the chances of crossover regression [32, 44, 47].

$$A_{t\phi,\theta}[k] = \left| \frac{1 + T_{A_1} j\omega_{t\phi,\theta}}{1 + T_{A_2} j\omega_{t\phi,\theta}} \right|^2 \quad (4)$$

In Eq. (4)  $T_{A_1} = 0.1$  s and  $T_{A_2} = 0.8$  s. The amplitude distributions  $A_{t\phi,\theta}[k]$  were scaled to attain variances for  $f_{t\phi,\theta}$  of  $1.5 \text{ deg}^2$ .

To ensure that describing functions resembled real-life control behavior as closely as possible, target forcing functions with a Gaussian magnitude distribution were desired. Also, to prevent peaks which cause sudden moments of high workload, target signals were required to have an average Crest Factor (CF) [47, 48]. The CF depends on the choice of the respective phases  $\phi_{t\phi,\theta}$  of the individual sinusoids. To determine the forcing function phase distributions, 10,000 random sets of phases were generated. Sets that yielded signals with a Gaussian-like distribution and an average CF were selected [48]. For both the roll and pitch target forcing functions, five different realizations were used, differing only in their phase distributions  $\phi_{t\phi,\theta}$  (see Table 1). These five different forcing function realizations in roll and pitch yielded five different forcing function settings, as the  $m^{\text{th}}$  realization in roll was always paired with the  $m^{\text{th}}$  realization in pitch. The different forcing function realizations were used to assure that it was virtually impossible for participants to detect patterns and be able to anticipate the signal, in which case participants would introduce feedforward behavior and thereby change the control structure to a system including additional feedforward paths for both roll and pitch [32, 43].

## D. Apparatus

The experiment was performed in the fixed-base simulator setup in the HMILab at TU Delft, as shown in Fig. 3. To give roll and pitch control inputs, participants used a control-loaded hydraulic sidestick with  $\pm 30^\circ$  excursion in roll and  $\pm 22^\circ$  excursion in pitch. The sidestick was installed on the right-hand side of the participants' seat, which was a fully adjustable aircraft seat. Participants could adjust this seat to their preferred position. The compensatory display was shown on the Primary Flight Display (PFD) directly in front of the participants. The display update rate was 100 Hz and the time delay of the image generation was in the order of 20-25 ms. The size of the compensatory display was 11.0 cm  $\times$  11.2 cm (width $\times$ height). Besides this display, no other visual information (e.g., outside visual) was displayed during the experiment. Measurement data were logged at a sampling frequency of 100 Hz.

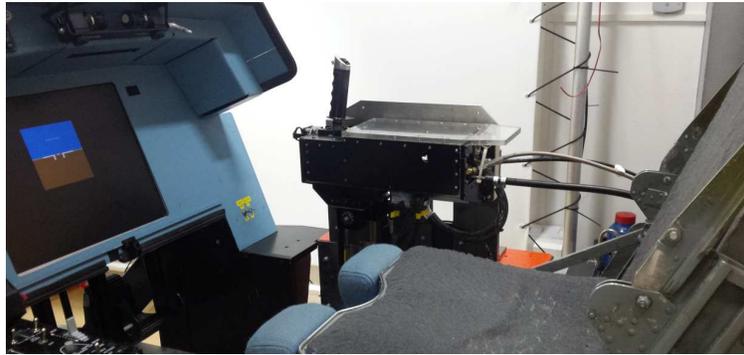


Fig. 3 Experiment setup in the HMILab at TU Delft.

## E. Participants and Instructions

A total of 38 fully task-naïve participants completed the experiment and all gave written informed consent for their participation. They also agreed to refrain from participation in all other tracking or flying experiments until having completed their participation. All participants were students at TU Delft, except for one, who had graduated from the university five months prior to the training phase of the experiment. The majority of students were from the Faculty of Aerospace Engineering. The participants were between 18 and 32 years old at the time of training, with an average age of 21.1 years and a standard deviation of 2.9 years. Twenty-eight participants were male, and ten female. All participants were comfortable operating the sidestick with their right hand.

The participants received a briefing before the start of the experiment. In this briefing the objective of the study, the dual-axis tracking task and the experiment procedures were explained, without disclosing the research questions and accompanying hypotheses. Additionally, participants filled out a pre-experimental questionnaire to determine participants' previous experience with tracking tasks. The answers to this questionnaire were used as reference when analyzing the experiment results. The experiment protocol and all forms participants received or were required to fill in had been approved by the Human Research Ethics Committee of the TU Delft prior to the experiment.

## F. Experiment Procedures

To evaluate the retention of manual control skills, a human-in-the-loop experiment was conducted consisting of two phases, referred to as the *training* phase and the *retention* phase. During the training phase, all participants received *ab initio* training in the dual-axis tracking task under the same conditions. After several months of no practice, the period referred to as the *retention interval* (RI), participants returned for the retention phase, where the same tracking task was performed as during the training phase.

The training phase of the experiment consisted of a fixed number of 100 tracking runs. These 90-second runs were performed in four sessions of 25 runs each, with each session on a separate (successive) day. This was done in order to enable skill improvement between training sessions, an effect known as offline learning (i.e., consolidation of learned control skills while not physically performing the task), as sleep enables offline skill improvement following explicit (intentional) learning [49]. Although there is no solid consensus yet on the optimum amount of time between consecutive training sessions, in a meta-analysis by Katak and Winstein [50] it was found that for low-level motor skills a retention time between training sessions of 24 hours can be considered close to an optimum. It was not possible

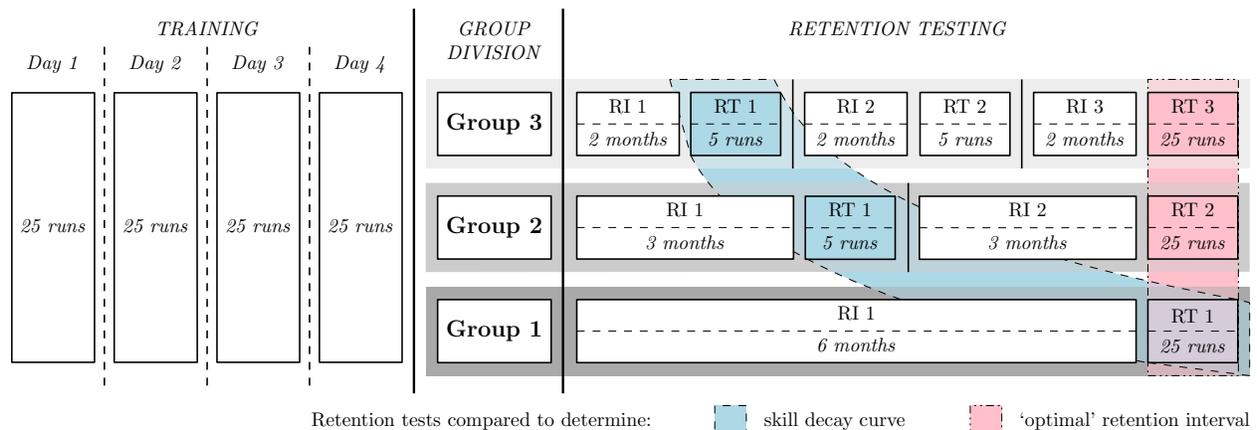
to completely honor the 24-hour break between training sessions by having all participants perform their training sessions at the same time every day. However, at least 14 hours of rest were scheduled between consecutive training sessions, including a night's sleep. During each training session, a five-minute break, in which participants left the simulator, was held after the first 15 runs. After the break, participants performed the last ten runs of the session. These breaks within training sessions were held to promote the participant's concentration during the training runs.

After each run, participants received feedback on their performance in roll and pitch by displaying their scores (the root mean square (RMS) of the tracking error signals in roll and pitch, respectively) on the PFD. Participants were encouraged to improve (i.e., lower) their scores with each tracking run. After each run, participants were asked if they were ready for the next run. In case of a positive response, the next run was started. Otherwise, participants were offered a brief break in order to ensure that their concentration levels were high and as constant as possible throughout the training session.

Although no actual evidence has been found favoring spaced practice, i.e., training with reasonably long rest intervals between separate training sessions, over massed practice, i.e., training with no or only short rest intervals between successive training sessions, for the retention of control skills [51–53], a spaced practice schedule was applied. Individuals training with a massed practice schedule often show worse performance than the performance level that would reflect their actual learning due to the effects of boredom and fatigue [54–56]. Therefore, spaced practice was preferred for this human-in-the-loop experiment, to be able to accurately capture the true learning curves of participants.

After all participants had completed the training phase, they were divided into three matched groups based on two criteria: their task performance and control behavior during the training phase, and their availability for retention testing. Training performance and control behavior of participants were evaluated based on two criteria: (1) by averaging their values of the tracking performance and control behavior metrics over the last ten training runs (runs 91–100), and (2) by fitting learning curve models (see Section II.H.1) to the RMS values of their tracking error signals throughout training to quantify the *learning rate*. These evaluations were made separately for pitch and roll. Subsequently, groups were formed such that there were no significant differences in the performance and control behavior metrics between the three groups to allow for a fair comparison of the retention performance of the different groups.

As illustrated in Fig. 4, the three groups differed in their RI length and in the number of *retention tests* (RTs) they performed. The first group, *Group 1*, only performed a single RT after a RI of six months. The second group, *Group 2*, performed two RTs with RIs of three months in between. The last group, *Group 3*, performed a total of three RTs with RIs of two months in between. This means that all participants performed their final RT six months after the end of training. At every RT participants were asked whether they had been involved in any activities during the RI that could either positively or negatively affect their retention performance. The final RT of each group was structured in the same manner as the training phase, meaning that the test consisted of a session of 25 90-second runs performed on a single day. The other RTs – i.e., the first RT of Group 2 and the first two RTs of Group 3 – consisted of only five 90-second runs. These five-run tests were kept short on purpose, to be able to capture the participants' performance at that moment in time, while at the same time minimizing additional learning.



**Fig. 4** Experiment setup (RI = Retention Interval, RT = Retention Test).

Whereas the first RT of each group was used to identify the trend of the skill decay curve, as indicated in blue in Fig. 4, the use of all final, 6-month, RTs was twofold: (1) to identify the 'optimal' *RI* (red highlights in Fig. 4) to

prevent skill decay, and (2) to establish participants' *relearning rate* of lost skills, if any skill decay had occurred after six months. The latter objective was also the reason for the 6-month RTs to be longer again compared to the 2-month, 3-month, and 4-month RTs.

## G. Dependent Measures

To quantify the acquisition, decay, and retention of participants' skill-based manual control behavior, a number of different objective dependent measures were determined from the measurement data. These dependent measures were analyzed for each axis of control separately. First, tracking performance and control activity in the roll and pitch axes were evaluated in terms of the RMS of the error and control signals,  $RMS(e)$  and  $RMS(u)$ , respectively, per experiment run. Next, to investigate the presence of crossfeed and its training and retention effects, the fraction of the crossfeed contribution to the total error and control variances over the course of the experiment was calculated.

## H. Data Analysis Methods

### 1. Learning Curve Modeling

To quantitatively describe how human operators' performance and control behavior varies during training and after a period of non-practice, exponential learning curves were fitted to the considered dependent measures of tracking performance and control activity. The considered exponential learning curve model is given by Eq. (5) and has successfully been applied in earlier training studies to quantify learning trends [31, 41, 46, 57].

$$y_{lc}(i) = p_a + (1 - F)^i(p_0 - p_a) \quad (5)$$

In Eq. (5)  $y_{lc}$  is the learning curve value for tracking run  $i$ ,  $p_0$  the initial value,  $p_a$  the asymptotic value and  $F$  the learning rate. The parameters  $p_0$ ,  $p_a$  and  $F$  were determined using a nonlinear optimization method to minimize the summed squared error between the experimental data and the learning curve model. Separate learning curves were fitted to the training phase data and data from the final 6-month RT. No learning curves were fitted to the intermediate RT data due to the low number of tracking runs performed in those tests. To assess the quality-of-fit of the learning curves, Pearson's correlation coefficient  $\rho$  was calculated for each fitted curve. However, only in case  $\rho$  was higher than 0.5, learning curves are shown in Sec. III. This is done to ensure that the data for which learning curves with a correlation coefficient lower than 0.5 were found, are clearly visible, as the data themselves might give an indication of why it could be less suitable to fit a learning curve to it.

