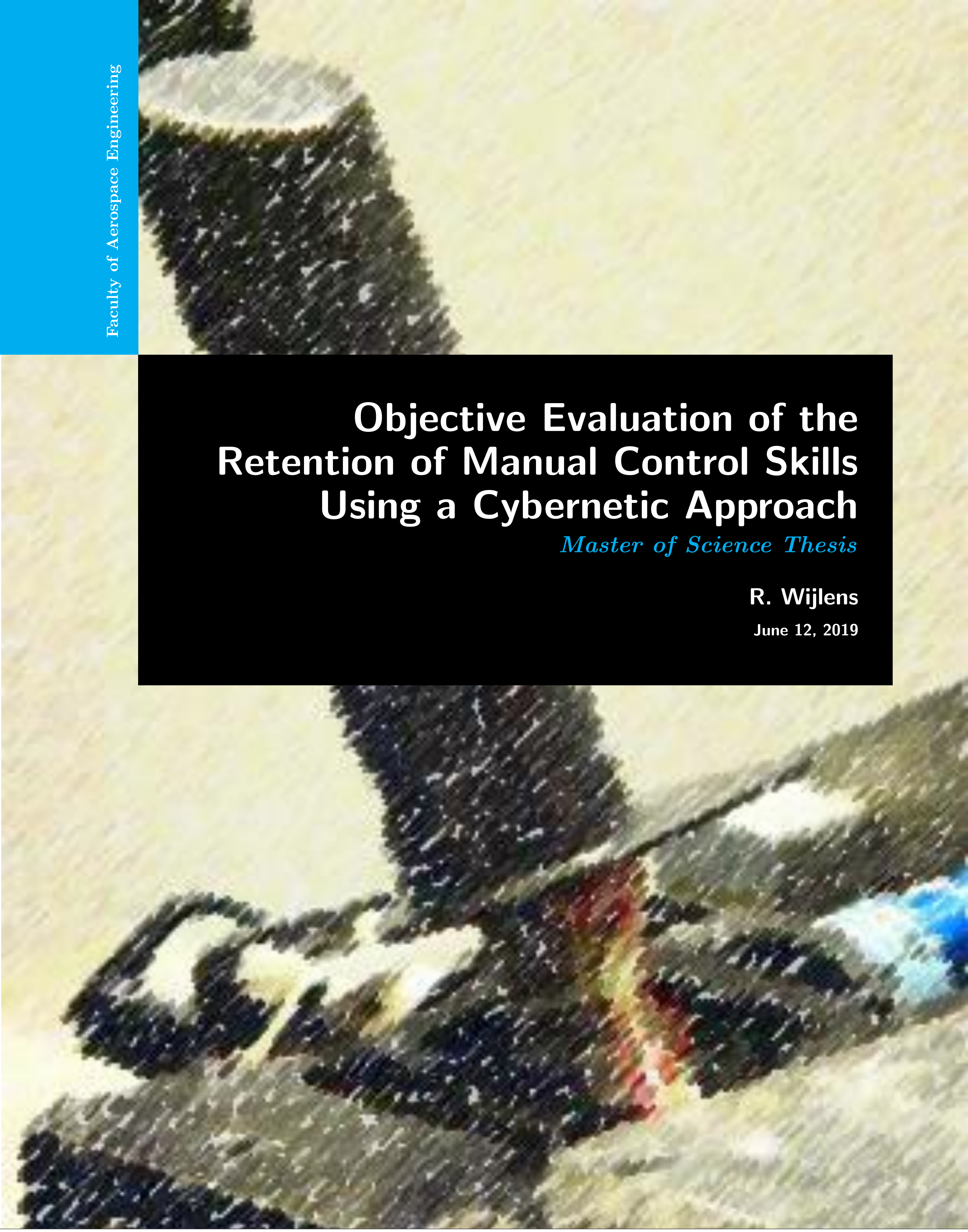
Faculty of Aerospace Engineering



Objective Evaluation of the Retention of Manual Control Skills Using a Cybernetic Approach

Master of Science Thesis

R. Wijlens

June 12, 2019

Delft University of Technology

Challenge the future

Objective Evaluation of the Retention of Manual Control Skills Using a Cybernetic Approach

Master of Science Thesis

For obtaining the degree of Master of Science in Aerospace Engineering
at Delft University of Technology

R. Wijlens

June 12, 2019

Cover Image [1]

Faculty of Aerospace Engineering · Delft University of Technology

Delft University of Technology

Delft University Of Technology
Department Of
Control and Simulation

The undersigned hereby certify that they have read and recommend to the Faculty of Aerospace Engineering for acceptance a thesis entitled **"Objective Evaluation of the Retention of Manual Control Skills Using a Cybernetic Approach"** by **R. Wijlens** in partial fulfillment of the requirements for the degree of **Master of Science**.

Dated: June 12, 2019

Professor:

prof. dr. ir. M. Mulder

Supervisors:

dr. ir. D. M. Pool

dr. ir. P. M. T. Zaal

Readers:

dr. ir. M. M. van Paassen

dr. ir. O. A. Sharpanskykh

Acknowledgments

Although this thesis might be regarded as the product of my work over the past 14 months, it would definitely not have been here today without the support of my family, friends and supervisors. Whereas the past 14 months have only been a part of the six years I have studied in Delft, I can honestly say that they have been the best 14 months thanks to the dedication and enthusiasm of the staff at the Control & Simulation section, and especially the support of my supervisors.

First of all, I would like to thank my daily supervisors, without whom this thesis would not have reached this quality. Daan, I have most certainly not been the easiest student to supervise. You have had to be really patient not only with the enormous amounts of text you have had to read, but also with my indecisiveness, not only in choosing my research topic, but throughout my entire thesis research. I must admit I have not always been the best listener, but I could probably not have had a more approachable supervisor than you; never going crazy about my million emails in which I asked you to change my lab reservation for the hundredth time or about my phone calls on your days off when the simulator was not working, sometimes even in the very early mornings. Peter, our Skype calls so late in the afternoon for me, but so early in the morning for you have provided me with so many insights over the course of my research. Your critical look has definitely made my long writings so much better. Thank you for all the effort you have put into supervising me, albeit from the other side of the world.

Next, I would like to express my gratitude to Max. Max, the time you take to talk to your students, including me, despite your busy schedule, is admirable. You took the time to write long emails when I was having difficulty deciding which research topic to choose. You also spent an ample amount of time giving me advice in person, sometimes about my thesis work, while other times about life in general.

Another person I would sincerely like to thank is Ren, who asked the best critical questions which always provided me with new insights. Ren was also the person who made sure my software would start working again if I could not get it to run anymore.

I would further like to thank my fellow students in SIM 0.08. Without you guys the past 14 months would have been a lot less fun. You guys were there at times when I was stuck, but also when I desperately needed a break.

Last but not least, I am extremely grateful to my parents, my brother, and, of course, Niels, for supporting and encouraging me throughout all these years in whatever I wanted to do.

Although there are so many more people I would like to thank for their contribution to this thesis research, in one way or the other, I simply cannot. My thank you words would, just like all my other writings, turn into a true “Rowenna”-sized piece of text... (I must admit I have already had to “cross out” some of my thank you words.)

Rowenna Wijlens Delft, June 2019

Contents

List of Figures

List of Tables

Nomenclature

Roman Symbols

Symbol	Description	Unit
A_t	Target forcing function sinusoid amplitude	[deg]
e	Tracking error signal	[deg]
f_t	Target forcing function	[deg]
F	Learning rate of learning curve	[-]
$H(s)$	Transfer function	[-]
$H(j\omega)$	Frequency response function	[-]
H_c	Controlled aircraft dynamics	[-]
H_{nm}	Neuromuscular actuation dynamics	[-]
H_{p_c}	Pilot crossfeed dynamics	[-]
H_{p_e}	Pilot response to error signal e	[-]
i	Data point index	[-]
j	Imaginary unit	[-]
k	Forcing function sinusoid index	[-]
K_p	Pilot gain	[-]
K_s	Stick gain	[-]
m	Forcing function realization index	[-]
n	Tracking run index	[-]
n	Pilot remnant signal	[deg]
n_t	Integer multiple of measurement time base frequency ω_m in target forcing function sinusoid frequency ω_t	[-]
N	Number of data points	[-]
N_t	Number of sinusoids in target forcing function	[-]
p_a	Asymptotic value of learning curve	[MDU[1]]

[1]MDU = Metric-Dependent Unit. The unit is equal to the unit of the dependent measure for which the learning curve is modeled.

Symbol	Description	Unit
p_0	Initial value of learning curve	[MDU[1]]
s	Laplace variable	[-]
t	Time	[s]
T_{A_1}, T_{A_2}	Target forcing function filter time constants	[s]
T_I	Visual lag time constant	[s]
T_L	Visual lead time constant	[s]
T_m	Tracking run measurement time	[s]
u	Pilot control signal	[deg]
u	Measured signal	[deg]
$\hat{u}$	Modeled signal	[deg]
u_e	Pilot control signal response to error signal e	[deg]
x_{peak}	Peak amplitude of signal	[deg]
x_{rms}	Root mean square of signal	[deg]
y_{lc}	Vertical coordinate of learning curve	[MDU[1]]

Greek Symbols

Symbol	Description	Unit
δ_a	Aileron deflection	[deg]
δ_e	Elevator deflection	[deg]
ζ_{nm}	Neuromuscular damping ratio	[-]
θ	Pitch attitude	[deg]
ρ	Pearson's correlation coefficient	[-]
σ_t^2	Variance of target forcing function f_t	[deg^2]
τ_e	Pilot error response delay	[s]
ϕ	Roll attitude	[deg]
ϕ_m	Phase margin	[deg]
ϕ_t	Target forcing function sinusoid phase shift	[rad]
$\phi_{t,m}$	Sinusoid phase shift in m^{th} target forcing function realization	[rad]
ω	Frequency	[rad/s]
ω_B	Forcing function input bandwidth	[rad/s]
ω_c	Crossover frequency	[rad/s]
ω_m	Measurement time base frequency	[rad/s]
ω_{nm}	Neuromuscular frequency	[rad/s]
ω_t	Target forcing function sinusoid frequency	[rad/s]

Subscripts

Symbol	Description
e	Error
m	Measurement
nm	Neuromuscular
p	Pilot
t	Target
θ	Pitch
ϕ	Roll

Acronyms

AE	Aerospace Engineering
ANOVA	Analysis of Variance
ARX	Auto-Regressive models with an eXogeneous input
CF	Crest Factor
EFIS	Electronic Flight Instrument System
FAA	Federal Aviation Administration
FC	Fourier Coefficient
HMI	Human-Machine Interaction
LTI	Linear, Time-Invariant
MDU	Metric-Dependent Unit
MLE	Maximum Likelihood Estimation
NASA	National Aeronautics and Space Administration
PE	Preliminary Experiment
PFD	Primary Flight Display
PT	Preliminary Thesis
RI	Retention Interval
RMS	Root Mean Square
RP	Retention Period
RT	Retention Test
SAFO	Safety Alert For Operators
SOP	Successive Organization of Perception
SRS	SIMONA Research Simulator
STD	Standard Deviation
TD	Training Day
TS	Test Subject
VAF	Variance Accounted For

Chapter 1

Introduction

1-1 Background

In the early aviation days, pilots controlled aircraft completely manually. However, it soon became clear that flight control assistance was needed as a means to improve safety and reduce pilot workload. In 1914, Lawrence Sperry demonstrated the use of the first autopilot [2]. Since then, aircraft flight decks have undergone a tremendous transformation. The first commercial aircraft with an Electronic Flight Instrument System (EFIS) incorporated were introduced during the late 1970s and early/mid 1980s, and after that, innovations in flight deck automation have succeeded one another rapidly [3]. This has led to pilots being assigned a different role in the cockpit. Nowadays, pilots take on a mainly supervisory role in the cockpit: they monitor the automated systems, keeping watch for deviations and failures, and they only take over when necessary [4].

Flight deck automation offers significant operational advantages. Wiener [5] has identified the following: increased capacity and productivity, reduction of manual work load and fatigue, relief from routine operations and small errors, more precise handling of routine operations, as well as economical advantages such as a reduction in fuel costs. However, unintended side effects of automation have been identified as well, including, but not limited to, the introduction of unanticipated failure modes, complacency, lack of vigilance and boredom in pilots, reluctance of pilots to take over malfunctioning automation as well as difficulty in recovering from automation failures [6]. Bainbridge [7] has identified several "ironies of automation". Instead of eliminating problems with the human operator, automation seems to actually expand some of the problems. One of these is that automation takes away the opportunity for humans to practice skills. Failures often occur in abnormal circumstances and the human needs to be able to take over when their skills are needed most. Taking away this practice opportunity for pilots has led to a decay in pilots' manual flying skills over the last decades. This skill decay has been proven in several studies.

In 1995, the work of Veillette [8] revealed significant differences in manual control inputs between commercial airline pilots flying aircraft with automated flight decks and commercial pilots flying aircraft with conventional flight decks, especially during abnormal operations. Pilots of automated aircraft, flying manually during training, *"consistently exhibited greater deviations from assigned courses and parameters and greater deviations from nominal pitch-and-bank attitudes. Occasional deviations were great enough to present a hazard to the safety of that aircraft and others in the terminal area."*

Veillette was not the only one identifying problems with the manual flying skills of pilots of automated aircraft. In a 2013 study conducted by the Federal Aviation Administration (FAA), vulnerabilities were identified in pilot skills for manual flight operations, including the retention of manual flying skills [9], where retention indicates the ability to perform these skills after a period of non-practice. In recent years, the FAA has added additional manual flight maneuvers to pilot training program requirements [10]. Additionally, a Safety Alert For Operators (SAFO) was issued recommending operators to promote the practicing of manual flying skills to their pilots [11].

The first retention studies concerning manual flying skills were conducted by the U.S. Armed Forces in the 1960s [12–18]. In the Air Force, proficiency of manual flying skills is critical for operational missions. Pilots must maintain a high state of combat readiness. Because of the high costs involved in aviation training, over the past decades the Air Force has been researching ways to best maintain manual flying skills. An investigation into U.S. Air Force accident rates during training has uncovered that the accident rate temporarily spikes immediately following leave periods, meaning that to maintain proficiency of flying skills, these skills must be practiced at regular intervals [19]. However, the question remains how pilot resources can best be managed with maximum cost effectiveness [16].

Not only in aviation, but also in space flight retaining proficiency of manual control skills becomes ever more important. Mission success and safety of especially long-duration deep space missions depend on the autonomy of the astronaut crew. The two historic risk mitigation factors for (long-duration) missions in low Earth orbit or to the Moon, which are (1) rotating new crew members into the operational environment for complex, mission-critical tasks, and (2) monitoring and supporting space operations in real time from the ground, do not support deep space operations. Communication delays between Earth and deep space make it challenging for Mission Control to provide continuous support from the ground and real-time communication will not be possible at all. Furthermore, because of the transit time to deep space sending spare parts or replacement crew members will not be feasible [20]. Therefore, the astronaut crew must be able to operate autonomously and retaining skill proficiency is of utmost importance. According to figures released in 2007, almost two thirds of all mishaps across NASA are attributed at least partially to human error [21]. Therefore, *"it is necessary to develop an understanding of how training can be tailored to better support long-duration deep space operations (including the extent to which materials, procedures, and schedules of training should be changed from current practices)"* [20]. Initial ground training will be a very long process, causing skills to possibly already decay before launch. The current International Space Station training-to-mission ratio is 10 to 1 [20]. With, for example, a Mars mission length of 32 months [22], even though it is almost impossible to expect an astronaut crew to train for 320 months prior to their mission, one can imagine that ground training will be a lengthy process. Skills might even further decay in transit. To prevent skill decay, onboard refresher training could be applied. However, onboard refresher training will place demands upon the volumetric space, weight and electrical power capacities available within the spacecraft. Smart use of the spacecraft's limited volumetric space, weight and power budgets is of utmost importance to mission success. Therefore, onboard refresher training should only provide the specific type, extent, and fidelity of training required to satisfy the goals of mission safety, reliability, and success, because each upgrade in the level of onboard training system complexity will add volume, weight and additional power use to the training system. Training devices which are more elaborate or complex than required to satisfy these goals without compromise are a waste of resources, but may also compromise the potential range of mission objectives [23]. However, the design of training devices is not the only question that is being faced: *"Questions about which topic should be refreshed and at what interval require a systematic methodology for determination"* [20].

1-2 Problem Statement

This section will elaborate on the motivation for the current research and the research scope. Additionally, this section will define the overall research question that is to be answered in the remainder of this Final Thesis Report.

Motivation It has been proven that with the increase in cockpit automation over the last decades [3], pilots' proficiency of manual flying skills has decreased [8,9]. Despite this established decrease, the FAA has not yet designed a method to determine how often pilots use manual flying skills [24]. Although in a 2013 study conducted by the FAA themselves [9] it was recommended that the FAA should develop and implement standards and guidance for retaining and improving manual flying skills, including (a) practice opportunities for pilots, and (b) training and checking, additional research is required to be able to implement scientifically substantiated standards. Research conducted by NASA [20–22] has also proven that investigating skill decay and retention is of utmost importance to be able to design the training for long-duration deep space missions.

Therefore, this research will focus on the retention of manual flying/control skills. The knowledge gained through the current research will be used to support the development of optimal recurrent training procedures for skill-based manual control in aviation and optimal astronaut crew ground training and onboard refresher training in space flight. This research is conducted in close collaboration with Peter Zaal, contractor with San Jose State University at NASA Ames Research Center.

Scope of research This research will try to set a new standard for measuring manual control skill retention by objectively and explicitly quantifying skill development, decay and retention using a cybernetic approach. This cybernetic approach has been used before in training studies [25], but not yet in retention studies. To facilitate this approach, the current research will focus on skill-based control behavior in manual tracking tasks. Skill-based behavior can be defined as the most basic form of manual control, during which tasks are often executed subconsciously as smooth, automated and highly integrated responses[1] [26]. It has been demonstrated that the control behavior adopted by well-trained individuals in skill-based tracking tasks is usually sufficiently stationary and time-invariant to be able to model it with quasi-linear, time-invariant control-theoretical models [27–29]. A tracking task closely related to flying will be chosen. This research will be narrowed down to the retention of manual control skills of novices. Since it is of utmost importance to the outcome of the research that during the course of the experiment participants refrain from all activities that could (heavily) influence their skill retention, it is not feasible to conduct this research with general aviation pilots. The scope of this Final Thesis Report is dived into two main parts, corresponding to Part I and Part II:

1. *Paper* – The scope of the paper is to provide the setup of a training and retention experiment conducted with 38 participants to investigate the retention of manual control skills of novices using a cybernetic approach. The results of the experiment are presented and discussed in light of the scope defined in the Preliminary Thesis (PT).
2. *Preliminary Thesis* – The scope of the Preliminary Thesis is to provide an overview of the current state-of-the-art literature on the retention of manual control skills and to report on a preliminary experiment conducted to be able to refine the design of the final experiment. The Preliminary Thesis also describes the design and training results of this final experiment.

Research question The main research question to be answered in this research is as follows:

> "To what extent do manual control skills of novices decay during periods of non-practice?"

Objectives The objective of this research is to *objectively and quantitatively evaluate the retention of manual control skills of novices using a cybernetic approach.* The objectives for this Final Thesis Report are stated in fourfold below, corresponding to Parts I – IV.

1. *Paper* – the goal of the Paper is to investigate the retention of skill-based manual control behavior of novices. This is done by conducting a training and retention experiment with 38

[1]The Skills-Rules-Knowledge taxonomy, three categories of human behavior based on different ways of information-processing, was defined by Rasmussen in 1983 [26]. Rule-based and knowledge-based behavior, the two higher, more complex categories of human behavior, are not suitable for identification using a cybernetic approach.

participants, in which the fully task-naive participants are first trained in a specific tracking task and subsequently divided into groups. After a period of non-practice – also referred to as the retention interval – of which the length depends on the group participants are in, the participants are retested to quantitatively analyze their skill decay.

2. *Preliminary Thesis* – the goal of the Preliminary Thesis is to formulate the context for the overall research objective. This is done by conducting an extensive literature study to gain insight into the current knowledge on the retention of manual control skills, as well as by conducting a preliminary experiment to ensure the applicability of the cybernetic approach and to be able to refine the final experiment design. The setup and training results of the final experiment are discussed as well.

3. *Preliminary Thesis Appendices* – the goal of the Preliminary Thesis Appendices is to provide additional insights into the design of the training and retention experiment as well as additional results of the training phase of the experiment.

4. *Final Thesis Appendices* – the goal of the Final Thesis Appendices is to provide additional insights into the training and retention experiment results presented in the Paper.

1-3 Methodology

To objectively and quantitatively evaluate the retention of manual control skills, a human-in-the-loop experiment was conducted in the Human-Machine Interaction (HMI) Laboratory at the Faculty of Aerospace Engineering at Delft University of Technology. The analysis of the experiment data was performed in MATLAB, while statistical analysis was conducted in SPSS.

Relevant literature for the literature review conducted during the preliminary research phase was mainly gathered using the following online databases: AIAA, APA PsycNET, Defense Technical Information Center, Delft University of Technology Control & Simulation Reference Database, Google Scholar, IEEE Xplore, Library of Delft University of Technology, ResearchGate, ScienceDirect, Scopus and Taylor & Francis. The following keywords have been used in the literature search process in various combinations:

Retention, Long-Term, Skill Decay, Manual Control Skills, Manual Flying Skills, Skill-Based Control Behavior, Cybernetics, Cybernetic Approach, Human Operator Model, Time-Varying Identification, Crossfeed, Cross-Coupling, Pilot Remnant

1-4 Report Outline

This Final Thesis Report is divided into four parts. Prior to the first part, an introduction into the background, motivation and research objective for the current research was provided. Part I presents the research paper that was written based on the results of the training and retention experiment with 38 participants conducted in the Human-Machine Interaction Laboratory. Part II is the Preliminary Thesis which contains an extensive literature review into the retention of manual control skills, the design and results of a preliminary experiment conducted to be able to refine the final experiment design as well as the design and training results of the final experiment. Part III consists of a set of appendices providing additional information on, among others, previous experiments concerning the retention of manual control skills, the final experiment design and training results not presented in Part II. Lastly, Part IV presents a set of appendices further specifying, among others, the details of the complete experiment design as well as additional results not presented in Part I.

Part I

Paper

Objective Evaluation of the Retention of Manual Control Skills Using a Cybernetic Approach

Rowenna Wijlens
Delft University of Technology, Delft, The Netherlands

Supervisors: Daan M. Pool, Peter M. T. Zaal, Marinus M. van Paassen, Max Mulder

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 groups based on their training performance. Performance of the first group was re-evaluated after a period of non-practice of six months, while 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 quantify changes in manual control skills, participants' control behavior was modeled using quasi-linear human operator models. 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. However, to construct a decay curve which is able to accurately model skill decay over an extended period of time, a larger-scale experiment should be conducted with a larger number of participants and periods of non-practice ranging from a few hours or days up to several years.