### 2. Between-Participant Variability

To assess the between-participant variability in tracking performance and control activity, the approach developed by Pool and Zaal [58] to assess training effectiveness in transfer-of-training experiments was used. To apply this approach to the current experiment, exponential learning curves as described above were fitted to the training and 6-month RT  $RMS(e)$  and  $RMS(u)$  data of individual participants. To compare the retention performance between participants, the instantaneous changes in tracking performance and control activity between the end of training and the start of the 6-month RT are defined by  $\bar{\Delta}RMS(e)$  and  $\bar{\Delta}RMS(u)$ , respectively. These measures were determined by subtracting the RMS values at the end of training ( $RMS(e)_{tr,end}$  and  $RMS(u)_{tr,end}$ ) from the RMS values at the start of the 6-month RT ( $RMS(e)_{6M,st}$  and  $RMS(u)_{6M,st}$ ), and subsequently dividing by the RMS values at the end of training, as described by Eq. (6). Whereas Eq. (6) shows the calculation for  $\bar{\Delta}RMS(e)$ , the same calculation can be made for  $\bar{\Delta}RMS(u)$ . The changes in tracking performance and control activity are in this case expressed as non-dimensional numbers to facilitate an easier comparison between individuals.

$$\bar{\Delta}RMS(e) = \frac{RMS(e) - RMS(e)_{tr,end}}{RMS(e)_{tr,end}} \quad (6)$$

In Eq. (6)  $RMS(e)$  without any subscripts is used to indicate that the change in RMS value compared to the end of training can be calculated for any tracking run. For all groups, the retention performance during the 6-month RT was analyzed instead of during the first RT of each group, as the 2-month and 3-month RTs of Groups 3 and 2, respectively, did not consist of a large enough number of runs to estimate learning curves.

### 3. Statistical Analysis

For statistical analysis of the variation in tracking performance and control activity throughout the training and retention phases, two-run averages of  $RMS(e)$  and  $RMS(u)$ , respectively, at several moments during the experiment were subjected to pairwise comparisons (dependent  $t$  tests). For all groups, these moments included the start and end of both the training phase and the 6-month RT. Additionally, two-run averages were taken at the start and end of the 3-month RT for Group 2 and at the start and end of the 2-month and 4-month RTs for Group 3. Sec. III presents the results of four types of pairwise comparisons, namely (1) between the start and end of the training phase, (2) between the end of training and the start of every RT, (3) between the start and end of every RT, and (4) between the end of training and the end of every RT. In case at least one of the compared samples could not be considered sufficiently normally distributed, a nonparametric Wilcoxon signed-rank test was applied instead of a dependent  $t$  test.

For statistical analysis of the differences in tracking performance and control activity between the three groups, the two-run averages of  $RMS(e)$  and  $RMS(u)$ , respectively, of the three groups were subjected to a one-way ANOVA. In case at least one of the compared samples was not sufficiently normally distributed, a nonparametric Kruskal-Wallis test was performed instead.

### 4. Human Operator Crossfeed

To contribute to the understanding of crossfeed in multi-axis manual control by verifying if training or retention effects would show, the presence of crossfeed and its training/retention effects were investigated in this dual-axis compensatory control task performed by task-naive participants. To gain a preliminary understanding of the amount of crossfeed present and its training/retention effects, the tracking error and control input variances were decomposed into individual contributions from the target signal of the principal axis, the target signal of the other axis (i.e., crossfeed), and the human operator remnant. To be able to separate these individual contributions, the variances of the measured error and control signals were calculated from spectral analysis, as the separate contributions provide power at independent frequencies [59].

Estimates of the variance contributions of the principal and off-axis target signals were obtained by integrating the power spectral density (PSD) only over the respective forcing function input frequencies, and then subtracting the remnant contribution at those input frequencies. The remnant contribution at a specific input frequency was estimated from the remnant signal power at adjacent non-excited frequencies, as the remnant signal power is continuously distributed over the frequency spectrum [32, 44]. The total remnant contribution was estimated by integrating over the remaining frequencies and then adding the remnant contributions found at the input frequencies of the principal and off-axis target signals. An example PSD illustrating the above is shown in Fig. 5. This figure shows a smoothed remnant spectrum to give an indication of the remnant contribution at the principal and off-axis target forcing functions' input frequencies.

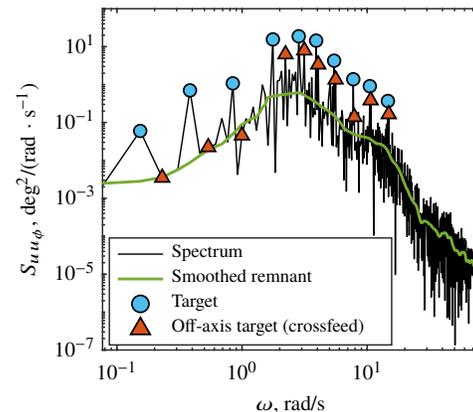


Fig. 5 Example PSD of roll control signal  $u_\phi(t)$  (participant 2, training run 68).

## I. Experiment Limitations

Due to the complex experiment procedures described earlier the introduction of several experiment limitations was inevitable. The most important limitations which had to be considered during data analysis were as follows:

### 1. Participants did not train to the same relative performance level

Some individuals have a more 'natural' ability than others in performing a task without prior practice, as a result of which they generally require less practice to reach a particular performance level [60]. Since all participants received the same number of training runs, this led to a situation in which at the end of the training phase, some participants had reached asymptotic performance, whereas others were still in the learning phase. This meant that at the end of training, part of the participants had 'overlearned' the tracking task, whereas others had actually 'underlearned' the task. The level of original learning is often seen as the most important determinant of the retention of control skills

[12, 14, 61–63]. Since overlearning is known to enhance retention [16, 19, 64, 65], this had to be taken into account when analyzing the experiment results.

In an ideal situation all participants would have been trained until they had just reached asymptotic performance, meaning that they would neither have overlearned nor underlearned the task, since the experiment was designed to only look at skill retention as a function of time, not at the effects of overlearning. However, training all participants to asymptotic performance would have meant that the number of training runs would have needed to be tailored to the individual and could only have been determined while training was taking place. Unfortunately, this scheduling uncertainty could not be accommodated for, because of simulator availability, having to avoid scheduling training on the weekends, as well as the large number of participants required.

### 2. Participants did not perform training at the same time every day

Training had to be scheduled around the individual (study) schedules of participants. This meant that training could not take place at the same time every day, which introduced a circadian confound. Because of the large number of participants required and the limited time available for the entire training phase of all participants, this limitation could not be avoided.

### 3. Retention intervals were not exact

The real RIs differed slightly from the ‘ideal’ ones due to participant availability. Especially the RIs of Group 3 (2-month RIs) contained some more variability. This was, in the first RI, caused by holidays, and was carried through in the second and third RIs in an attempt to ensure that the number of days between the end of training and the final, 6-month, RT of this group was as similar as possible to the number of days for Groups 1 and 2. The exact RIs can be found in Table 2\*.

**Table 2 Individual retention intervals (Partic. = Participant, RI = Retention Interval, RP = Retention Period, STD = Standard Deviation)**

Group 1		Group 2				Group 3				
Partic.	RI 1 / Total RP, days	Partic.	RI 1, days	RI 2, days	Total RP, days	Partic.	RI 1, days	RI 2, days	RI 3, days	Total RP, days
2	181	1	91	92	183	3	55	63	63	181
4	183	16	90	91	181	5	55	64	64	183
7	181	17	91	90	181	6	60	63	67	190
9	182	21	90	92	182	8	57	62	68	187
11	181	29	90	93	183	12	55	64	63	182
13	182	31	89	93	182	19	60	59	63	182
14	181	33	90	96	186	22	61	63	58	182
18	187	34	89	92	181	24	61	62	58	181
23	181	35	89	96	185	26	61	59	61	181
25	181	39	89	93	182	27	53	64	65	182
28	181	40	90	91	181	38	60	62	59	181
30	182	41	90	92	182	43	60	59	63	182
37	182	42	90	92	182					
Mean	<b>181.9</b>		<b>89.8</b>	<b>92.5</b>	<b>182.4</b>		<b>58.2</b>	<b>62.0</b>	<b>62.7</b>	<b>182.8</b>
STD	<b>1.6</b>		<b>0.7</b>	<b>1.7</b>	<b>1.5</b>		<b>2.8</b>	<b>1.9</b>	<b>3.1</b>	<b>2.7</b>
Ideal	182		91	91	182		60.7	60.7	60.7	182

## J. Hypotheses

Based on the findings of previous (dual-axis) tracking task experiments as well as several experiments concerning the retention of manual control skills, five main hypotheses were formulated for the current research. No hypotheses were formulated concerning the presence of crossfeed and its training/retention effects, as the limits and capabilities of the data analysis methods used to identify crossfeed are still being explored [28].

As observed in a number of earlier training experiments [41, 46, 66], clear effects of training were expected to occur. We hypothesized that training causes an improvement in performance and task proficiency (lower RMS( $e$ )) (**hypothesis 1**).

As also found in previous dual-axis tracking task experiments [28, 31, 38, 40, 42], it was expected that participants perform better in pitch than in roll both during training and retention testing (**hypothesis 2**). This will be visible through a lower RMS( $e$ ) in pitch than in roll.

\*Participant numbers range from 1 to 43 instead of from 1 to 38 as participants 10, 15, 20, 32, and 36 dropped out before the experiment was completed.

When comparing the retention results of the three groups, it was hypothesized that skill decay can be captured by a positively accelerating decay curve, meaning that at first, skills are retained fairly well, but at some point start to deteriorate at an increasing rate (**hypothesis 3**). This skill-decay trend has been found in two flying task experiments [67, 68] that are most comparable to the control task used in the current research.

During the last RTs, six months after training, the best performance and task proficiency was expected for Group 3, whereas the worst performance was anticipated for Group 1 (**hypothesis 4**). This expectation was based on the fact that individuals perform better at retention testing if they are provided with some form of practice during the RI [67, 69–71]. When comparing the final, 6-month, RTs of the three groups, the experiment setup can also be seen as if all groups have a RI of six months, during which Group 1 receives no practice at all, Group 2 receives one practice moment mid-interval and Group 3 receives two practice moments.

Additionally, during the final RTs, six months after training, degraded control skills of all three groups were predicted to be re-acquired at a higher rate than the initial acquisition rate during the training phase (**hypothesis 5**). Earlier retention experiments concerning motor skills have consistently shown that retraining after a RI up to performance levels achieved at the end of training requires less time than initial training, hardly ever exceeding 50% of the initial training time [12, 63, 72].

### III. Results

This section presents the experiment results. For results figures that show data from all experiment runs, solid black vertical lines indicate the interval between training and the first RT, as well as the intervals between subsequent RTs. At the top of each figure, the experiment phase is indicated, where ‘Training’ indicates the 100 training runs performed by all groups, ‘2’ and ‘4’ represent the 2-month and 4-month RTs of Group 3, ‘3’ indicates the 3-month RT of Group 2 and ‘6M’ the 6-month, final, RTs of all groups.

#### A. Tracking Performance

Tracking performance is defined in terms of the RMS of the pitch and roll error signals  $e$ , i.e., the errors presented to the human operator on the PFD. The lower the value of  $\text{RMS}(e)$ , the better the task performance is. Figure 6 shows the average pitch and roll RMS( $e$ ) per experiment run throughout the experiment. Average results per run are indicated with blue squares for Group 1, red triangles for Group 2 and yellow circles for Group 3. Gray error bars present the 95% confidence intervals of the mean data. Pearson’s correlation coefficients for comparison of the fitted learning curves and the data are presented in the figure legends for both the training phase and the 6-month RT as  $\rho = [\rho_{\text{training}}, \rho_{\text{retention}}]$ . The parameters of the fitted learning curves are presented in Table 3. Tables 4 and 5 present the statistical analysis results of the training and retention effects within groups and between groups, respectively.