Nomenclature

$A_t[k]$	=	amplitude of kth target sinusoid, deg
E	=	tracking error signal Fourier transform
e	=	tracking error signal, deg
F	=	learning curve learning rate
f_t	=	target forcing function, deg
$H(s)$	=	transfer function
$H(j\omega)$	=	frequency response function
H_c	=	controlled aircraft dynamics
H_n	=	remnant filter
H_{nm}	=	neuromuscular actuation dynamics
H_{p_c}	=	human operator response to crossfeed
H_{p_e}	=	human operator response to error signal e
i	=	tracking run index
j	=	imaginary unit
K_n	=	remnant filter gain
K_p	=	human operator gain
K_s	=	stick gain
k	=	target forcing function sinusoid index
N	=	remnant signal Fourier transform
N_t	=	number of sinusoids in target forcing function
n	=	human operator remnant signal, deg
$n_t[k]$	=	frequency integer factor for kth target sinusoid
p_a	=	learning curve asymptotic value
p_0	=	learning curve initial value
s	=	Laplace operator
T_{A_1}, T_{A_2}	=	target forcing function filter time constants, s
T_L	=	visual lead time constant, s
T_m	=	tracking run measurement time, s
t	=	time, s
u	=	human operator control input, deg
y_{lc}	=	learning curve value
δ_a	=	aileron deflection, deg
δ_e	=	elevator deflection, deg
ζ_n	=	remnant filter damping ratio
ζ_{nm}	=	neuromuscular damping ratio
θ	=	pitch attitude, deg
ρ	=	Pearson's correlation coefficient
σ^2	=	variance, deg^2
τ_e	=	human operator error response delay, s
ϕ	=	roll attitude, deg
$\phi_t[k]$	=	phase shift of kth target sinusoid, rad
$\phi_{t,m}$	=	phase shift of mth realization of target sinusoid, rad
φ_m	=	phase margin, deg
ω	=	frequency, rad $\cdot$ s^{-1}
ω_c	=	crossover frequency, rad $\cdot$ s^{-1}
ω_m	=	measurement time base frequency, rad $\cdot$ s^{-1}
ω_n	=	remnant filter break frequency, rad $\cdot$ s^{-1}
ω_{nm}	=	neuromuscular frequency, rad $\cdot$ s^{-1}
$\omega_t[k]$	=	frequency of kth target sinusoid, rad $\cdot$ s^{-1}

Subscripts

c	=	crossfeed
e	=	error signal
t	=	target forcing function
θ	=	pitch
ϕ	=	roll

I. Introduction

PILOTS' manual flying skills have notably reduced due to the increase in flight deck automation over the last decades [1, 2]. This has resulted in a growing concern that today pilots lack the skills to safely and successfully prevent or recover from unexpected aircraft upset events or to take over controls after a sudden transition to manual flying [3]. Although the necessity to reverse this decline in manual flying skills, by developing and implementing 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 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 of the astronaut crew, as traditional risk mitigation factors are inaccessible. Real-time ground support will be impossible due to communication delays between Earth and deep space, and sending up spare parts or replacement crew members will be unfeasible due to the extensive transit time. For the astronaut crew to be able to operate autonomously, retaining skill proficiency is of utmost importance. 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 – due to the lengthy process of ground training –, it is necessary to investigate how (onboard refresher) training should be designed to better support long-duration deep space operations [9].

Although the first retention studies concerning manual flying skills were already conducted in the 1960s and 1970s [11–18], understanding skill decay has remained a challenging task up to this day. Besides the fact that many who have tried believe a universal skill decay curve 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 long retention periods are involved. Earlier research was often comprised of short retention intervals, i.e., periods of non-practice, 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 skill decay curves [19]. Both positively accelerating [22–24], as well as linear [25], and negatively accelerating [13] skill decay curves have been identified by measuring performance either in terms of absolute errors [22, 23], training time required to regain proficiency [24], time-on-target [25], or on the basis of pilots' self ratings [13].

The goal of this paper is to assess the retention of skill-based manual control behavior using a cybernetic approach [26]. To accomplish this, a human-in-the-loop experiment was performed in a fixed-base setting in the Human-Machine Interaction Laboratory at Delft University of Technology with 38 fully task-naive participants. In the first phase of the experiment, all participants were trained under the same conditions in a compensatory dual-axis roll and pitch attitude tracking task. This task was similar to the task performed in a number of earlier tracking, but not necessarily training, experiments in which the cybernetic approach was successfully applied [27–31]. After the training phase of a fixed number of 100 tracking runs, the participants were divided into three groups based on their 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 differentiated from one another in the length of the retention interval and the number of retention tests they performed.

This experiment setup of three experiment groups with different retention intervals 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 retention interval 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.

In addition to metrics of task performance and control activity, changes in human operator control behavior over the course of the experiment were quantified using quasi-linear human operator models obtained for each individual tracking run. In this paper, the use of this cybernetic approach provides 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 contributes to a systematic optimization of training associated with manual control [33]. Additionally, by being the first retention study to objectively and explicitly quantify skill development, decay and retention using this cybernetic approach, this paper attempts to solve the issue of the use of different performance measures which made the results of earlier retention studies difficult to compare. However, given the expected lack of stationary control behavior during the initial training of fully task-naive participants and the fact that human operator models were fitted to individual run data, the validity of the human operator modeling results was assessed through the quality-of-fit. Finally, the learning characteristics visible in tracking performance, control activity and the estimated human operator model parameters throughout the experiment were quantified using fitted exponential learning curve models.

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 in the current study 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_ϕ and e_θ, respectively. Participants controlled the roll and pitch attitudes, ϕ and θ, 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. This display is depicted in Fig. 2. The arrows indicating the magnitude of the roll and pitch errors as well as the error symbols themselves were not depicted on the display during the experiment.

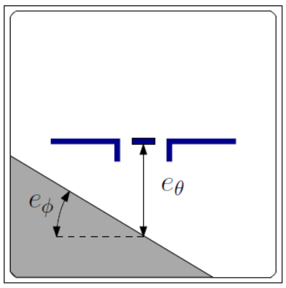


Fig. 2. Dual-axis compensatory display.

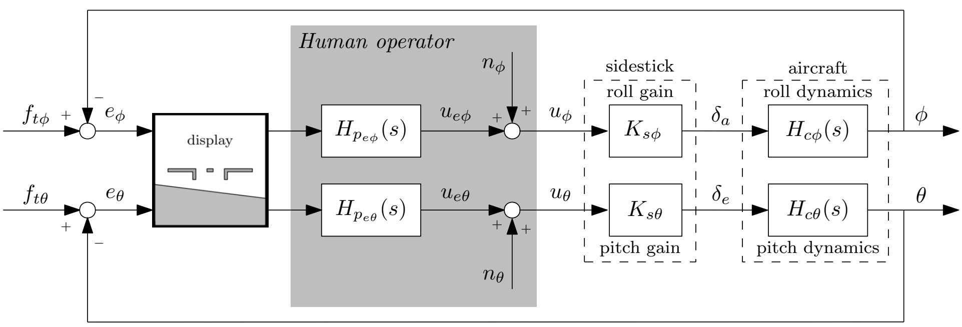
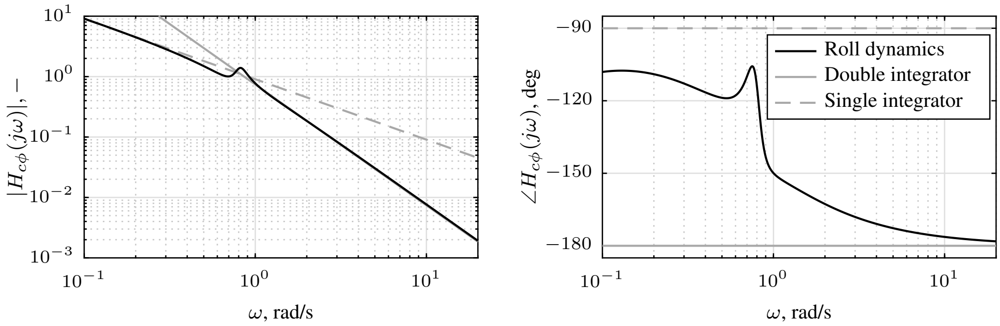
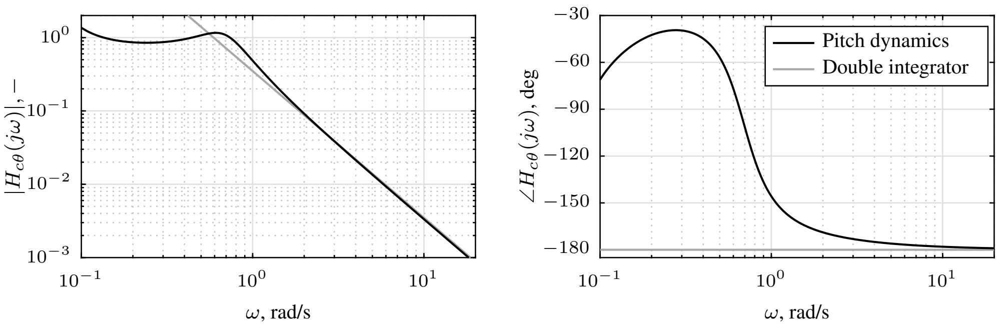


Fig. 1. Schematic representation of the compensatory dual-axis roll and pitch attitude tracking task.



(a) Roll dynamics



(b) Pitch dynamics

Fig. 3. Frequency response of the linearized aircraft dynamics.

The pilot-aircraft system in Fig. 1 is assumed to consist of two independent control loops, since no cross-coupling is present between the roll and pitch dynamics [34,35]. This allows for participant control behavior to be modeled using two independent quasi-linear models, one for the roll response and one for the pitch response. The roll and pitch control inputs, u_ϕ and u_θ, respectively, both consist of a linear error response, $u_{e\phi}$ and $u_{e\theta}$, and a remnant, n_ϕ and n_θ, accounting for nonlinear behavior and measurement noise [36]. The linear human operator error response functions in roll and pitch are represented by $H_{p_{e\phi}}$ and $H_{p_{e\theta}}$, respectively.

B. Controlled Aircraft Dynamics

To make the tracking task feel as realistic as possible different aircraft roll and pitch dynamics were used, as in real nonlinear aircraft dynamics the roll and pitch axes have different dynamics as well. 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 [30].

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

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

The linearized roll dynamics of Eq. (1) have a mildly unstable pole (spiral) in this flight condition. In Fig. 3(a), it is shown that 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. 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. 3(b). Both the roll and pitch dynamics required the operator to adopt lead equalization, making the task rather challenging to perform [32,36].

C. Forcing Functions

To facilitate reliable identification and modeling 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 [32,35]:

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

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^{th} sine in $f_{t\phi}$ or $f_{t\theta}$, respectively.

To allow for estimating frequency domain describing functions for $H_{p_{e\phi}}$ and $H_{p_{e\theta}}$ [35], 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 [27–31,37–41]. 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.

To form sufficiently unpredictable forcing functions [32, 42], 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 [43], 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 [44].

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 [28,31,37,38,40,41,44,45] to reduce

Table 1. Experiment forcing function data

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

amplitudes at higher frequencies, yielding a more feasible control task and minimizing the chances of crossover regression [32, 43, 46].

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

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 $\sigma^2_{t\phi,\theta}$ = 1.5 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) [46, 47]. 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 [47]. 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^{th} realization in roll was always paired with the m^{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 roll and pitch control loops to systems including an additional feedforward path [32, 42].

D. Apparatus

The experiment was performed in the fixed-base flight simulator setup in the Human-Machine Interaction Laboratory at the Faculty of Aerospace Engineering at Delft University of Technology, as shown in Fig. 4. To give roll and pitch control inputs, participants used a control-loaded hydraulic sidestick with $\pm 30°$ excursion in roll and $\pm 22°$ 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 11.0×11.2 cm (width×height). Besides this display, no other visual information (e.g., outside visuals) was displayed during the experiment. Measurement data were logged at a sampling frequency of 100 Hz.

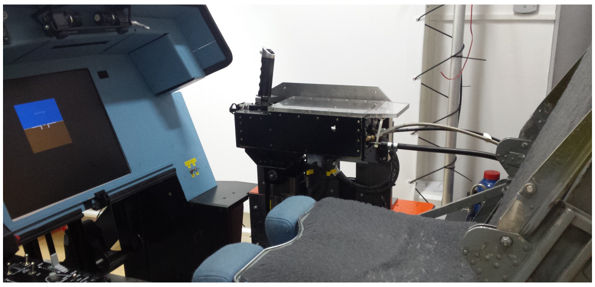


Fig. 4. Flight simulator setup in Human-Machine Interaction Laboratory.

E. Participants and Instructions

A total of 38 fully task-naive 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 Delft University of Technology, 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.

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*, 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. For four consecutive working days, participants performed one session per day 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 [48]. Although there is no solid consensus yet on the optimum amount of time between consecutive training sessions, in a meta-analysis by Kantak and Winstein [49] 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

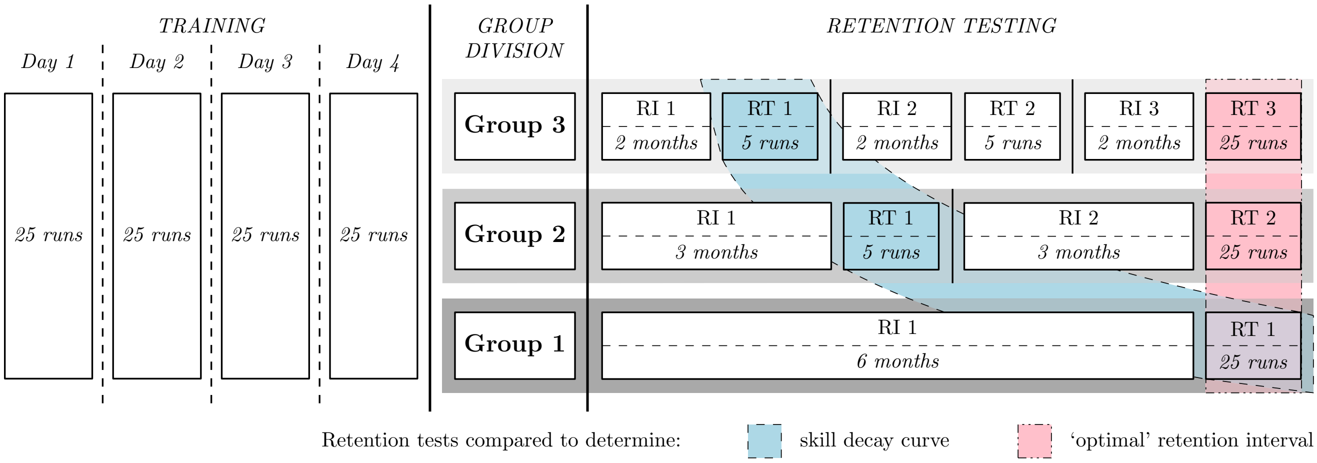


Fig. 5. Experiment setup (RI = Retention Interval, RT = Retention Test).

(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 [50–52], a spaced practice schedule was applied during the current study. 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 [53–55]. 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 experiment groups based on two criteria: their training performance and their availability for retention testing. Training performance of participants was determined based on two criteria as well: (1) by averaging their tracking errors, control inputs and human operator model parameters of the last ten training runs (runs 91-100), and (2) by fitting learning curves to the RMS values of their tracking error signals throughout training. Participants' training performance was evaluated for roll and pitch separately. Subsequently, groups were formed such that there were no significant differences in error scores, control activity, pilot model parameters or any of the learning curve parameters between the three groups to allow for a fair comparison of the retention performance of the different groups.

As illustrated in Fig. 5, the three groups were differentiated from one another in their Retention Interval (RI) length and the number of Retention Tests (RT) they performed. The first group, *group 1*, only performed a single retention test after a retention interval of six months. The second group, *group 2*, performed two retention tests with retention intervals of three months in between. The last group, *group 3*, performed a total of three retention tests with retention intervals of two months in between. This means that all participants performed their final retention test six months after the end of training. At every retention test participants were asked whether they had been involved in any activities during the retention interval that could either positively or negatively affect their retention performance. The final retention test of each group was structured in the same manner as the individual training sessions, meaning that the test consisted of 25 90-second runs with a five-minute break between the first 15 and last ten runs. The other retention tests, i.e., the first retention test of group 2 and the first two retention tests of group 3, consisted of only five 90-second runs without a break. 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 avoiding additional learning as much as possible.

While the first retention test of each group was used to identify the trend of the skill decay curve, as indicated in blue in Fig. 5, the use of all final, 6-month, retention tests was twofold: (1) to identify the 'optimal' retention interval (red highlights in Fig. 5) 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 these retention tests to be longer again compared to the 2-month, 3-month and 4-month retention tests.

G. Dependent Measures

To quantify the acquisition, decay and retention of participants' skill-based manual control behavior in the current experiment, a number of different objective dependent measures were determined from the measurement data. These dependent measures were analyzed for each axis of control separately. The data analysis methods were split up into two parts: the established data analysis methods and the experimental data analysis methods.

Using the established methods, which have successfully been applied in numerous previous tracking and training studies [26, 41, 45, 56, 57] and are regarded as the current state-of-the-art in manual control [33, 58], the following dependent measures were analyzed. 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. Next, the human operator model parameters of the operator's error response H_{p_e} were estimated. The crossover frequency ω_c and phase margin φ_m were calculated as measures of the human operator's tracking ability and the stability of the combined pilot-aircraft system, respectively. Then, the human operator remnant was modeled using its power spectrum. The exact remnant parameters are defined in Sec. II.H.

Using the experimental data analysis methods, which have only been used once or twice before [31, 59] and are meant to facilitate the analysis of the (often time-varying) manual control behavior observed in more realistic control tasks than the tightly controlled and stationary conditions for which the current state-of-the-art methods are only truly valid [33], an attempt was made to identify both human operator crossfeed, i.e., a task interference phenomenon observed in multi-axis manual control and explained more elaborately in Sec. II.I, and time-varying manual control behavior. To identify crossfeed, first the fraction of the crossfeed contribution to the total error and control variances was calculated. Then, the crossfeed dynamics were modeled by parameter estimation. An exact model structure is proposed in Sec. II.I. To gain insight into the time-varying characteristics of manual control behavior, time-varying model parameters of the human operator error response H_{p_e} were estimated.

H. Established Data Analysis Methods

1. Human Operator Modeling and Model Performance

To explicitly assess skill development, decay and retention, participants' control behavior was modeled using a cybernetic approach, which is a system-theoretical, model-based approach to mathematically describe how human operators perform continuous manual control tasks [26,33,36,60–64]. Using the cybernetic approach, operators' control behavior was modeled in terms of distinct contributions that are physically interpretable. The model used for the separate human operator error responses in roll and pitch, $H_{p_{e\phi}}$ and $H_{p_{e\theta}}$ (see Fig. 1), respectively, has been used successfully in previous studies concerning human operator control in either fixed-base or moving-base settings [27,29–31,45]:

$$H_{p_e}(s) = K_p(T_L s + 1)e^{-\tau_e s} H_{nm}(s) \tag{5}$$

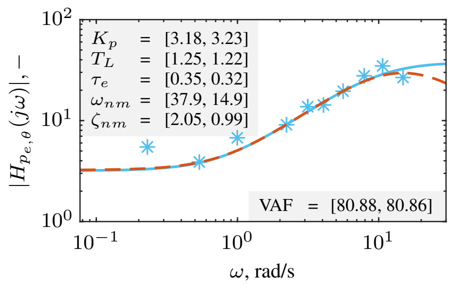
$$H_{nm}(s) = \frac{\omega_{nm}^2}{s^2 + 2\zeta_{nm}\omega_{nm}s + \omega_{nm}^2} \tag{6}$$

The human operator model has been defined such that the combined pilot-aircraft system, $H_{p_e}(j\omega)H_c(j\omega)$, approximates single-integrator dynamics for the frequency range around the crossover frequency [36,65]. The equalization dynamics in Eq. (5), including the gain K_p and lead time constant T_L, capture the human operator's control strategy influenced by the controlled aircraft dynamics [26]. The limitation dynamics, accounting for some of the physical limitations that affect manual control behavior [32,36], include the human operator response delay τ_e to account for the time delays incurred in the perception and processing of the visual information, and the neuromuscular actuation dynamics H_{nm}, which are modeled as a second-order mass-spring damper system with a neuromuscular frequency ω_{nm} and a neuromuscular damping ratio ζ_{nm}. The dynamics of the visual perception sensors, i.e., the eyes that perceive the tracking error on the visual display, are often modeled by a unity gain, and are therefore not explicitly represented in the human operator error response model [65].

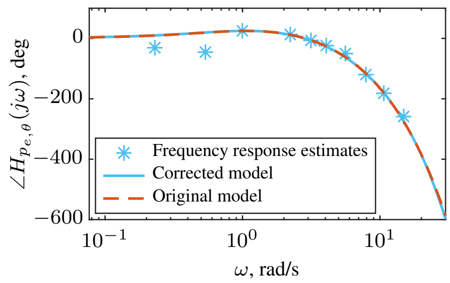
To objectively evaluate human operator control dynamics, the five free parameters of the model defined in Eqs. (5) and (6), were estimated in roll and pitch separately by fitting the model to the time-domain signals $e_{\phi,\theta}$ and $u_{\phi,\theta}$ using a Genetic Maximum Likelihood Estimation (MLE) method. This time-domain estimation technique, described in [56], has been shown to yield more accurate and reliable results than those obtained with two-step frequency-domain identification techniques.

The human operator model accuracy in describing pilots' control behavior was assessed using the Variance Accounted For (VAF). The VAF indicates the amount of variance in the measured human operator control signal which is accounted for by the estimated model, and can be seen as a measure of how well the model is able to describe the human operator data (the higher the VAF, the better). In many tracking studies [29–31,37,56,66] human operator data are averaged between consecutive runs to increase the accuracy of the human operator model, as data averaging results in a decrease in the amount of remnant noise and an increase in the linearity of the measured human control behavior. In that case, the VAF is usually around 80% to 90% for single-axis compensatory tracking tasks [56,66]. However, changes in pilot model parameters throughout the training and retention trials are crucial for evaluating the development, decay and retention of control skills. Averaging results between tracking runs would mask the training and retention effects. Therefore, in the current study, pilot models were fitted to individual tracking runs, resulting in lower VAF values. In previous single-axis tracking task experiments in which individual experiment runs were evaluated [41,45], the majority of pilot models had VAF values between 60% and 80%. However, in dual-axis tracking task experiments, such as the current one, slightly lower VAF values can be expected, as human operators have to divide their attention between two axes, causing more operator remnant and nonlinearities in control behavior [67].

Unfortunately, in the current experiment current-day cybernetics did not always allow for accurate and reliable modeling, as its theory and methods only include accepted, universal, models for single-axis compensatory tracking performed by well-trained human operators [33,58]. The neuromuscular parameters are often the first parameters to be affected by insufficiently consistent or nonlinear operator behavior. Therefore, when extreme neuromuscular parameter estimates occurred, the neuromuscular frequency ω_{nm} was fixed at the maximum input frequency of the principal axis' target forcing function and the other model parameters were estimated once more with the neuromuscular frequency treated as a constant. If a sensitivity analysis revealed that the change in VAF was minimal, the lower neuromuscular frequency and the newly estimated model parameters were adopted. A similar method has successfully been applied in earlier research [68]. Figure 6 shows an example of the effect of limiting the neuromuscular frequency to the target forcing function's maximum input frequency on the human operator model parameters and the VAF. The initial and corrected parameter values and VAF are provided as p = [p_{initial}, p_{corr}]. Limiting the neuromuscular frequency mainly influences the estimates of the neuromuscular actuation parameters.