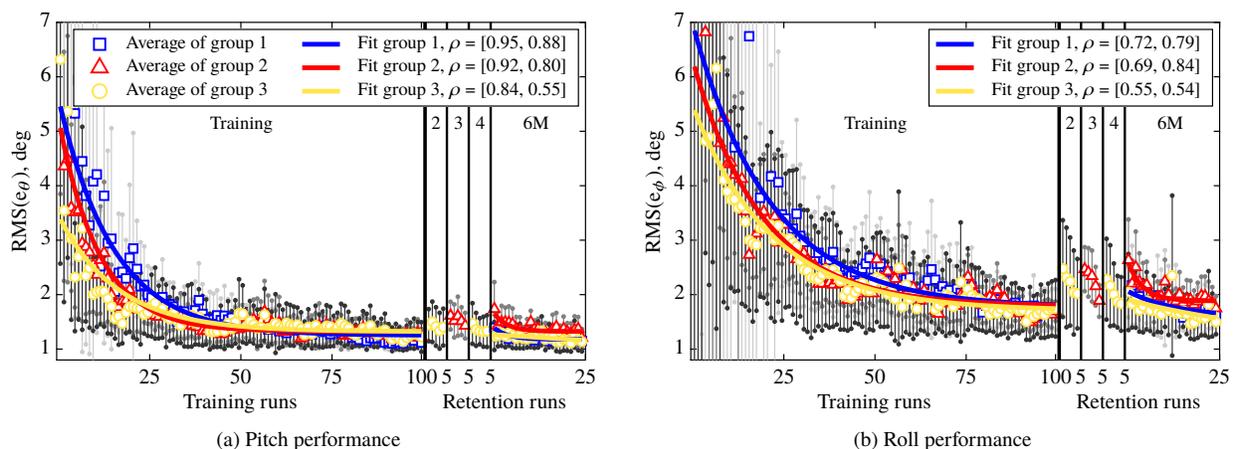


Fig. 6 Average pitch and roll tracking error with corresponding learning curves.

Figure 6 shows that at the start of training, average tracking errors varied between 3.5 deg and 5.8 deg in pitch and between 5.6 deg and 7.1 deg in roll for the three groups. All groups showed a steep initial learning curve for the first

**Table 3 Learning curve parameters for the pitch and roll tracking error RMS data**

RMS( $e$ ), deg	Pitch $\theta$						Roll $\phi$					
	Training phase			Retention phase			Training phase			Retention phase		
	$p_0$ , deg	$p_a$ , deg	$F(\times 10^{-2})$	$p_0$ , deg	$p_a$ , deg	$F(\times 10^{-2})$	$p_0$ , deg	$p_a$ , deg	$F(\times 10^{-2})$	$p_0$ , deg	$p_a$ , deg	$F(\times 10^{-2})$
Group 1	5.75	1.25	6.45	1.44	1.18	22.62	7.09	1.77	4.55	2.09	1.58	7.01
Group 2	5.41	1.32	8.61	1.75	1.33	23.77	6.43	1.79	5.14	2.76	1.88	21.61
Group 3	3.52	1.32	5.99	1.27	-2.03	0.13	5.56	1.71	4.51	1.92	1.14	2.34

**Table 4 Statistical analysis results within groups for tracking error (TR = Training, St = Start, M = Month)**

RMS( $e$ )	Pitch $\theta$			Roll $\phi$		
	Group 1	Group 2	Group 3	Group 1	Group 2	Group 3
Training	**	** <sup>a</sup>				
End TR - St 2M			* <sup>a</sup>			** <sup>a</sup>
St 2M - End 2M			— <sup>a</sup>			* <sup>a</sup>
End TR - End 2M			— <sup>a</sup>			— <sup>a</sup>
End TR - St 3M		** <sup>a</sup>			**	
St 3M - End 3M		* <sup>a</sup>			**	
End TR - End 3M		*			*	
End TR - St 4M			— <sup>a</sup>			* <sup>a</sup>
St 4M - End 4M			— <sup>a</sup>			** <sup>a</sup>
End TR - End 4M			— <sup>a</sup>			— <sup>a</sup>
End TR - St 6M	* <sup>a</sup>	* <sup>a</sup>	—	* <sup>a</sup>	**	— <sup>a</sup>
St 6M - End 6M	** <sup>a</sup>	* <sup>a</sup>	—	** <sup>a</sup>	**	* <sup>a</sup>
End TR - End 6M	—	— <sup>a</sup>	—	— <sup>a</sup>	—	— <sup>a</sup>

<sup>a</sup> At least one sample not normally distributed, Wilcoxon signed-rank test applied instead of dependent  $t$  test.

\*\* = highly significant ( $p < 0.01$ )

Legend: \* = significant ( $0.01 \leq p < 0.05$ )

— = not significant ( $p > 0.05$ )

**Table 5 Statistical analysis results between groups for tracking error (G = Group, M = Month, RT = Retention Test, St = Start)**

RMS( $e$ )	Pitch $\theta$	Roll $\phi$
Start training	— <sup>a</sup>	— <sup>a</sup>
End training	— <sup>a</sup>	— <sup>a</sup>
Start 6-month test	— <sup>a</sup>	— <sup>a</sup>
End 6-month test	— <sup>a</sup>	— <sup>a</sup>
Start RT1 of each group (St 2M G3, St 3M G2, St 6M G1)	— <sup>a</sup>	— <sup>a</sup>

<sup>a</sup> At least one sample not normally distributed, Kruskal-Wallis test applied instead of one-way ANOVA.

25 runs in pitch and the first 50 runs in roll, followed by a more gradual decrease in RMS( $e$ ). The observed learning rates were around  $7 \times 10^{-2}$  in pitch and  $5 \times 10^{-2}$  in roll (see Table 3). Still, over the course of training, the differences in average tracking errors between the groups decreased. At the end of training, average RMS( $e$ ) values of around 1.30 deg in pitch and 1.76 deg in roll were observed for all three groups, as shown in Fig. 6 and Table 3. Although the differences in tracking performance between the groups were larger at the start than at the end of training, the statistical analysis results in Table 5 show that both at the start and at the end of training these differences were not significant, which was a desired result of the group division. The average tracking errors observed during training were slightly higher than those observed in an earlier training experiment with a comparable dual-axis tracking task [31]. However, this was not surprising, as the earlier experiment was performed with motion feedback, and task proficiency is often better when motion feedback is present [35, 36]. The statistical analysis results in Table 4 show that performance improvement during training was significant in both pitch and roll for all three groups. Nonetheless, over the course of the experiment pitch tracking performance was consistently better than roll tracking performance for all three experiment groups. This observation is consistent with earlier dual-axis tracking task experiments [28, 31, 38, 40, 42] and shows a general emphasis on pitch control. The fact that roll errors are more difficult to perceive on a PFD than pitch errors due to a lower pixel resolution might be the cause of this [28]. However, the difference in performance between pitch and roll decreased throughout training. Whereas at the start of training, the performance difference in pitch and roll was around 1.5 deg on average, this difference decreased to around 0.5 deg, i.e., around 30-35% of the initial difference, at asymptotic performance, as shown in Table 3.

For the retention of the final ‘learned’ performance level at the end of training, Fig. 6 clearly shows that to a large extent the end-of-training performance was retained until the 6-month RTs by all three groups. Though the RMS( $e$ ) values during the RTs were overall much lower than those at the start of the training phase, performance in roll was found to degrade more compared to the end of training than performance in pitch, after a period of inactivity. This observation is consistent with previous studies on skill retention [16, 19, 64, 65], as overlearning is known to enhance

retention. With steeper learning curves and earlier stabilization in pitch than in roll during the training phase, pitch control was evidently more ‘overlearned’ than roll control. As a result of the larger performance degradation in the roll axis, larger performance improvements were observed in roll compared to pitch during retention testing.

At the start of each group’s first RT (the 2-month RT of Group 3, the 3-month RT of Group 2 and the 6-month RT of Group 1), the  $RMS(e)$  increased on average by 0.19 deg in pitch and 0.58 deg in roll compared to the end of training. All of the instantaneous increases at the start of the groups’ first RTs compared to the end of training were significant (see Table 4). However, no significant performance differences were observed between the different groups at the start of their first RTs (see Table 5). This suggests that tracking performance follows a negatively accelerating decay curve, as performance decreases rapidly during the first months after training, after which the decrease starts to slow down.

During the 6-month RT, Group 2 performed consistently worse than Groups 1 and 3, which exhibited similar tracking performance (see Fig. 6). This can be considered a curious result, as from the earlier finding that operators perform better during retention testing if they have received some form of practice during the RI [67, 69–71], it was expected that Group 1 would show the worst performance during the 6-month RT. However, statistical analysis results in Table 5 show that tracking performance of the three groups was not significantly different from one another, neither at the start, nor at the end of the 6-month RT. As a result, the groups’ performances during the 6-month RTs did not give an indication of what the ‘optimal’ RI is while at the same time minimizing the amount of refresher training, as was one of the ideas behind this experiment setup. However, Table 4 indicates that when ‘refresher’ training was provided to Group 1 after two months, five ‘refresher’ runs were sufficient to decrease  $RMS(e)$  again to end-of-training values, whereas when ‘refresher’ training was given to Group 2 after three months, performance in both pitch and roll was still significantly different from end-of-training values after an equal number of five tracking runs.

When comparing learning rates between the training phase and the 6-month RT, it is shown in Table 3 that Groups 1 and 2 had higher learning rates during retention testing than during training, whereas Group 3 exhibited the opposite behavior, i.e., higher learning rates during training than during the 6-month RT. For Group 3, a very low learning rate of  $0.13 \times 10^{-2}$  in pitch during the 6-month RT even resulted in a negative asymptotic  $RMS(e)$  (see Table 3). The difference between Groups 1 and 2 and Group 3 can be explained by the fact that Groups 1 and 2 exhibited significant performance decrements at the start of the 6-month RT when compared to the end of training, both in pitch and roll, whereas Group 3 did not show any significant decrements compared to the end of training due to its earlier ‘practice’ opportunities in the 2-month and 4-month RTs. However, the higher learning rates during the 6-month RT compared to the training phase for both pitch and roll of Groups 1 and 2 suggest that lost control skills are re-acquired at a higher rate than their initial acquisition rate.

## B. Control Activity

Control activity is measured in terms of the RMS of the pitch and roll control signals,  $u_\theta$  and  $u_\phi$ , respectively. A lower  $RMS(u)$  indicates less control effort. Operator control activity for pitch and roll are shown in Fig. 7. The parameters of the fitted learning curves are provided in Table 6. Tables 7 and 8 present the statistical analysis results of the training and retention effects within groups and between groups, respectively.

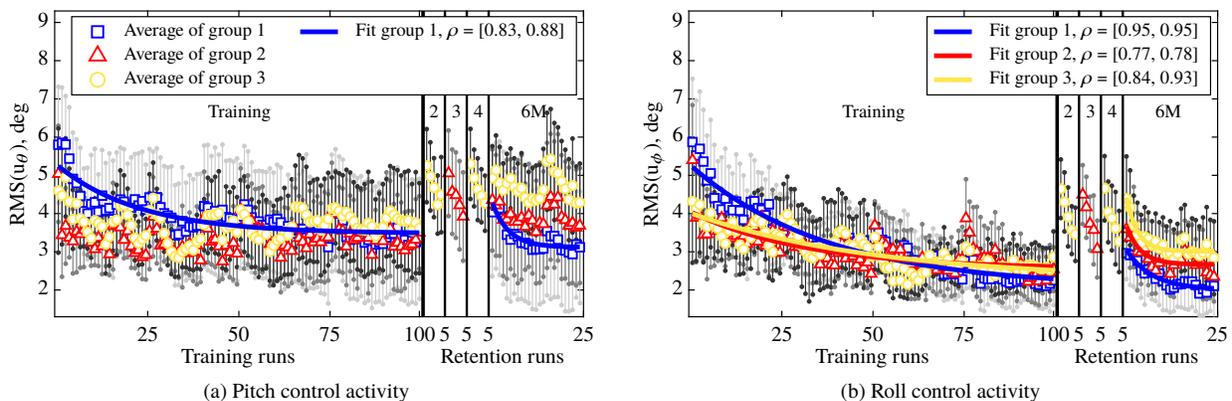


Fig. 7 Average pitch and roll control input with corresponding learning curves.

**Table 6 Learning curve parameters for the pitch and roll control input RMS data**

RMS( $u$ ), deg	Pitch $\theta$						Roll $\phi$					
	Training phase			Retention phase			Training phase			Retention phase		
	$p_0$ , deg	$p_a$ , deg	$F(\times 10^{-2})$	$p_0$ , deg	$p_a$ , deg	$F(\times 10^{-2})$	$p_0$ , deg	$p_a$ , deg	$F(\times 10^{-2})$	$p_0$ , deg	$p_a$ , deg	$F(\times 10^{-2})$
Group 1	5.32	3.48	4.31	4.56	3.12	19.65	5.30	2.06	2.60	3.20	2.01	13.07
Group 2	n/a	n/a	n/a	n/a	n/a	n/a	4.07	2.55	3.05	3.96	2.67	22.62
Group 3	n/a	n/a	n/a	n/a	n/a	n/a	4.03	2.26	1.92	4.96	3.02	27.97

**Table 7 Statistical analysis results within groups for control input (TR = Training, St = Start, M = Month)**

RMS( $u$ )	Pitch $\theta$			Roll $\phi$		
	Group 1	Group 2	Group 3	Group 1	Group 2	Group 3
Training	* <sup>a</sup>	—	— <sup>a</sup>	**	* <sup>a</sup>	*
End TR - St 2M			* <sup>a</sup>			**
St 2M - End 2M			— <sup>a</sup>			**
End TR - End 2M			* <sup>a</sup>			*
End TR - St 3M		**			** <sup>a</sup>	
St 3M - End 3M		* <sup>a</sup>			** <sup>a</sup>	
End TR - End 3M		— <sup>a</sup>			* <sup>a</sup>	
End TR - St 4M			* <sup>a</sup>			**
St 4M - End 4M			* <sup>a</sup>			**
End TR - End 4M			— <sup>a</sup>			*
End TR - St 6M	* <sup>a</sup>	— <sup>a</sup>	— <sup>a</sup>	*	— <sup>a</sup>	**
St 6M - End 6M	** <sup>a</sup>	— <sup>a</sup>	—	* <sup>a</sup>	** <sup>a</sup>	**
End TR - End 6M	— <sup>a</sup>	— <sup>a</sup>	— <sup>a</sup>	— <sup>a</sup>	— <sup>a</sup>	—

<sup>a</sup> At least one sample not normally distributed, Wilcoxon signed-rank test applied instead of dependent  $t$  test.

**Table 8 Statistical analysis results between groups for control input (G = Group, M = Month, RT = Retention Test, St = Start)**

RMS( $u$ )	Pitch $\theta$	Roll $\phi$
Start training	—	—
End training	— <sup>a</sup>	— <sup>a</sup>
Start 6-month test	— <sup>a</sup>	—
End 6-month test	— <sup>a</sup>	— <sup>a</sup>
Start RT1 of each group (St 2M G3, St 3M G2, St 6M G1)	— <sup>a</sup>	**

<sup>a</sup> At least one sample not normally distributed, Kruskal-Wallis test applied instead of one-way ANOVA.

Over the course of the training phase, the control input in pitch varied in a different manner than the control input in roll, as shown in Fig. 7. Whereas in roll the performance improvement during training was achieved with a significant decrease in control input for all three groups, the control input in pitch only decreased significantly for Group 1 (see Table 7). At the start of training, Group 1 exhibited the highest control activity in pitch and roll of around 5.8 deg and 5.5 deg, respectively, compared to around 4 deg for Groups 2 and 3 in both pitch and roll. However, at the end of training, Group 1 actually showed a control activity in pitch similar to that of Groups 2 and 3 (around 3.5 deg) and an even slightly lower control input in roll (around 2.3 deg for Group 1 compared to 2.6 deg for Groups 2 and 3). The aforementioned shows that the between-group differences in control input decreased throughout training for both pitch and roll. Also, whereas at the start of training control activity was very similar in pitch and roll, at the end of training control activity in roll had decreased to below RMS( $u_\theta$ ). The significant decreases in control input throughout training were achieved despite a significant spread in control input data. Although the spread was less prominent in roll compared to pitch, this spread is consistent with earlier findings [35, 36, 41].

As a desired result of the group division, the control input differences between the groups were not significant in pitch or roll, neither at the start, nor at the end of training (see Table 8). However, a clear “sawtooth shape” was observed in the control input group averages throughout training, which can be explained by the motivation and attention span of operators. Control activity started out relatively high at the start of each training day and reduced as motivation or attention gradually decreased. Control activity sometimes rose again in the last few runs before the break within the training sessions, as controllers regained motivation due to the upcoming break. After the break, control activity started out higher again due to increased attention and then gradually decreased throughout the second segment of the training sessions. Control activity increased once more during the last few runs of the sessions as controllers regained motivation as the end of the session neared. This effect was, however, more prominent in pitch than in roll and more noticeable for Groups 2 and 3 compared to Group 1.

During the retention phase, control activity was higher in pitch than in roll, as was also the case at the end of training. When comparing retention and training performance, Fig. 7 shows that for all groups and RTs control activity in both pitch and roll started out higher than at the end of training. Although due to the large spread, not all of these

instantaneous increases were statistically significant (see Table 7), the increases at the start of each group's first RT were. During the retention phase, the between-group differences in control activity were larger again than at the end of training. When comparing the first RTs of each group, the statistical analysis results in Table 8 indicate that a significant between-group difference in control input was observed in roll, as post-hoc tests indicated that the control activity of Groups 2 and 3 was significantly higher than that of Group 1. However, the between-group differences during the 6-month RTs were not statistically significant. Also, during the RTs, the same "sawtooth shape" was observed as during training. Again, the effect was more prominent in pitch than in roll, and more evident for Groups 2 and 3 compared to Group 1. The data of the 6-month RTs followed the same trend as the training phase data; in roll, all groups exhibited a clear and significant decrease in control activity, whereas in pitch only Group 1 demonstrated a statistically significant decrease (see Fig. 7 and Table 7). Similar to the  $RMS(e)$  data of Groups 1 and 2, for  $RMS(u)$  much higher learning rates were observed during the 6-month RTs compared to training.

### C. Between-Participant Variability

Figure 8 presents the retention performance of individual participants in the 6-month RT in terms of the relative change in tracking performance and control activity with respect to the end of training,  $\bar{\Delta}RMS(e)$  plotted against  $\bar{\Delta}RMS(u)$ . The left column of graphs concerns pitch performance, whereas the right one presents roll performance. Each row of graphs presents the results of a single group. The instantaneous changes in tracking error and control input, as found in the first run of the 6-month RT, are indicated with square markers with the participant numbers in them. The evolution of the tracking errors and control inputs throughout the 6-month RT are indicated with solid lines. These lines are terminated by a cross, representing the error and control RMS differentials of the last run of the RT. Finally, the gray oval area in each figure indicates the maximum variation in tracking error and control input over the last ten training runs for all participants in the group, serving as an indication of the overall end-of-training spread.

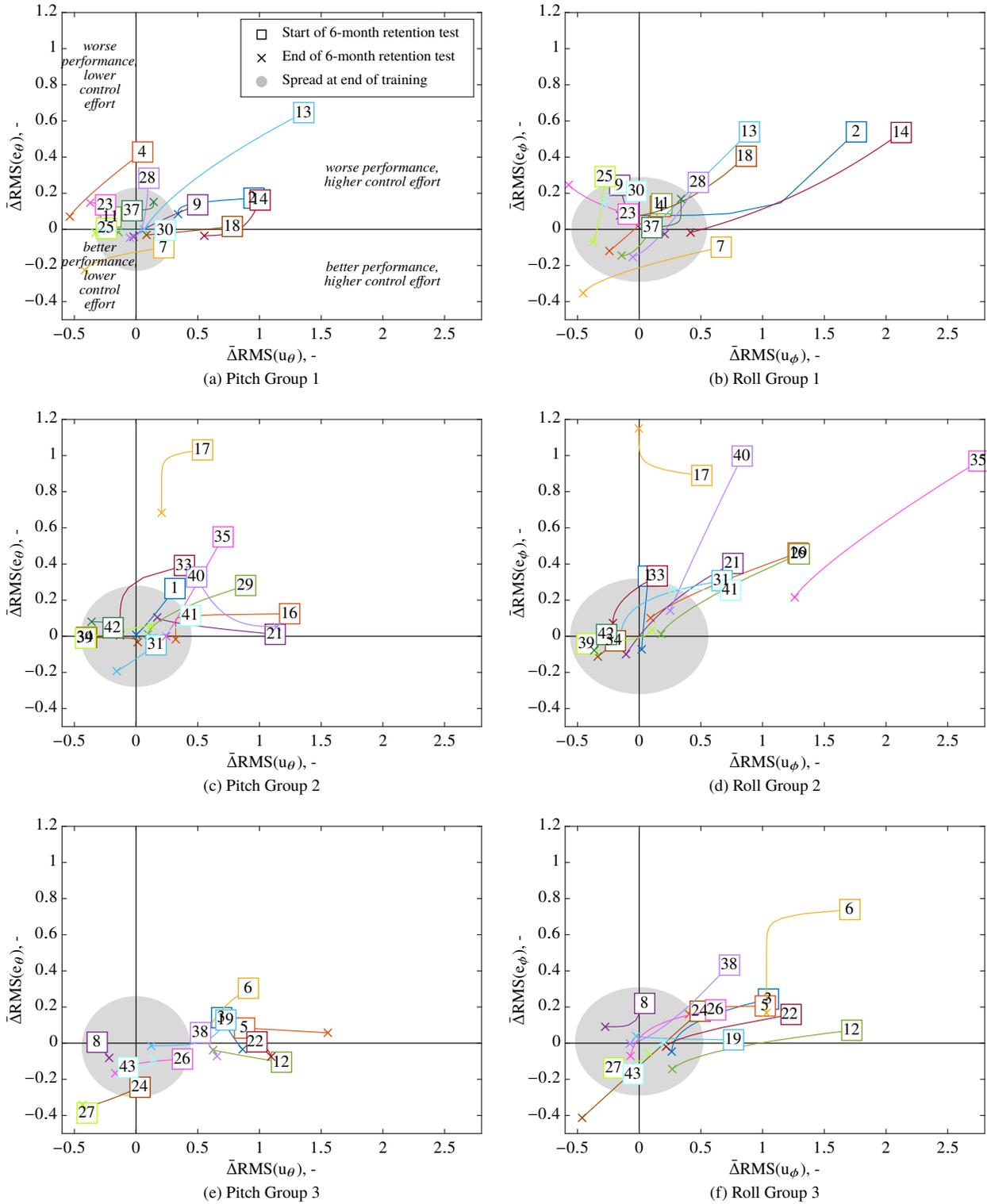
After a period of non-practice, a decrease in performance (positive  $\bar{\Delta}RMS(e)$ ) is expected. Most graphs of Fig. 8 indeed show that a notable number of squares, indicating the behavior of participants in the first run of the RT, are in the upper half of the graph. An exception to this, however, is seen in Fig. 8e, in which around half of the squares are located at the divider between the upper and lower halves or in the lower half itself, indicating that these participants of Group 3 exhibited either no change or an instantaneous improvement in pitch tracking performance at the start of the RT compared to the end of training.

At the end of the 6-month RT, the majority of participants exhibited tracking performance and control effort similar to at the end of training, as most of the crosses lie within the gray oval areas. Some exceptions to this can be found in Figs. 8a, 8b, 8e and 8f, showing that at the end of the RT participant 7 in Group 1 and participant 24 in Group 3 exhibited slightly better tracking performance and lower control activity compared to the end of training in both pitch and roll. This also holds for participant 27 in Group 3, but in this case, only for the pitch axis. On the contrary, Figs. 8c and 8d show that at the end of the 6-month RT participant 17 in Group 2 still performed considerably worse than at the end of training, meaning that after 25 runs of practice he/she had not fully regained the skills lost during the RT.

In every graph, several squares fall within the gray oval area, meaning that those participants did not show a real difference in performance and control effort between the end of training and the start of the 6-month RT. In a considerable number of these cases, the crosses also fall within these gray oval areas, illustrating that at the end of the RT these participants still performed and behaved similar to at the end of training.

It must be noted from Figs. 8a to 8f that the majority of participants started off the 6-month RT with a higher control effort than at the end of training. A possible explanation for this is that participants were motivated to perform to the best of their abilities after a few months of inactivity. Additionally, it is likely that concentration levels were high, as participants had just started. Overall, Group 3 seemed to have experienced the least degradation in manual control skills, as their square markers can be found the closest to the horizontal divider between 'worse' and 'better' performance. This was to be expected, as Group 3 had the most practice opportunities between the end of training and the 6-month RT. Participants of Group 2 had clearly experienced the largest degradation in skills, as Figs. 8c and 8d have the most square markers close to the top of the graph. Whereas most participants showed an increase in performance again during the 6-month RT (solid line going in a downward direction from square to cross), Fig. 8d shows that the tracking performance in roll of participant 17 in Group 2 even worsened during the RT, as its cross is located higher in the graph than the square marker. Although from the experiment setup it was not expected that Group 2 would perform worst in the 6-month RT, it is consistent with the tracking error results in Fig. 6.