(a) Human operator magnitude



(b) Human operator phase

Fig. 6. Example pilot model with and without corrected maximum neuromuscular frequency (participant 2, training run 66, pitch axis).

2. Learning Curve Modeling

To quantitatively describe how human operators' control behavior changes during training and after a period of non-practice, exponential learning curves were fitted to all dependent measures obtained using the *established* data analysis methods. The exponential learning curve model is given by Eq. (7) and has successfully been applied in earlier training studies [30,41,45,69].

$$y_{lc}(i) = p_a + (1 - F)^i (p_0 - p_a) \tag{7}$$

In Eq. (7) 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, retention test. No learning curves were fitted to the intermediate retention test data due to the small number of tracking runs performed in those tests. To assess the quality-of-fit of the learning curves, Pearson's correlation coefficient ρ was calculated for each fitted curve. However, only in case ρ 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 itself might give an indication of why it could be less suitable to fit a learning curve to it.

3. Between-Participant Variability

To assess the between-participant variability in tracking performance and control activity, the approach developed by Pool and Zaal [57] 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 retention test 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 retention test 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 retention test (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. (8). While Eq. (8) 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.

$$\bar{\Delta}\text{RMS}(e) = \frac{\text{RMS}(e) - \text{RMS}(e)_{tr,end}}{\text{RMS}(e)_{tr,end}} \tag{8}$$

In Eq. (8) 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 retention test was analyzed instead of during the first retention test of each group, as the 2-month and 3-month retention tests of groups 3 and 2, respectively, did not consist of a large enough number of runs to identify learning curves.

4. Human Operator Remnant

The human operator remnant n is defined as the difference between the modeled output u_e of the linear human operator model H_{p_e} and the measured operator control signal u, as illustrated in Fig. 1, and thus represents the part of the human operator control input u to the aircraft dynamics that is not related to the input to the human operator model, in this case the tracking error e, by the linear time-invariant human operator model. This remnant includes, among others, time-varying and nonlinear control behavior, observation noise and motor noise [70]. As the tracking task in the current experiment extends beyond the single-axis compensatory tracking task for which accepted, universal, models exist, it is imaginable that the control signal u has a stronger contribution of human operator remnant than is usually observed in human operator modeling. Also, a training effect can be expected to occur, as human operator control behavior becomes more linear as training progresses [41]. Therefore, an attempt was made to model the human operator remnant, in addition to the human operator error response. This remnant was estimated from its power spectrum, which can be modeled by a third-order low-pass filter with damping:

$$H_n(j\omega) = \frac{K_n \omega_n^3}{\left((j\omega)^2 + 2\zeta_n\omega_n j\omega + \omega_n^2\right)(j\omega + \omega_n)} \tag{9}$$

where K_n is the remnant filter gain, ω_n the remnant filter break frequency and ζ_n the remnant filter damping ratio. Similar remnant characteristics have been found in earlier research [56]. In Sec. III it will be shown that this remnant model does indeed model the remnant data in the current experiment well.

5. Statistical Analysis

For statistical analysis of the change 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 retention test. Additionally, two-run averages were taken at the start and end of the 3-month retention test for group 2 and at the start and end of the 2-month and 4-month retention tests 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 retention test, (3) between the start and end of every retention test, and (4) between the end of training and the end of every retention test. 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 difference 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.

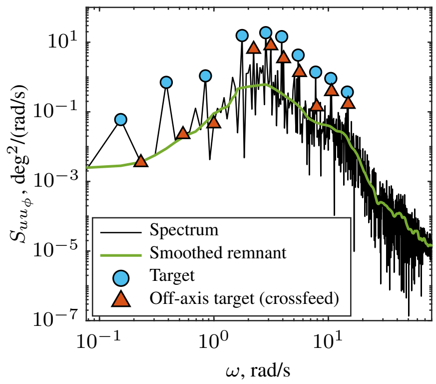


Fig. 7. Example PSD of roll control signal $u_\phi(t)$ (participant 2, training run 68).

I. Experimental Data Analysis Methods

1. Human Operator Crossfeed Identification

Following the current state-of-the-art cybernetics, multi-axis human manual control is often modeled as if multiple fully independent single axes are controlled [33, 58], as shown in Fig. 1 and as also applied in the *established* data analysis methods in the current experiment. However, earlier research has shown that manual control in multi-axis tasks is actually markedly different from single-axis control [31, 67, 71–73]. One of the observed phenomena in multi-axis human manual control is the presence of crossfeed [31, 71–74]. Crossfeed can be described as a form of task interference in which the human operator is not able to completely decouple two tasks [72]. 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 [75].

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 it can be assumed that the remnant signal power is continuously distributed over the frequency spectrum [32,43]. 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. 7. 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.

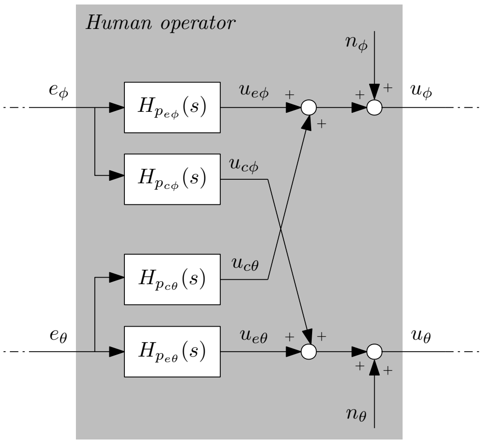


Fig. 8. Schematic representation of human operator in compensatory dual-axis roll and pitch attitude tracking task with crossfeed.

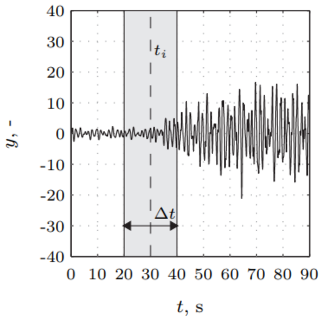


Fig. 9. MLE time window [59].

After the signal variance decomposition, the presence of crossfeed and its training/retention effects were further analyzed by modeling the crossfeed response. To do this, the human operator model in Fig. 1 was replaced by the human operator model in Fig. 8. To determine the model structure of the human operator crossfeed response H_{p_c}, an extended frequency-domain Fourier coefficient (FC) method was used [31, 76]. In this method, the total pitch control input in the frequency domain is defined as follows:

$$U_\theta(j\omega) = E_\theta(j\omega)H_{p_{e\theta}}(j\omega) + E_\phi(j\omega)H_{p_{c\phi}}(j\omega) + N_\theta(j\omega) \quad (10)$$

A similar expression for the roll control input was obtained using the same derivation method described in [31]. For human operator identification, Eq. (10) had to be solved for its two unknowns: $H_{p_{e\theta}}(j\omega)$ and $H_{p_{c\phi}}(j\omega)$. To be able to identify the additional unknown crossfeed response $H_{p_{c\phi}}$ the independent target forcing function components from the other axis had to be used in addition to the target forcing function components of the principal axis. The requirement for this approach to be successful is that the target forcing functions of the two axes are independent, i.e., are constructed of sines with different frequencies. The complete procedure used to obtain the FC frequency response estimates of the error and crossfeed responses is described in [31, 76].

Using the above frequency response estimate method, the model structure of the human operator crossfeed response H_{p_c} was determined. As found in the first study to have successfully used objective human operator identification techniques to verify the presence [74] and dynamics [31] of the human operator crossfeed response, the crossfeed dynamics can be modeled using the same model structure as the error response:

$$H_{p_c}(s) = K_{p_c}(T_{L_c}s + 1)e^{-\tau_{e_c}s}\frac{\omega_{nm_c}^2}{s^2 + 2\zeta_{nm_c}\omega_{nm_c}s + \omega_{nm_c}^2} \quad (11)$$

In Sec. III it will be shown that this model structure does indeed model the crossfeed dynamics well. In the current study, the model parameters were estimated by adapting the Genetic MLE method [37] such that this parameter estimation method estimated the parameters of both the human operator error and crossfeed responses.

2. Time-Varying Human Operator Model Identification

Although the current state-of-the-art cybernetics theory and methods model human manual control behavior as being (quasi-)linear and time-invariant, in real life manual control behavior is often nonlinear as well as time-varying [33, 58]. Although not yet proven, it is likely that some human operator parameters will change faster than others [33], i.e., different parameters have different "life expectancies", especially during (initial) training. To gain a better insight into the temporal scales of learning, time-varying manual control identification and modeling methods should be applied. Therefore, in the current study the time-varying characteristics of manual control behavior in the dual-axis tracking task performed by task-naive participants was investigated by applying a windowed MLE method to obtain a time-varying human operator error response model. This windowed parameter estimation method is the time-varying implementation of the MLE method used for the time-invariant estimation of the human operator model parameters [56]. In the time-varying implementation, which has been applied before [59], the MLE optimization is performed at every time step t_i using a sliding time window of length Δt, as visualized in Fig. 9. Setting a too small time window would negatively impact the accuracy of the estimated model parameters associated with low-frequency dynamics, whereas a too large window would reduce the ability to detect small variations in human operator model parameters. In the current study, the length of the time window was set to 20 s, equal to the length used in [59].

To reduce the required amount of computational power, the human operator model parameters found for a particular experiment run using the time-invariant MLE method were used as initial parameters for the time-varying model identification applied to that tracking run. Due to the high computational effort, time-varying model identification was performed on a subset of the experiment data. The time-varying model parameters were estimated for the last five training runs and the retention runs of eleven participants (four from group 1, three from group 2 and four from group 3), where the number of retention runs depended on the group the participant was in. The eleven participants were selected based on the time-varying nature of their control behavior observed in the time traces of their control signals, their overall control activity and on the VAF values of the human operator models estimated using the time-invariant MLE method. Both the control activity and the VAF values were desired to be high to increase the chances of accurate and reliable identification. Furthermore, the experiment runs at the end of training and during retention testing were identified, as the goal was to investigate the influence of a period of non-practice on the time-varying characteristics of manual control behavior.

J. 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 [77]. 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, 22, 78, 79]. Since overlearning is known to enhance retention [16, 19, 80, 81], this had to be taken into account when analyzing the experiment results.

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

	Group 1		Group 2				Group 3				
	Participant	RI 1 / Total RP (days)	Participant	RI 1 (days)	RI 2 (days)	Total RP (days)	Participant	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		**181.9**		**89.8**	**92.5**	**182.4**		**58.2**	**62.0**	**62.7**	**182.8**
STD		**1.6**		**0.7**	**1.7**	**1.5**		**2.8**	**1.9**	**3.1**	**2.7**
Ideal		*182*		*91*	*91*	*182*		*60.7*	*60.7*	*60.7*	*182*

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 retention intervals differed slightly from the "ideal" ones due to participant availability. Especially the retention intervals of group 3 (2-month retention intervals) contained some more variability. This was caused by holidays due to which the first retention interval of this group differed more from the "ideal" retention interval. Consequently, the other two retention intervals of this group also differed slightly more from the "ideal" ones in an attempt to ensure that the number of days between the end of training and the final, 6-month, retention test of this group was as similar as possible as the number of days for groups 1 and 2. The exact retention intervals can be found in Table 2*.

K. 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. These hypotheses relate to the experiment results obtained using the *established* data analysis methods only. No hypotheses were formulated for data that were analyzed using the experimental data analysis methods, as the definition 'experimental' already implies that the limits and capabilities of these data analysis methods are still being explored.

As observed in a number of earlier training experiments [26, 41, 45], clear effects of training were expected to occur. We hypothesized that training causes an improvement in performance and task proficiency (**hypothesis 1**). Training will be evident from improved performance (lower RMS(e)) and higher crossover frequencies and phase margins. In the human operator models, parameters known to be related to improved performance were expected to evolve (increased K_p, decreased T_L and τ_e). Finally, an increase in human operator linearity was expected to occur (increased VAF).