Based on Figs. 8a to 8f, skill retention of participants could be categorized into five different groups, as depicted in Table 9. These five different groups are based on participants' tracking performance during the 6-month RT



**Fig. 8 Individual performance at 6-month retention test expressed in  $\bar{\Delta}RMS(e)$  vs.  $\bar{\Delta}RMS(u)$ .**

**Table 9 Participants' skill retention at 6-month retention test categorized based on tracking performance and control activity. (CE = control effort)**

	Group 1			Group 2			Group 3		
	Only pitch	Only roll	Both axes	Only pitch	Only roll	Both axes	Only pitch	Only roll	Both axes
<b>1. No change</b> remains within end-of-training spread	25	4, 9	11, 30, 37	31		34, 39, 42	26	27	8, 43
<b>2. Always worse</b> than at the end of training	a) Higher CE	2		40	35	17	5		6
	b) Lower CE		23						
	c) Higher → lower CE	4							
<b>3. Worse</b> (higher CE) → <b>no change</b> when compared to the end of training	9	2, 14	13, 18, 28	35	31, 40	1, 16, 21, 29, 33, 41		3, 5, 12, 22, 26, 38	19
<b>4. Worse → better</b> than at the end of training	a) Higher CE	14					3, 22, 38		
	b) Lower CE		25						
	c) Higher → lower CE							24	
<b>5. Always better</b> than at the end of training	a) Higher CE						12		
	b) Lower CE						27		
	c) Higher → lower CE			7			24		

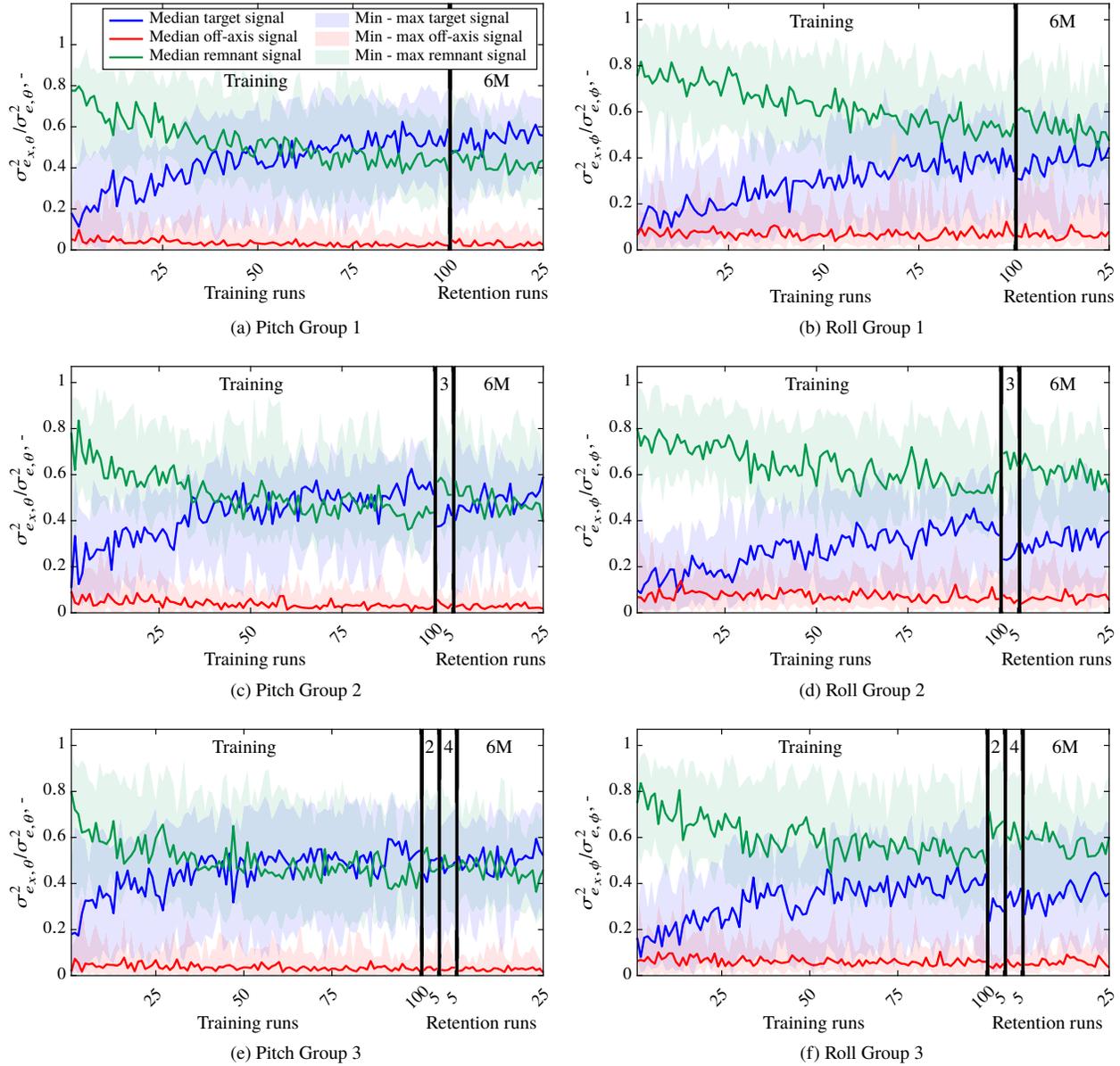
compared to their end-of-training performance ( $\Delta\text{RMS}(e)$ ). Subcategories are based on participants' control activity compared to their end-of-training control inputs ( $\Delta\text{RMS}(u)$ ). A right arrow indicates that during the first run of the RT participants' tracking performance or control activity, respectively, was at one end of its spectrum when compared to the end of training, after which it evolved and in run 25 ended at the other end of its spectrum. Table 9 indicates that Group 2 showed the most consistency between its participants, as all participants fell within only three out of eleven subcategories and 10 out of the 13 participants showed the same trend in retention performance in both pitch and roll. Group 3, on the other hand, despite showing the best performance, was the least consistent as eight out of eleven subcategories were used to classify the participants' retention behavior and only 4 out of the 12 participants exhibited the same behavior in both axes. Whereas for Group 1 eight subcategories were required to describe participants' retention behavior, this group includes eight participants who showed the same kind of retention behavior in both axes.

When comparing the group average tracking error and control input results in Figs. 6 and 7, respectively, with the individual results in Fig. 8, it is clear that notably more different variations in behavior were observed than the group average results show. Although the group average results often showed similar changes in control behavior for the pitch and roll axes, a considerable number of individuals in fact showed different changes in control behavior in both axes. Thus, the group average results actually mask these different retention effects.

#### D. Human Operator Crossfeed

To gain insight into both the amount of crossfeed present in the dual-axis tracking task performed by task-naive participants, as well as possible related training and retention effects, the decomposition of tracking error and control input variance into contributions from the target signal of the principal axis, the target signal of the other axis (i.e., crossfeed), and human operator remnant was examined, as described in Section II.H.4. These contributions are shown as a fraction of the total variance in Figs. 9 and 10 for the tracking error and control input, respectively. Again, the left column of graphs concerns the pitch axis, whereas the right one shows the roll axis results. Each row of graphs presents the results of a single group. In these figures, the contribution from the target signal of the principal axis is shown in blue, the crossfeed contribution is shown in red, and the remnant signal contribution is shown in green. Solid lines indicate the variations in the group medians over the course of the experiment, whereas transparently colored areas define the boundaries of the minimum and maximum contributions throughout the experiment.

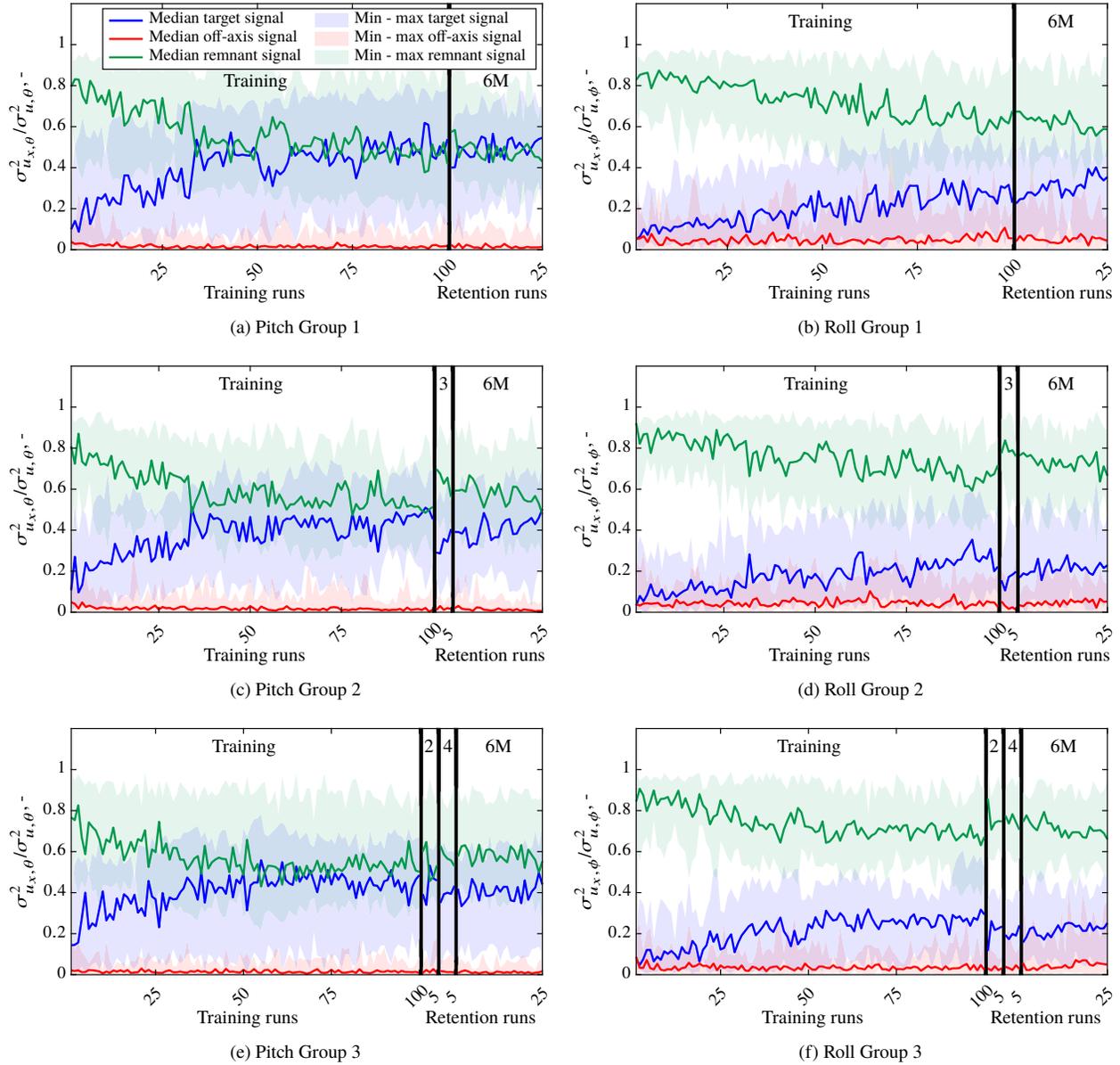
Figure 9 shows that at the start of the training phase the largest contribution to the error variance was from the human operator remnant, making up around 80% of the total variance in both pitch and roll for all three experiment groups, whereas the principal-axis target signal and the off-axis target signal (i.e., crossfeed) contributed around 15% and 5%, respectively, in the pitch axis and around 12% and 8%, respectively, in the roll axis. Throughout the training phase,



**Fig. 9** Normalized principal-axis target, off-axis target, and remnant signal contributions to tracking error variance.

the crossfeed contribution remained relatively constant, whereas the contribution from the remnant signal decreased and that from the principal-axis target signal increased. At the end of training, all groups exhibited a slightly larger contribution from the principal-axis target signal compared to the remnant signal in the pitch axis (around 52% opposed to around 43% for the principal-axis target and remnant signals, respectively). The opposite was observed in the roll axis, where at the end of training the contribution from the remnant signal remained larger than the contribution from the principal-axis target signal (around 37% opposed to around 55% for the principal-axis target and remnant signals, respectively). The values of the remnant signal contribution to the total error variance confirmed the generally accepted notion that in dual-axis tracking, the contribution of the human operator remnant signal is larger than is usually the case in single-axis compensatory tracking. For single-loop tasks, remnant signal contributions up to 40% for fully task-naive human operators and contributions around 20% for more experienced operators are generally reported [28, 41, 73].

An instantaneous increase in the remnant signal contribution was observed at the start of the retention phase, whereas the contribution of the principal-axis target signal showed an instantaneous decrease. The crossfeed contribution



**Fig. 10 Normalized principal-axis target, off-axis target, and remnant signal contributions to control input variance.**

remained similar to its contribution during the training phase. Figure 9 shows that the instantaneous increase of the remnant signal contribution at the start of the retention phase was the largest for Group 2, as was also the case for the instantaneous increase in  $RMS(e)$  (see Fig. 6). Throughout the retention phase, the contribution of the remnant signal decreased again while that of the principal-axis target signal increased.

In the decomposition of the control input variance in Fig. 10, similar observations can be made as for the tracking error variance. Early on in the training phase, the human operator remnant was responsible for the largest contribution to the total control input variance as it composed around 80% of the total variance in the pitch axis and around 85% in the roll axis. The contributions of the principal-axis target and off-axis target signals were around 16% and 4%, respectively, in the pitch axis, whereas they provided very similar contributions in the roll axis of around 7% or 8%. Throughout the training phase, the principal-axis target signal contribution increased while the remnant signal contribution decreased. Similar to the error variance, the crossfeed contribution remained relatively constant throughout training. Towards the

end of training, both the principal-axis target and remnant signals had contributions of around 48% in the pitch axis, whereas in the roll axis the remnant signal contributed around 65% and the principal-axis target signal around 28%.

Whereas at the start of the retention phase, Groups 2 and 3 showed an instantaneous increase and decrease in the contributions of the remnant and principal-axis target signals, respectively, for both pitch and roll, Group 1 did not show any noteworthy instantaneous contribution changes in roll at the start of the retention phase. In addition, the instantaneous changes of Group 1 in the pitch axis were small. A possible explanation for this marked result is that remnant is known to scale with overall control input [74] and Group 1 exhibited the smallest instantaneous change in control input of the three groups at the start of the retention phase. Similar to the observations made concerning the  $RMS(e)$  and the component contributions of the error variance, the largest changes were observed for Group 2. In the retention phase, the crossfeed contribution was again similar to its contribution in the training phase.

#### IV. Discussion

The goal of this paper was to objectively and explicitly quantify the acquisition, decay, and retention of skill-based manual control behavior. This was done by analyzing changes in performance and control behavior during training and retention testing of a compensatory dual-axis roll and pitch attitude tracking task in a human-in-the-loop experiment with 38 fully task-naïve participants. All participants were trained under the same conditions and subsequently divided into three matched groups based on their training performance and control behavior. These groups differed in their RI length and the number of RTs they performed. The first group had their performance re-evaluated after a period of non-practice of six months, whereas the second group was retested at both three and six months after training, and the third group at two, four, and six months after training. The aim of this study was threefold: (1) to determine the trend of the skill decay curve, (2) to determine the optimal RI which ensures that manual control skills do not decay significantly, while simultaneously minimizing the amount of required refresher training, and (3) to evaluate how the re-acquisition rate of manual control skills during retention testing compares to their initial acquisition rate. To objectively evaluate and quantify changes in manual control behavior, exponential learning curve models were fitted to metrics of tracking performance and control activity. In addition, to contribute to the understanding of crossfeed in multi-axis manual control, the presence of crossfeed and its training and retention effects were explicitly investigated.

Based on the findings of previous training experiments [41, 46, 66], participants were expected to show considerable skill development during the training phase of the experiment (**hypothesis 1**). Throughout training, tracking performance indeed improved significantly (i.e., lower  $RMS(e)$ ) for all three groups in both pitch and roll.

In accordance with previous dual-axis tracking task experiments [28, 31, 38, 40, 42], better tracking performance was anticipated in pitch compared to roll (**hypothesis 2**). Tracking errors were indeed consistently lower in pitch compared to roll during the training phase and higher learning rates were observed for pitch as well. The difference in performance between pitch and roll might be caused by a decreased visibility of errors in roll due to a lower pixel resolution in roll compared to pitch. Similar performance differences between the pitch and roll axes were observed in the retention phase.

As the aim of the current experiment was to model the decay curve of skill-based human operator control behavior, it was expected that skill decay could be captured by a positively accelerating decay curve, with fairly little skill loss right after training and a higher rate of decay later on (**hypothesis 3**). Such a decay curve was found in previous experiments with flying tasks comparable to the current tracking task [67, 68]. The first RT of each of the groups showed that all groups exhibited significantly degraded tracking performance compared to the end of training in both pitch and roll. However, no significant differences were observed between these first RTs (the 6-month RT of Group 1, the 3-month RT of Group 2 and the 2-month RT of Group 3). These findings actually suggest the opposite of the hypothesis, i.e., that skill-based manual control behavior in fact decays according to a negatively accelerating decay curve, as tracking performance degraded quickly the first few months after training, after which the degradation slowed down and performance stabilized at a new, lower level than end-of-training performance. Also, the retention results showed that the rate of skill decay was influenced by the rate of learning during training. Task performance deteriorated more in roll than in pitch due to the lower learning rate in the roll axis during training, resulting in less additional practice beyond the point of asymptotic performance compared to the pitch axis. The observed difference in skill decay between the pitch and roll axes is consistent with previous retention research which identified the level of original learning as one of the most important factors for skill retention [12, 14, 61–63]; overlearning is well-known to enhance retention [16, 19, 64, 65].

For the 6-month RTs of all groups, the best performance was expected to be shown by Group 3, whereas the worst performance was foreseen for Group 1 (**hypothesis 4**), based on the finding from earlier retention experiments that

operators perform better during retention testing if they have received practice opportunities of some form [67, 69–71]. In this case, the 2-month, 3-month, and 4-month RTs of Groups 2 and 3 can be regarded as additional practice moments scheduled during a 6-month RI. Surprisingly, the worst performance during the 6-month RT was shown by Group 2, whereas Groups 1 and 3 exhibited very similar performance. However, statistical analysis results revealed that the performance of none of the groups was significantly different from that of the other groups, due to which no solid conclusions could be drawn on the best and the worst performing group during the final RT, and the hypothesis could neither be accepted, nor rejected.

Based on previous retention studies [12, 63, 72], skill re-acquisition during the retention phase was expected to occur at a higher rate than initial skill development (**hypothesis 5**). Learning curve fitting indeed showed that the 6-month RT learning curves of Groups 1 and 2 had higher learning rates than the learning curves fitted to their training phase data. The 6-month RT learning curve of Group 3, however, had lower learning rates than the group's initial development learning curves due to the group's relatively good performance immediately from the start of the 6-month RT, which could be explained by the fact that the group had already had two practice opportunities in the past six months. Therefore, based on the learning curve parameters of Groups 1 and 2, the results, in accordance with the hypothesis, still suggest that relearning occurs at a higher rate than initial development.

The experimental results presented in this paper suggest that manual control skills decay following a negatively accelerating decay curve. However, the remarkable good performance of Group 1 and the comparatively worse performance of Group 2 in the 6-month RT suggest that a larger-scale experiment should be conducted to reliably and quantitatively model the curve of skill decay. A larger number of participants per experiment group is required to reduce the influence of single individuals with extraordinarily good or poor performance and to increase statistical power. At least twice as many participants per group as in the current experiment would be preferred, as with a sample size larger than 25 the distribution of the sample means tend to be normally distributed around the population mean [75]. Additionally, the number of experiment groups should be increased to be able to evaluate RIs ranging from a few hours or days up to several years and to capture (almost) the complete curve of skill decay.

Although the groups' performance during the 6-month RT did not give an indication of what the 'optimal' RI is while at the same time minimizing the amount of refresher training, the 2-month RT of Group 3 and 3-month RT of Group 2 did give a slight indication. When 'refresher' training was provided to Group 3 after two months, five 'refresher' runs were sufficient to decrease  $RMS(e)$  again to end-of-training values, whereas when 'refresher' training was given to Group 2 after three months, performance in both pitch and roll was still significantly different from end-of-training values after an equal number of five tracking runs. This somewhat suggests that refresher training can better be provided in the form of shorter refresher training sessions on a more regular basis. Although over a longer period of time this might lead to the same total amount of refresher training as when longer refresher training sessions are provided on a larger between-session interval basis, the likelihood of skill decay occurrence might be smaller due to the shorter RIs.

The presence of crossfeed and its training and retention effects were investigated by examining the decomposition of the tracking error and control input variance into contributions from the target signal of the principal axis, the target signal of the other axis (i.e., crossfeed) and human operator remnant. The results showed that the identified crossfeed could explain up to 8% of the measured control inputs. However, no training or retention effects were observed, as the relative crossfeed contribution to the total tracking error and control input variances remained relatively constant over the course of the experiment. In an earlier study [28] it was found that in a similar task the identified crossfeed contributed up to 20% to the total human operator control response. These differences demonstrate that the presence of crossfeed noticeably depends on the manipulator. Whereas the previous study [28] was performed in the SIMONA Research Simulator (SRS) at the TU Delft, the current experiment was completed in the HMILab at the TU Delft. The position of the sidestick in the HMILab allows for the participant's arm to be placed more in line with the sidestick's pitch axis, opposed to in the SRS, in which due to the position of the sidestick the operator's arm is not aligned with either the pitch or roll axis. The location of the sidestick in the simulator setup in the HMILab resulted in a lower likelihood of introducing cross-coupled control inputs compared to in the previous experiment.

To further contribute to the understanding of skill decay and retention, in future work we aim to quantify the changes in human control behavior over the course of this training and retention experiment using quasi-linear human operator models obtained for each individual tracking run. The use of this *cybernetic approach* [66] will provide a unique insight into the development of manual control skills during training and retention testing by modeling operators' control behavior in terms of distinct contributions that are physically interpretable [32], which in turn will enable a systematic optimization of manual control skill training [22]. Additionally, by objectively quantifying skill

development, decay, and retention using this cybernetic approach, our future work will attempt to solve the issue of the use of different performance measures which made the results of earlier retention studies difficult to compare.

Finally, it should be noted that the findings of the current experiment apply only to skill-based control behavior, as it is generally believed that different types of skills and behavior decay in different ways [19]. Although evaluating the retention of manual control skills in more complex, realistic flying tasks is of utmost importance to be able to design training procedures such that pilot flying proficiency can be developed, maintained, and improved, the control task in the current experiment was chosen based on the fact that no quantitative analysis techniques exist yet for evaluating manual control behavior in more complex tasks, requiring the higher levels of control behavior known as rule-based and knowledge-based behavior [76]. The advancement to human operator modeling techniques suitable for identification of the higher levels of control behavior and thus for human operator control behavior identification in more complex and realistic flying tasks is, however, essential for the advancements in skill decay and retention research and the determination of optimal (recurrent) training procedures.

## V. Conclusions

The goal of this paper was to objectively evaluate and quantify the acquisition, decay, and retention of skill-based manual control behavior in a training and retention experiment with 38 fully task-naïve participants. Participants were trained in a compensatory dual-axis roll and pitch attitude tracking task and were divided into three groups, differing in their RI length and the number of RTs they performed. The performance of the first group was re-evaluated after a period of non-practice of six months, whereas the second group performed the task again at both three and six months after training, and the third group was retested after two, four, and six months. To explicitly quantify skill development, decay, and retention, exponential learning curve models were fitted to metrics of task performance and control activity.

The results of the experiment suggest that manual control skills decay following a negatively accelerating decay curve. The rate of decay is influenced by the learning rate during training. Performance in roll decayed more than performance in pitch due to participants' higher learning rates and earlier stabilization in pitch during the training phase. In addition, fitted learning curve models to metrics of task performance and control activity showed that during the retention tests the lost skills were re-acquired at a higher rate than their initial development rate. Finally, the comparison between the different experiment groups with different RIs suggests that short refresher training sessions can be effective for limiting skill decay over periods of inactivity. However, to be able to implement scientifically substantiated standards to ensure pilots receive sufficient training opportunities to develop, maintain, and improve flying proficiency as well as to be able to design (e.g., onboard refresher) training to prevent skill decay during long-duration deep space operations, a larger-scale experiment should be conducted with at least twice the number of participants per experiment group as in the current experiment and with periods of non-practice ranging from a few hours or days up to several years to be able to capture the complete skill decay curve.