As also found in previous dual-axis tracking task experiments [27–31], 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.

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 shape of skill decay curve has been found in two flying task experiments [23, 24] that are most comparable to the control task used in the current research.

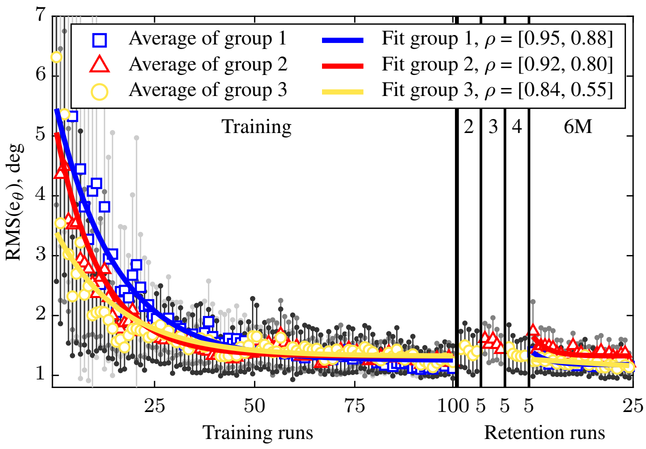
During the last retention tests, six months after training, the best performance and task proficiency was expected for group 3, whereas the worst performance will be exhibited by group 1 (**hypothesis 4**). This expectation was based on the fact that individuals perform better at retention testing if they are provided with some form of practice during the retention interval [23, 82–84]. When comparing the last retention tests of each group, the experiment setup can also be seen as if all groups have a retention interval 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 retention tests, six months after training, degraded control skills of all three groups were predicted to be reacquired at a faster rate than the initial acquiring rate during the training phase (**hypothesis 5**). Earlier retention experiments concerning motor skills have consistently shown that retraining after a retention interval 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, 79, 85].

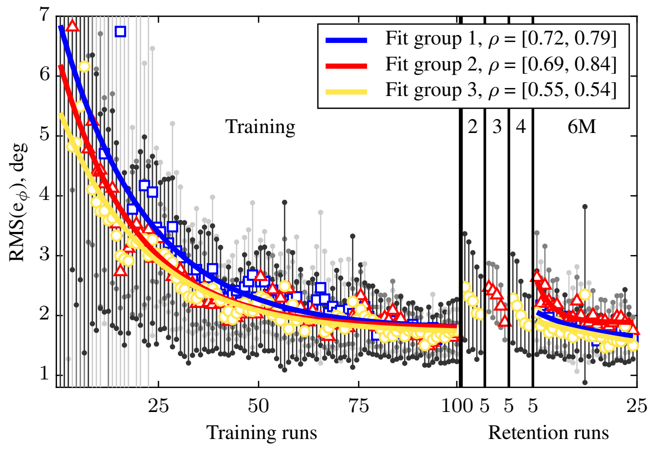
III. Results

This section presents the experiment results. For plots showing data from all experiment runs, 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. Solid black vertical lines indicate the interval between training and the first retention test, as well as the intervals between subsequent retention tests. At the top of each figure, the experiment phase is indicated, where 'Training' indicates the 100 training runs for all groups, '2' and '4' represent the 2-month and 4-month retention tests of group 3, '3' the 3-month retention test of group 2 and '6M' the 6-month retention test of all groups. 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 retention test as $\rho = [\rho_{\text{training}}, \rho_{\text{retention}}]$.

*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.



(a) Pitch performance



(b) Roll performance

Fig. 10. Average pitch and roll tracking error with corresponding learning curves.

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

RMS(e), deg	Pitch θ						Roll ϕ					
	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 θ			Roll ϕ		
Comparison	Group 1	Group 2	Group 3	Group 1	Group 2	Group 3
Training	$**$	$**^a$	$**^a$	$**^a$	$**^a$	$**^a$
End TR - St 2M			$*^a$			$**^a$
St 2M - End 2M			$—^a$			$*^a$
End TR - End 2M			$—^a$			$—^a$
End TR - St 3M		$**^a$			$**$	
St 3M - End 3M		$*^a$			$**$	
End TR - End 3M		$*$			$*$	
End TR - St 4M			$—^a$			$*^a$
St 4M - End 4M			$—^a$			$**^a$
End TR - End 4M			$—^a$			$—^a$
End TR - St 6M	$*^a$	$*^a$	—	$*^a$	$**$	$—^a$
St 6M - End 6M	$**^a$	$*^a$	—	$**^a$	$**$	$*^a$
End TR - End 6M	—	$—^a$	—	$—^a$	—	$—^a$

[a] 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 θ	Roll ϕ
Start training	$—^a$	$—^a$
End training	$—^a$	$—^a$
Start 6-month test	$—^a$	$—^a$
End 6-month test	$—^a$	$—^a$
Start RT1 of each group (St 2M G3, St 3M G2, St 6M G1)	$—^a$	$—^a$

[a] At least one sample not normally distributed, Kruskal-Wallis test applied instead of one-way ANOVA.

A. Tracking Performance

Tracking performance is defined in terms of the RMS of the roll and pitch error signals e, i.e., the errors presented to the human operator on the PFD. The lower the value of RMS(e), the better the task performance is. Figures 10(a) and 10(b) show the average pitch and roll RMS(e) per experiment run, respectively. The parameters of the fitted learning curves are presented in Table 3. Table 4 presents the statistical analysis results of the training and retention effects within groups.

At the start of training, average tracking errors differed 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 exhibited steep initial learning for the first 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 7×10^{-2} in pitch and 5×10^{-2} in roll (see Table 3). Still, over the course of training, the difference in average tracking errors between the groups decreased. At the end of training, average values of around 1.30 deg in pitch and 1.76 deg in roll were reached by all three groups, as shown in Fig. 10 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 [30]. 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 [37, 38]. 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 training pitch tracking performance was consistently better than roll tracking performance for all three experiment groups, as expected from earlier dual-axis tracking task experiments [27–31]. 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 [31]. However, the difference in performance between pitch and roll decreased throughout training. While 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.

During the retention phase, performance was again consistently better in pitch compared to roll. However, larger performance improvements were observed in roll compared to pitch. When comparing retention and training performance, it can be seen that performance in roll relapsed more than performance in pitch. This is consistent with previous studies on skill retention [16, 19, 80, 81], as overlearning is known to enhance retention and participants exhibited steeper learning curves and earlier stabilization in pitch than in roll during the training phase. At the start of each group's first retention test (the 2-month retention test of group 3, the 3-month retention test of group 2 and the 6-month retention test 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 retention tests 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 retention tests. 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.

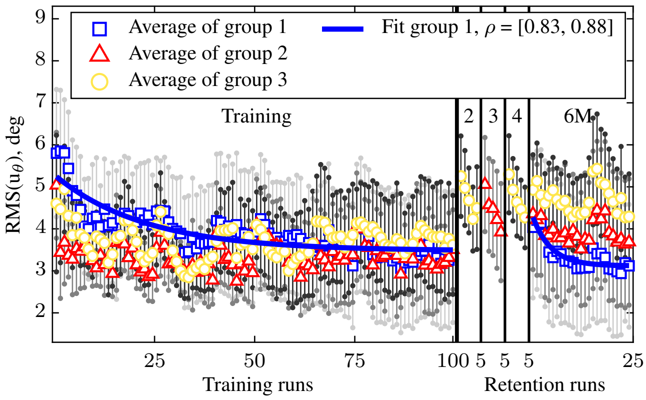
During the 6-month retention test, group 2 performed consistently worse than groups 1 and 3, which exhibited similar tracking performance (see Fig. 10). This can be considered a curious result, as from the earlier finding that operators perform better during retention testing if they have received some form of practice during the retention interval [23, 82–84], it was expected that group 1 would show the worst performance during the 6-month retention test. 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 retention test. As a result, the groups' performances during the 6-month retention tests did not give an indication of what the 'optimal' retention interval is while at the same time minimizing the amount of refresher training, as was one of the ideas behind the 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 retention test, 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 retention test. For group 3, a very low learning rate of 0.13×10^{-2} in pitch during the 6-month retention test 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 retention test when compared to end of training, both in pitch and roll, whereas group 3 did not show any significant decrements compared to end of training due to its earlier 'practice' opportunities in the 2-month and 4-month retention tests. However, the higher learning rates during the 6-month retention test 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.

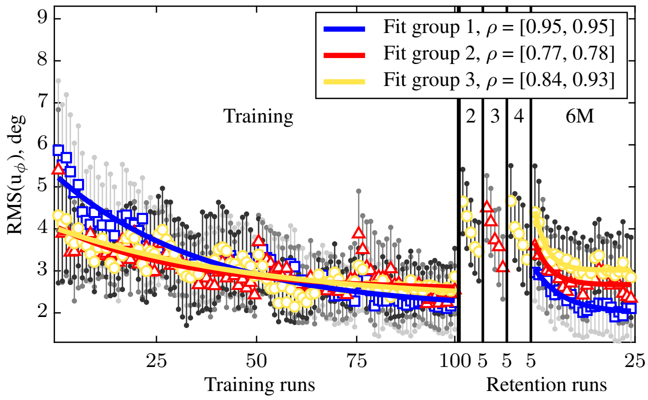
B. Control Activity

Control activity is measured in terms of the RMS of the pitch and roll control signals, u_θ and u_ϕ, respectively. A lower RMS(u) indicates less control effort. Operator control activity for pitch and roll are shown in Figs. 11(a) and 11(b), respectively. The parameters of the fitted learning curves are provided in Table 6.

Over the course of the training phase, the change in control input was different in pitch compared to roll, as shown in Fig. 11. Whereas in roll, the performance improvement during training was achieved with a significant decrease in control input for all three groups, 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 similar control activity as groups 2 and 3 in pitch (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). A decrease in between-group differences in control input was demonstrated throughout training for both pitch and roll. While 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_θ). The significant decreases in control input through-



(a) Pitch control activity



(b) Roll control activity

Fig. 11. Average pitch and roll control input with corresponding learning curves.

Table 6. Learning curve parameters for control input

RMS(u), deg	Pitch θ						Roll ϕ					
	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

RMS(u)	Pitch θ			Roll ϕ		
Comparison	Group 1	Group 2	Group 3	Group 1	Group 2	Group 3
Training	*[a]	—	—[a]	**	*[a]	*
End TR - St 2M			*[a]			**
St 2M - End 2M			—[a]			**
End TR - End 2M			*[a]			*
End TR - St 3M		**			**[a]	
St 3M - End 3M		*[a]			**[a]	
End TR - End 3M		—[a]			*[a]	
End TR - St 4M			*[a]			**
St 4M - End 4M			*[a]			**
End TR - End 4M			—[a]			*
End TR - St 6M	*[a]	—[a]	—[a]	*	—[a]	**
St 6M - End 6M	**[a]	—[a]	—	*[a]	**[a]	**
End TR - End 6M	—[a]	—[a]	—[a]	—[a]	—[a]	—

[a] 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 θ	Roll ϕ
Start training	—	—
End training	—[a]	—[a]
Start 6-month test	—[a]	—
End 6-month test	—[a]	—[a]
Start RT1 of each group (St 2M G3, St 3M G2, St 6M G1)	—[a]	**

[a] At least one sample not normally distributed, Kruskal-Wallis test applied instead of one-way ANOVA.

out training were achieved despite a significant spread in control input data, which is consistent with earlier findings [37, 38, 41], although the spread was less prominent in roll compared to pitch.

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. 11 shows that for all groups and retention tests 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 retention test 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 retention tests 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 retention tests could not be considered significant. Also, during the retention tests, 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 retention tests 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. 11 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 retention tests compared to training.

C. Between-Participant Variability

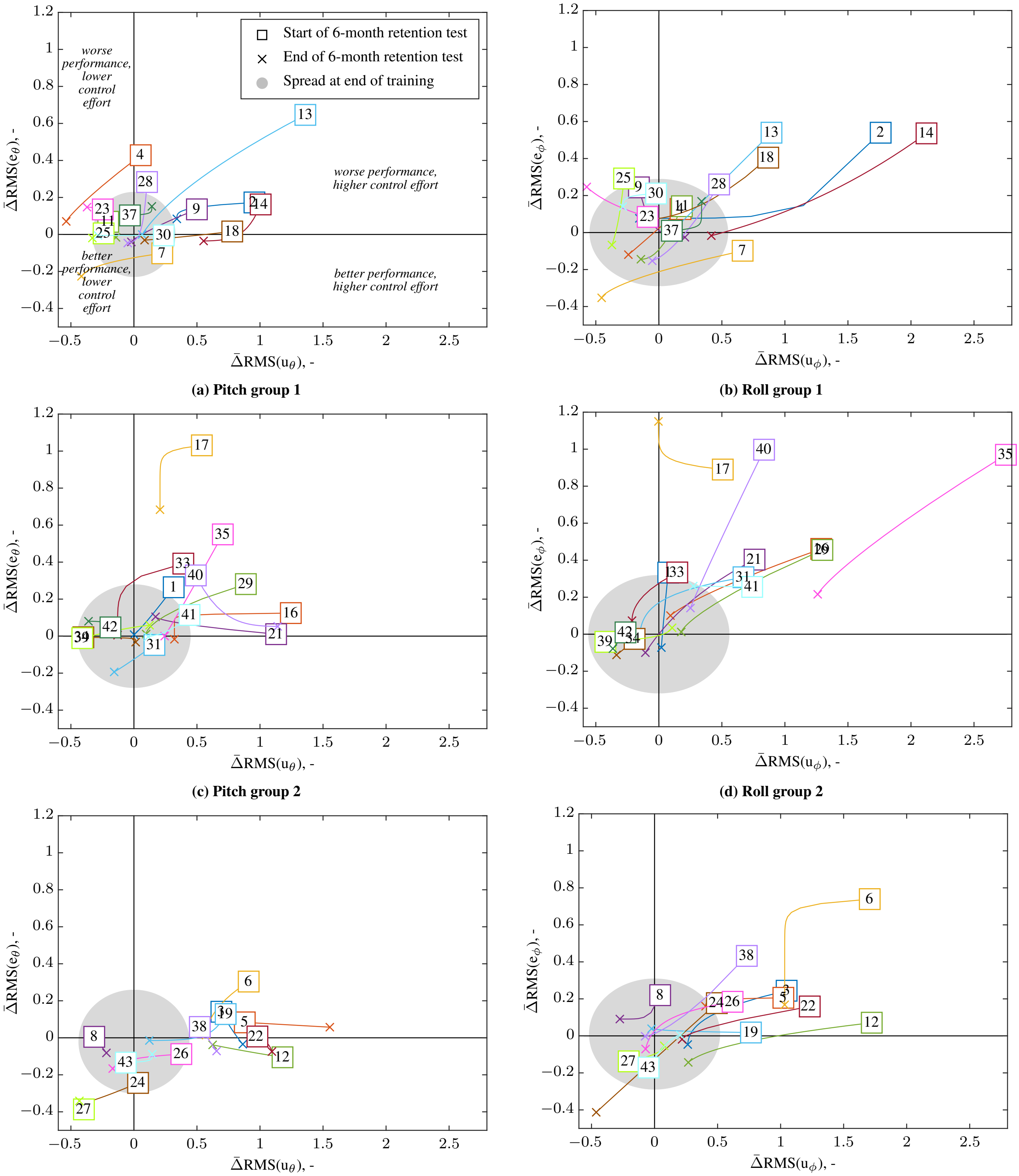
Figure 12 presents the retention performance of individual participants in the 6-month retention test in terms of the relative change in tracking performance and control activity with respect to end of training, $\bar{\Delta}$RMS(e) plotted against $\bar{\Delta}$RMS(u). The left column of graphs concerns 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 retention test, are indicated with square markers with the participant numbers in it. The evolution of the tracking errors and control inputs throughout the 6-month retention test 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 retention test. Finally, the gray oval area in each plot defines 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. 12 indeed show that a notable number of squares, indicating the behavior of participants in the first run of the retention test, are in the upper half of the graph. An exception to this, however, is seen in Fig. 12(e), 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 retention test compared to end of training.

At the end of the retention test, 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. 12(a), 12(b), 12(e) and 12(f), showing that at the end of the retention test 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. 12(c) and 12(d) show that at the end of the 6-month retention test 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 retention interval.

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 retention test. In a considerable number of these cases, the crosses also fall within these gray oval areas, illustrating that at the end of the retention test these participants still performed and behaved similar to at the end of training.

It must be noticed in Figs. 12(a) - 12(f) that the majority of participants started off the retention test 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 not performing the task. Additionally, it is likely that concentration levels were high, as participants had just started. Overall, group 3 seemed to have lost the least amount of skills, as the 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 end of training and the 6-month retention test. Participants of group 2 had clearly lost the most skills, as Figs. 12(c) and 12(d) have the most square markers close to the top of the graph. Whereas most participants showed an increase in performance again during the retention test (solid line going in a downward direction from square to cross), Fig. 12(d) shows that the tracking per-



(a) Pitch group 1

(b) Roll group 1

(c) Pitch group 2

(d) Roll group 2

(e) Pitch group 3

(f) Roll group 3

Fig. 12. Individual performance at 6-month retention test expressed in $\bar{\Delta}$RMS(e) vs. $\bar{\Delta}$RMS(u).

formance in roll of participant 17 in group 2 even worsened during the retention test, 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 retention test, it is consistent with the tracking error results in Fig. 10.

Based on Figs. 12(a) - 12(f), 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 retention test compared to their end-of-training performance (ΔRMS(e)). Subcategories are based on participants' control

Table 9. Participants' skill retention at 6-month retention test categorized based on tracking performance and control activity

		Group 1			Group 2			Group 3		
		Only pitch	Only roll	Both axes	Only pitch	Only roll	Both axes	Only pitch	Only roll	Both axes
1. No change remains within end-of-training spread		25	4, 9	11, 30, 37	31		34, 39, 42	26	27	8, 43
2. Always worse than at the end of training	a) Higher control effort	2			40	35	17	5		6
	b) Lower control effort			23						
	c) Higher → lower control effort	4								
3. Worse (higher control effort) → **no change** 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
4. Worse → better than at the end of training	a) Higher control effort	14						3, 22, 38		
	b) Lower control effort		25							
	c) Higher → lower control effort								24	
5. Always better than at the end of training	a) Higher control effort							12		
	b) Lower control effort							27		
	c) Higher → lower control effort			7				24		

activity compared to their end-of-training control inputs ($\Delta \mathrm{RMS}(u)$). A right arrow indicates that during the first run of the retention test participants' tracking performance or control activity, respectively, was at one end of its spectrum when compared to 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. While 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. 10 and 11, respectively, with the individual results in Fig. 12, it is clear that there were notably more different types of control behavior than has become clear from the group average results. Although the group average results often showed similar changes in control behavior for the pitch and roll axes, a considerable number of individuals actually exhibited different kinds of control behavior in the two axes. Thus, the group average results actually mask many of the different retention effects.

D. Human Operator Modeling Results

To gain more insight in the development, decay and retention of skill-based manual control behavior, linear human operator models were identified for each individual run performed by every participant. Data from the pitch and roll axes were analyzed separately, resulting in two operator models per experiment run. Firstly, the human operator model accuracy in describing the operators' control behavior is assessed using the VAF. Then, the changes in the estimated model parameters throughout the training and retention phases are presented.

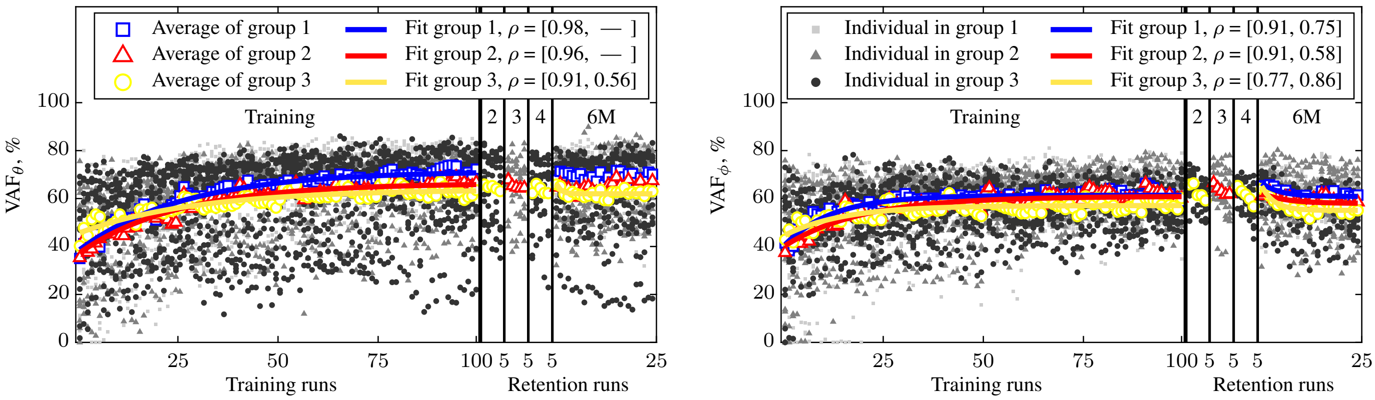
1. Variance Accounted For (VAF)

Figures 13(a) and 13(b) present the average group VAF values per experiment run in pitch and roll, respectively, together with fitted learning curves. Instead of showing the 95% confidence intervals, the VAF values of all individual runs are plotted with gray markers. For each group, the shape of the gray markers corresponds to the shape of the colored markers indicating the group's averages. The parameters of the fitted learning curves are provided in Table 10.

Figures 13(a) and 13(b) show that the average VAF values obtained for the models fitted to the individual experiment runs were around 65% in pitch and 58% in roll, respectively. While in previous single-axis tracking task experiments in which individual experiment runs were evaluated, nearly all estimated models had VAF values between 60% and 80% [41, 45], in the current experiment the VAF values were lower due to the dual-axis tracking task, as divided attention between two axes results in lower operator linearity compared to single-axis tracking tasks. The slightly lower VAF observed in roll compared to the pitch axis, as seen in both the training and retention phases, is consistent with previous experiments [27–31] and indicates that operators behaved less linearly in roll compared to pitch. This might be due to the difficulty in observing the roll error on the PFD, but it could also be caused by a (conscious) choice of participants to focus more on pitch instead of roll. Lower VAF values were also expected in this experiment due to the absence of motion feedback, as in previous experiments concerning the influence of motion feedback on human operator control behavior [38, 65], lower VAF values were obtained in fixed-base environments compared to motion-base settings, as well as due to the fact that participants were fully task-naive. The latter was especially reflected in the results obtained during early training, as average VAF values in both pitch and roll were around 40% at the start of training.

When comparing VAF values between the training and retention phases, Fig. 13 shows that in both pitch and roll VAF values obtained in retention runs were very similar to those obtained at the end of training. However, in closer detail, during retention tests VAF values started out slightly higher than at the end of training and then gradually decreased, which was similar to the trend in control activity in Fig. 11. A probable cause for this is a high attention or motivation at the start of the retention tests, followed by a decrease in attention or motivation throughout the tests, as the VAF values in the training phase also showed signs of a "sawtooth shape", similar to the control activity in Fig. 11.