## Acknowledgments

The authors would like to thank the many participants who volunteered to perform this extensive simulator experiment. In addition, they would like to express their gratitude to prof.dr.ir. Max Mulder and dr.ir. Marinus (René) M. van Paassen from the Section Control & Simulation, Faculty of Aerospace Engineering, TU Delft for their highly useful insights and comments throughout all phases of this research.

## References

- [1] Veillette, P. R., "Differences in Aircrew Manual Skills in Automated and Conventional Flight Decks," *Transportation Research Record*, No. 1480, 1995, pp. 43–50.
- [2] Nakamura, D., Abbott, K. H., McKenney, D., Railsback, P., et al., "Operational Use of Flight Path Management Systems," Performance-Based Operations Aviation Rulemaking Committee/Commercial Aviation Safety Team Flight Deck Automation Working Group, Federal Aviation Administration, Washington, DC, Sept. 2013. [https://www.faa.gov/aircraft/air\\_cert/design\\_approvals/human\\_factors/media/OUFPMS\\_Report.pdf](https://www.faa.gov/aircraft/air_cert/design_approvals/human_factors/media/OUFPMS_Report.pdf).
- [3] Anonymous, "Office of Inspector General Audit Report: Enhanced FAA Oversight Could Reduce Hazards Associated With Increased Use of Flight Deck Automation," United States Department of Transportation, Audit Rept. AV-2016-013, Washington, DC, Jan. 2016. [https://www.oig.dot.gov/sites/default/files/FAA%20Flight%20Deck%20Automation\\_Final%20Report%5E1-7-16.pdf](https://www.oig.dot.gov/sites/default/files/FAA%20Flight%20Deck%20Automation_Final%20Report%5E1-7-16.pdf).

- [4] Gillen, M. W., "Diminishing Skills?" *AeroSafety World*, Vol. 5, No. 6, 2010, pp. 30–34.
- [5] Anonymous, "SAFO 13002: Manual Flight Operations," Federal Aviation Administration, Washington, DC, Jan. 2013. [https://www.faa.gov/other\\_visit/aviation\\_industry/airline\\_operators/airline\\_safety/safo/all\\_safos/media/2013/SAFO13002.pdf](https://www.faa.gov/other_visit/aviation_industry/airline_operators/airline_safety/safo/all_safos/media/2013/SAFO13002.pdf).
- [6] Casner, S. M., Geven, R. W., and Williams, K. T., "The Effectiveness of Airline Pilot Training for Abnormal Events," *Human Factors*, Vol. 55, No. 3, 2013, pp. 477–485. <https://doi.org/10.1177/0018720812466893>.
- [7] Casner, S. M., Geven, R. W., Recker, M. P., and Schooler, J. W., "The Retention of Manual Flying Skills in the Automated Cockpit," *Human Factors*, Vol. 56, No. 8, 2014, pp. 1506–1516. <https://doi.org/10.1177/0018720814535628>.
- [8] Anonymous, "SAFO 17007: Manual Flight Operations Proficiency," Federal Aviation Administration, Washington, DC, May 2017. [https://www.faa.gov/other\\_visit/aviation\\_industry/airline\\_operators/airline\\_safety/safo/all\\_safos/media/2017/SAFO17007.pdf](https://www.faa.gov/other_visit/aviation_industry/airline_operators/airline_safety/safo/all_safos/media/2017/SAFO17007.pdf).
- [9] Barshi, I., and Dempsey, D. L., "Evidence Report: Risk of Performance Errors Due to Training Deficiencies," Lyndon B. Johnson Space Center, National Aeronautics and Space Administration, Houston, TX, Apr. 2016. <https://humanresearchroadmap.nasa.gov/Evidence/reports/TRAIN.pdf>.
- [10] Mars Architecture Steering Group, "Human Exploration of Mars: Design Reference Architecture 5.0," National Aeronautics and Space Administration Headquarters, Technical Rept. NASA/SP-2009-566, Washington, DC, July 2009. [https://www.nasa.gov/pdf/373665main\\_NASA-SP-2009-566.pdf](https://www.nasa.gov/pdf/373665main_NASA-SP-2009-566.pdf).
- [11] Mengelkoch, R. F., Adams, J. A., and Gainer, C. A., "The Forgetting of Instrument Flying Skills as a Function of the Level of Initial Proficiency," US Naval Training Center, Technical Rept. NAVTRADEVCCEN TR-71-16-18, Port Washington, NY, Sept. 1958.
- [12] Mengelkoch, R. F., Adams, J. A., and Gainer, C. A., "The Forgetting of Instrument Flying Skills," *Human Factors*, Vol. 13, No. 5, 1971, pp. 397–405. <https://doi.org/10.1177/001872087101300502>.
- [13] Wright, R. H., "Review of Behavioral Science Research Data Relevant to Army Proficiency Flying Programs," HumRRO Consulting Report, Human Resources Research Organization, Division No. 6 (Aviation), Fort Rucker, AL, Apr. 1969.
- [14] Wright, R. H., "Retention of Flying Skills and Refresher Training Requirements: Effects of Non-Flying and Proficiency Flying," Human Resources Research Organization, Technical Rept. HumRRO-TR-73-32, Alexandria, VA, Dec. 1973. <https://apps.dtic.mil/dtic/tr/fulltext/u2/774853.pdf>.
- [15] Armstrong, M. B., Bleymaier, J. S., Hinkel, L. F., Levins, R., and Sheppard, R. R., "Flying Skill Retention and Proficiency Flying," Air Command and Staff College, Air University, Maxwell Air Force Base, Research Rept. No. 0095-75, Montgomery, AL, May 1975.
- [16] Smith, J. F., and Matheny, W. G., "Continuation Versus Recurrent Pilot Training," Air Force Human Resources Laboratory, Brooks Air Force Base, Technical Rept. AFHRL-TR-76-4, San Antonio, TX, May 1976. <https://files.eric.ed.gov/fulltext/ED126346.pdf>.
- [17] Prophet, W. W., "Long-Term Retention of Flying Skills: A Review of the Literature," Human Resources Research Organization, Final Rept. HumRRO-FR-ED(P)-76-35, Alexandria, VA, Oct. 1976. <https://apps.dtic.mil/dtic/tr/fulltext/u2/a036077.pdf>.
- [18] Prophet, W. W., "Long-Term Retention of Flying Skills: An Annotated Bibliography," Human Resources Research Organization, Final Rept. HumRRO-FR-ED(P)-76-36, Alexandria, VA, Oct. 1976. <https://apps.dtic.mil/dtic/tr/fulltext/u2/a036114.pdf>.
- [19] Naylor, J. C., and Briggs, G. E., "Long-Term Retention of Learned Skills: A Review of the Literature," Aerospace Medical Laboratory, Wright-Patterson Air Force Base, ASD Technical Rept. 61-390, Canton, OH, Aug. 1961.
- [20] Smode, A. F., Hall, E. R., and Meyer, D. E., "An Assessment of Research Relevant to Pilot Training," Aerospace Medical Research Laboratories, Wright-Patterson Air Force Base, Technical Rept. AMRL-TR-66-196, Dayton, OH, Nov. 1966. <https://apps.dtic.mil/dtic/tr/fulltext/u2/804600.pdf>.
- [21] Naylor, J. C., Briggs, G. E., and Reed, W. G., "Task Coherence, Training Time, and Retention Interval Effects on Skill Retention," *Journal of Applied Psychology*, Vol. 52, No. 5, 1968, pp. 386–393. <https://doi.org/10.1037/h0026239>.
- [22] Mulder, M., Pool, D. M., Abbink, D. A., Boer, E. R., Zaal, P. M. T., Drop, F. M., Van der El, K., and Van Paassen, M. M., "Manual Control Cybernetics: State-of-the-Art and Current Trends," *IEEE Transactions on Human-Machine Systems*, Vol. 48, No. 5, 2018, pp. 468–485. <https://doi.org/10.1109/THMS.2017.2761342>.

- [23] Mulder, M., Pool, D. M., Abbink, D. A., Boer, E. R., and Van Paassen, M. M., "Fundamental Issues in Manual Control Cybernetics," *Proceedings of the 13th IFAC/IFIP/IFORS/IEA Symposium on Analysis, Design, and Evaluation of Human-Machine Systems, Kyoto, Japan*, Aug.-Sept. 2016. <https://doi.org/10.1016/j.ifacol.2016.10.429>.
- [24] Bekey, G. A., Meissinger, H. F., and Rose, R. E., "Mathematical Models of Human Operators in Simple Two-Axis Manual Control Systems," *IEEE Transactions on Human Factors in Electronics*, Vol. HFE-6, No. 1, 1965, pp. 42–52. <https://doi.org/10.1109/THFE.1965.6591255>.
- [25] Todosiev, E. P., Rose, R. E., and Summers, L. G., "Human Performance in Single and Two-Axis Tracking Systems," *IEEE Transactions on Human Factors in Electronics*, Vol. HFE-8, No. 2, 1967, pp. 125–129. <https://doi.org/10.1109/THFE.1967.233629>.
- [26] Levison, W. H., Elkind, J. I., and Ward, J. L., "Studies of Multivariable Manual Control Systems: A Model for Task Interference," National Aeronautics and Space Administration, Technical Rept. NASA CR-1746, Washington, DC, May 1971. <https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19710016684.pdf>.
- [27] Van Lunteren, A., "Identification of Human Operator Describing Function Models with One or Two Inputs in Closed Loop Systems," unpublished Ph.D. Thesis, Faculty of Mechanical Engineering, Delft University of Technology, Delft, The Netherlands, Mar. 1979. <https://repository.tudelft.nl/islandora/object/uuid%3A1466fac5-bc32-48f7-bf52-3d508d1d95f6>.
- [28] Barendswaard, S., Pool, D. M., Van Paassen, M. M., and Mulder, M., "Dual-Axis Manual Control: Performance Degradation, Axis Asymmetry, Crossfeed, and Intermittency," *IEEE Transactions on Human-Machine Systems*, Vol. 49, No. 2, 2019, pp. 113–125. <https://doi.org/10.1109/THMS.2019.2890856>.
- [29] Barendswaard, S., Pool, D. M., and Mulder, M., "Human Crossfeed in Dual-Axis Manual Control with Motion Feedback," *Proceedings of the 13th IFAC/IFIP/IFORS/IEA Symposium on Analysis, Design, and Evaluation of Human-Machine Systems, Kyoto, Japan*, Vol. 49, No. 19, 2016, pp. 189–194. <https://doi.org/10.1016/j.ifacol.2016.10.514>.
- [30] McRuer, D. T., and Jex, H. R., "A Review of Quasi-Linear Pilot Models," *IEEE Transactions on Human Factors in Electronics*, Vol. HFE-8, No. 3, 1967, pp. 231–249. <https://doi.org/10.1109/THFE.1967.234304>.
- [31] Zaal, P. M. T., and Mobertz, X. R. I., "Effects of Motion Cues on the Training of Multi-Axis Manual Control Skills," *Proceedings of the AIAA Modeling and Simulation Technologies Conference*, Denver, CO, AIAA Paper 2017-3473, June 2017. <https://doi.org/10.2514/6.2017-3473>.
- [32] McRuer, D., Graham, D., Krendel, E., and Reisener, W., Jr., "Human Pilot Dynamics in Compensatory Systems: Theory, Models, and Experiments with Controlled Element and Forcing Function Variations," Air Force Flight Dynamics Laboratory, Wright-Patterson Air Force Base, Technical Rept. AFFDL-TR-65-15, Dayton, OH, July 1965. <https://apps.dtic.mil/dtic/tr/fulltext/u2/470337.pdf>.
- [33] Stapleford, R. L., McRuer, D. T., and Magdaleno, R. E., "Pilot Describing Function Measurements in a Multiloop Task," *IEEE Transactions on Human Factors in Electronics*, Vol. HFE-8, No. 2, 1967, pp. 113–125. <https://doi.org/10.1109/THFE.1967.233628>.
- [34] Van Paassen, M. M., and Mulder, M., "Identification of Human Operator Control Behavior in Multiple-Loop Tracking Tasks," *Proceedings of the 7th IFAC/IFIP/IFORS/IEA Symposium on Analysis, Design and Evaluation of Man-Machine Systems, Kyoto, Japan*, Vol. 31, No. 26, Sept. 1998, pp. 455–460. [https://doi.org/10.1016/S1474-6670\(17\)40135-2](https://doi.org/10.1016/S1474-6670(17)40135-2).
- [35] Zaal, P. M. T., Pool, D. M., De Bruin, J., Mulder, M., and Van Paassen, M. M., "Use of Pitch and Heave Motion Cues in a Pitch Control Task," *Journal of Guidance, Control, and Dynamics*, Vol. 32, No. 2, 2009, pp. 366–377. <https://doi.org/10.2514/1.39953>.
- [36] Pool, D. M., Zaal, P. M. T., Van Paassen, M. M., and Mulder, M., "Effects of Heave Washout Settings in Aircraft Pitch Disturbance Rejection," *Journal of Guidance, Control, and Dynamics*, Vol. 33, No. 1, 2010, pp. 29–41. <https://doi.org/10.2514/1.46351>.
- [37] Zaal, P. M. T., Pool, D. M., Van Paassen, M. M., and Mulder, M., "Comparing Multimodal Pilot Pitch Control Behavior Between Simulated and Real Flight," *Journal of Guidance, Control, and Dynamics*, Vol. 35, No. 5, 2012, pp. 1456–1471. <https://doi.org/10.2514/1.56268>.
- [38] Zaal, P. M. T., and Sweet, B. T., "Identification of Time-Varying Pilot Control Behavior in Multi-Axis Control Tasks," *Proceedings of the AIAA Modeling and Simulation Technologies Conference*, Minneapolis, MN, AIAA Paper 2012-4793, Aug. 2012. <https://doi.org/10.2514/6.2012-4793>.

- [39] Nieuwenhuizen, F. M., Mulder, M., Van Paassen, M. M., and Bühlhoff, H. H., "Influences of Simulator Motion System Characteristics on Pilot Control Behavior," *Journal of Guidance, Control, and Dynamics*, Vol. 36, No. 3, 2013, pp. 667–676. <https://doi.org/10.2514/1.59257>.
- [40] Zaal, P. M. T., and Pool, D. M., "Multimodal Pilot Behavior in Multi-Axis Tracking Tasks with Time-Varying Motion Cueing Gains," *Proceedings of the AIAA Modeling and Simulation Technologies Conference*, National Harbor, MD, AIAA Paper 2014-0810, Jan. 2014. <https://doi.org/10.2514/6.2014-0810>.
- [41] Pool, D. M., Harder, G. A., and Van Paassen, M. M., "Effects of Simulator Motion Feedback on Training of Skill-Based Control Behavior," *Journal of Guidance, Control, and Dynamics*, Vol. 39, No. 4, 2016, pp. 889–901. <https://doi.org/10.2514/1.G001603>.
- [42] Zaal, P. M. T., "Manual Control Adaptation to Changing Vehicle Dynamics in Roll-Pitch Control Tasks," *Journal of Guidance, Control, and Dynamics*, Vol. 39, No. 5, 2016, pp. 1046–1058. <https://doi.org/10.2514/1.G001592>.
- [43] Damveld, H. J., "A Cybernetic Approach to Assess the Longitudinal Handling Qualities of Aeroelastic Aircraft," Ph.D. Thesis, Faculty of Aerospace Engineering, Delft University of Technology, Delft, The Netherlands, May 2009. <http://resolver.tudelft.nl/uuid:3869d074-4e3c-4738-b445-15dbdfd51cab>.
- [44] Van Paassen, M. M., and Mulder, M., "Identification of Human Control Behavior," *International Encyclopedia of Ergonomics and Human Factors*, 2nd ed., edited by W. Karwowski, Taylor & Francis, London, England, 2006, pp. 400–407.
- [45] Drop, F. M., De Vries, R., Mulder, M., and Bühlhoff, H. H., "The Predictability of a Target Signal Affects Manual Feedforward Control," *Proceedings of the 13th IFAC/IFIP/IFORS/IEA Symposium on Analysis, Design, and Evaluation of Human-Machine Systems, Kyoto, Japan*, Aug.-Sept. 2016. <https://doi.org/10.1016/j.ifacol.2016.10.482>.
- [46] Mendes, M. F. S., Pool, D. M., and Van Paassen, M. M., "Effects of Peripheral Visual Cues in Simulator-Based Training of Multimodal Control Skills," *Proceedings of the AIAA Modeling and Simulation Technologies Conference*, Denver, CO, AIAA Paper 2017-3671, June 2017. <https://doi.org/10.2514/6.2017-3671>.
- [47] Beerens, G. C., Damveld, H. J., Mulder, M., Van Paassen, M. M., and Van der Vaart, J. C., "Investigation into Crossover Regression in Compensatory Manual Tracking Tasks," *Journal of Guidance, Control, and Dynamics*, Vol. 32, No. 5, 2009, pp. 1429–1445. <https://doi.org/10.2514/1.43528>.
- [48] Damveld, H. J., Beerens, G. C., Van Paassen, M. M., and Mulder, M., "Design of Forcing Functions for the Identification of Human Control Behavior," *Journal of Guidance, Control, and Dynamics*, Vol. 33, No. 4, 2010, pp. 1064–1081. <https://doi.org/10.2514/1.47730>.
- [49] Robertson, E. M., Pascual-Leone, A., and Press, D. Z., "Awareness Modifies the Skill-Learning Benefits of Sleep," *Current Biology*, Vol. 14, No. 3, 2014, pp. 208–212. <https://doi.org/10.1016/j.cub.2004.01.027>.
- [50] Kantak, S. S., and Winstein, C. J., "Learning-Performance Distinction and Memory Processes for Motor Skills: A Focused Review and Perspective," *Behavioural Brain Research*, Vol. 228, No. 1, 2012, pp. 219–231. <https://doi.org/10.1016/j.bbr.2011.11.028>.
- [51] Reynolds, B., and Bilodeau, I. M., "Acquisition and Retention of Three Psychomotor Tests as a Function of Distribution of Practice during Acquisition," *Journal of Experimental Psychology*, Vol. 44, No. 1, 1952, pp. 19–26. <https://doi.org/10.1037/h0058929>.
- [52] Adams, J. A., and Reynolds, B., "Effect of Shift in Distribution of Practice Conditions Following Interpolated Rest," *Journal of Experimental Psychology*, Vol. 47, No. 1, 1954, pp. 32–36. <https://doi.org/10.1037/h0058919>.
- [53] Lewis, D., and Lowe, W. F., "Retention of Skill on the SAM Complex Coordinator," *Proceedings of the Iowa Academy of Science*, Cedar Falls, IA, Vol. 63, No. 1, 1956, pp. 591–599.
- [54] Holding, D. H., *Principles of Training*, 1st ed., Pergamon Press, Oxford, England, 1965, pp. 92–94.
- [55] Schmidt, R. A., *Motor Skills*, Harper & Row, New York, NY, 1975.
- [56] Singer, R. N., *Motor Learning and Human Performance: Application in Physical Education Skills*, 2nd ed., Macmillan Publishers, London, England, 1975.
- [57] Levison, W. H., Lancraft, R. E., and Junker, A. M., "Effects of Simulator Delays on Performance and Learning in a Roll-Axis Tracking Task," *Proceedings of the Fifteenth Annual Conference on Manual Control*, Dayton, OH, Mar. 1979, pp. 168–186.

- [58] Pool, D. M., and Zaal, P. M. T., “Between-Subject Variability in Transfer-of-Training of Skill-Based Manual Control Behavior,” *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics, Kowloon Tong, Hong Kong*, Oct. 2015, pp. 1094–1099. <https://doi.org/10.1109/SMC.2015.196>.
- [59] Jex, H. R., Magdaleno, R. E., and Junker, A. M., “Roll Tracking Effects of G-Vector Tilt and Various Types of Motion Washout,” *Proceedings of the 14th Annual Conference on Manual Control, Los Angeles, CA*, Apr. 1978, pp. 463–502. <https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19790007451.pdf>.
- [60] Vineberg, R., “A Study of the Retention of Skills and Knowledge Acquired in Basic Training,” Human Resources Research Organization, Technical Rept. HumRRO-TR-75-10, Alexandria, VA, June 1975. <https://apps.dtic.mil/dtic/tr/fulltext/u2/a012678.pdf>.
- [61] Fleishman, E. A., and Parker, J. F., Jr., “Factors in the Retention and Relearning of Perceptual-Motor Skill,” *Journal of Experimental Psychology*, Vol. 64, No. 3, 1962, pp. 215–226. <https://doi.org/10.1037/h0041220>.
- [62] Gardlin, G. R., and Sitterley, T. E., “Degradation of Learned Skills: A Review and Annotated Bibliography,” The Boeing Company, Technical Rept. D180-15080-1, Seattle, WA, June 1972. <https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19730001425.pdf>.
- [63] Ammons, R. B., Farr, R. G., Bloch, E., Neumann, E., Dey, M., Marion, R., and Ammons, C. H., “Long-Term Retention of Perceptual-Motor Skills,” *Journal of Experimental Psychology*, Vol. 55, No. 4, 1958, pp. 318–328. <https://doi.org/10.1037/h0041893>.
- [64] Hammerton, M., “Retention of Learning in a Difficult Tracking Task,” *Journal of Experimental Psychology*, Vol. 66, No. 1, 1963, pp. 108–110. <https://doi.org/10.1037/h0040296>.
- [65] Melnick, M. J., “Effects of Overlearning on the Retention of a Gross Motor Skill,” *Research Quarterly*, Vol. 42, No. 1, 1971, pp. 60–69. <https://doi.org/10.1080/10671188.1971.10615036>.
- [66] Pool, D. M., and Zaal, P. M. T., “A Cybernetic Approach to Assess the Training of Manual Control Skills,” *Proceedings of the 13th IFAC/IFIP/IFORS/IEA Symposium on Analysis, Design, and Evaluation of Human-Machine Systems, Kyoto, Japan*, Aug.-Sept. 2016. <https://doi.org/10.1016/j.ifacol.2016.10.588>.
- [67] Sitterley, T. E., and Berge, W. A., “Degradation of Learned Skills: Effectiveness of Practice Methods on Simulated Space Flight Skill Retention,” The Boeing Company, Technical Rept. D180-15081-1, Seattle, WA, July 1972. <https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19730001426.pdf>.
- [68] Ruffner, J., Wick, W., and Bickley, W., “Retention of Helicopter Flight Skills: Is There a ‘Critical Period’ for Proficiency Loss?” *Proceedings of the Human Factors Society 28th Annual Meeting, San Antonio, TX*, Vol. 28, No. 4, 1984, pp. 370–374.
- [69] Sitterley, T. E., Zaitzeff, L. P., and Berge, W. A., “Degradation of Learned Skills: Effectiveness of Practice Methods on Visual Approach and Landing Skill Retention,” The Boeing Company, Technical Rept. D180-15082-1, Seattle, WA, Oct. 1972. <https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19730014359.pdf>.
- [70] Sitterley, T. E., “Degradation of Learned Skills: Static Practice Effectiveness for Visual Approach and Landing Skill Retention,” The Boeing Aerospace Company, Technical Rept. D180-17876-1, Seattle, WA, May 1974. <https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19740024447.pdf>.
- [71] Leonard, R. L., Jr., Wheaton, G. R., and Cohen, F. P., “Transfer of Training and Skill Retention,” Unit Training and Educational Technology System Branch, American Institutes for Research, Technical Rept. TR-76-A3, Washington, DC, Oct. 1976. <https://apps.dtic.mil/dtic/tr/fulltext/u2/a036059.pdf>.
- [72] Hill, D. S., “Minor Studies in Learning and Relearning,” *Journal of Educational Psychology*, Vol. 5, No. 7, 1914, pp. 375–386. <https://doi.org/10.1037/h0073871>.
- [73] Van der El, K., Morais Almeida, J., Pool, D. M., Van Paassen, M. M., and Mulder, M., “The Effects of Motion Feedback in Manual Preview Tracking Tasks,” *Proceedings of the AIAA Modeling and Simulation Technologies Conference*, Denver, CO, AIAA Paper 2017-3472, June 2017. <https://doi.org/10.2514/6.2017-3472>.
- [74] Levison, W. H., Baron, S., and Kleinman D. L., “A Model for Human Controller Remnant,” *IEEE Transactions on Man-Machine Systems*, Vol. 10, No. 4, 1969, pp. 101–108. <https://doi.org/10.1109/TMMS.1969.299906>.
- [75] Hinton, P. R., *Statistics Explained: A Guide for Social Science Students*, 2nd ed., Taylor & Francis, London, England, 2004.
- [76] Rasmussen, J., “Skills, Rules, and Knowledge; Signals, Signs, and Symbols, and Other Distinctions in Human Performance Models,” *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. SMC-13, No. 3, 1983, pp. 257–266. <https://doi.org/10.1109/TSMC.1983.6313160>.