In the following section, the development, decay and retention in human manual control behavior are quantified by fitting learning curves to the estimated human operator model parameters. However, as clearly seen in Fig. 13, the VAF values of the estimated models of a fair number of experiment runs were extremely low, i.e., below 30%, a result of human operator behavior not being sufficiently stationary and linear to allow for reliable modeling results. As for unreliable modeling results, human operator model parameters often take on extreme values which could strongly bias the fit of learning curves, in the following section the human operator model parameters of experiment runs with a VAF value lower than 30% were excluded from the learning curve fitting process. While in previous research [45] a threshold of 40% was upheld, a lower threshold was used in the current experiment due to the already lower VAF values expected in the dual-axis



(a) Pitch **(b) Roll**

Fig. 13. Average and individual pitch and roll Variance Accounted For with corresponding learning curves.

Table 10. Learning curve parameters for VAF

VAF, %	Pitch θ						Roll ϕ					
	Training phase			Retention phase			Training phase			Retention phase		
	p_0, %	p_a, %	$F(\times 10^{-2})$	p_0, %	p_a, %	$F(\times 10^{-2})$	p_0, %	p_a, %	$F(\times 10^{-2})$	p_0, %	p_a, %	$F(\times 10^{-2})$
Group 1	37.4	71.5	3.73	n/a	n/a	n/a	39.0	61.4	7.97	67.8	60.9	20.0
Group 2	36.2	66.4	3.93	n/a	n/a	n/a	39.3	60.9	5.03	63.0	58.0	15.4
Group 3	45.1	63.8	3.40	72.3	62.0	59.1	43.9	57.1	6.63	69.0	55.1	27.3

tracking task. However, to present an honest view of the experiment results, these (sometimes) extreme model parameters are still included in the groups' average parameter values.

In total, 4.4% of the models for pitch and roll on average were excluded from the learning curve fitting process due to their low VAF. As expected, this percentage was higher than in previous single-axis tracking task experiments performed with fully task-naive participants in which human operator models were estimated for individual experiment runs [41, 45], due to the overall lower VAF values found in the current experiment as a result of the task configuration (dual-axis) and the absence of motion feedback. The majority of training runs (69%) with VAF values lower than 30% were performed on the first training day, as control behavior is often more inconsistent and nonstationary when human operators are still inexperienced.

2. *Human Operator Model Parameters*

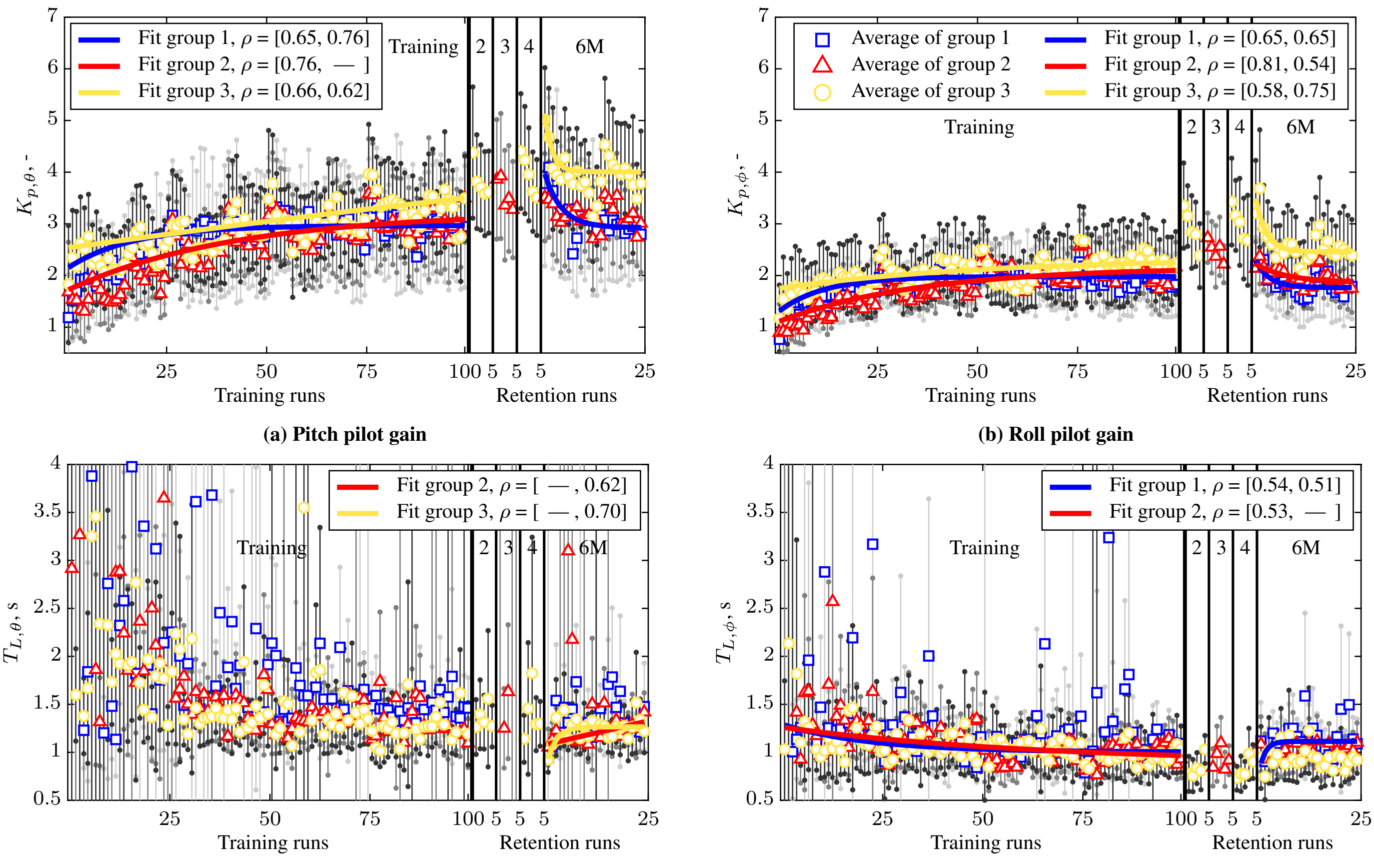
Figures 14 and 15 present the estimated parameters of the equalization and limitation dynamics, respectively. The parameters of the corresponding fitted learning curves are listed in Tables 11 and 12. In addition to excluding experiment runs with obtained VAF values lower than 30% for the fitting of the learning curve models, in Figs. 14(c) and 14(d) experiment runs with a lead time constant T_L higher than 3.0 s were excluded from the fitting process as well, resulting in an additional 3.5% of excluded experiment runs. In Figs. 15(a) and 15(b), experiment runs with a human operator response delay τ_e higher than 0.7 s were excluded from learning curve fitting, resulting in an additional 1.0% of excluded experiment runs. Parameters above these thresholds for T_L and τ_e were considered artefacts caused by a low signal-to-noise ratio of the measurement data which made reliable parameter estimation and model fitting impossible.

Figures 14(a) and 14(b) show that the human operator gain K_p increased throughout training in both pitch and roll, respectively, for all three groups (on average from around 2.2 to 3.1 in pitch and 1.4 to 2.1 in roll), as expected from earlier training studies [26, 41]. In addition, K_p was consistently higher in pitch than in roll for all three groups in both experiment phases. This indicates that pitch errors were corrected more strongly than errors in roll [26, 31, 41], which, in turn, led to better performance in the pitch axis. The higher gains in pitch compared to roll were expected, as human operators usually perform better in pitch than in roll in dual-axis tracking tasks [27–31]. This was also the case in the current experiment as indicated by the tracking performance shown in Fig. 10. At the start of the retention tests, K_p was instantaneously and considerably higher than at the end of training, especially for group 3, similar to the instantaneous increases in control activity in Fig. 11. A clear "sawtooth shape" was exhibited in K_p, both in the training and retention phases, where, similar to the control input, the effect was more prominent in pitch than in roll and more apparent for groups 2 and 3 compared to group 1. This "sawtooth shape" was caused by the changes in control effort throughout the training sessions and retention tests resulting from changing motivation and concentration levels.

While a slight decrease in T_L throughout the training phase was to be expected from earlier training studies [26, 41], Figures 14(c) and 14(d), respectively, show a very large spread in the pitch and roll group averages of T_L throughout the entire experiment, where the spread is even more extreme in pitch than in roll. As a result, most learning curves have been omitted due to their very low correlation coefficients. Especially early on in training, extreme T_L values were observed, often in pitch. In both the pitch and roll axes, the values obtained in the retention phase do not seem to be much different from the values obtained at the end of training.

Unrealistically high values for τ_e were estimated in both pitch and roll in the early stages of the training phase, as shown in Figs. 15(a) and 15(b), respectively. After excluding these experiment runs from the learning curve fitting process, as the high response delays estimated for these runs were considered artefacts, the fitted learning curves with relatively high correlation coefficients showed that training still resulted in a steady decrease in τ_e in both pitch and roll. At the end of training, average values of around 0.31 and 0.34 were achieved in pitch and roll, respectively, with slightly larger between-group differences in roll than in pitch. These values for τ_e are typical for tracking tasks with unpredictable quasi-random target signals in which the controlled element dynamics approximate a double integrator [36]. The slightly lower τ_e in pitch compared to roll was to be expected, as a lower τ_e is related to better performance [26, 41]. Figures 15(a) and 15(b) also reveal a marked difference in skill development between the groups. Whereas for groups 1 and 2 τ_e in the retention phase remained similar to its end-of-training value, group 3 continued to improve (i.e., lower) its response delay throughout the retention tests.

Examining the estimated models for the neuromuscular dynamics, Figs. 15(c) and 15(d) show that the neuromuscular frequency ω_{nm} de-



(a) Pitch pilot gain **(b) Roll pilot gain**

(c) Pitch lead time constant **(d) Roll lead time constant**

Fig. 14. Average pitch and roll equalization parameters with corresponding learning curves.

Table 11. Learning curve parameters for human operator gain and lead time constant

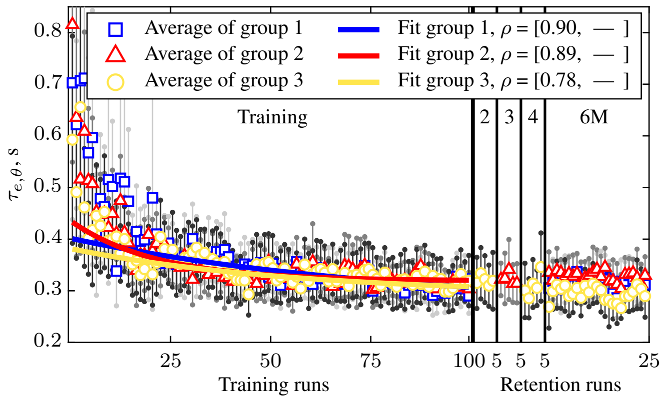
K_p, −	Pitch θ						Roll ϕ					
	Training phase			Retention phase			Training phase			Retention phase		
	p_0, −	p_a, −	$F(\times 10^{-2})$	p_0, −	p_a, −	$F(\times 10^{-2})$	p_0, −	p_a, −	$F(\times 10^{-2})$	p_0, −	p_a, −	$F(\times 10^{-2})$
Group 1	2.10	2.97	5.98	4.24	2.93	21.20	1.25	1.98	8.37	2.47	1.77	25.57
Group 2	1.69	3.23	2.36	n/a	n/a	n/a	1.08	2.16	2.78	2.27	1.86	11.83
Group 3	2.53	7.12	0.24	5.86	4.01	0.41	1.74	2.33	1.96	3.98	2.46	32.43

T_L, s	Pitch θ						Roll ϕ					
	Training phase			Retention phase			Training phase			Retention phase		
	p_0, s	p_a, s	$F(\times 10^{-2})$	p_0, s	p_a, s	$F(\times 10^{-2})$	p_0, s	p_a, s	$F(\times 10^{-2})$	p_0, s	p_a, s	$F(\times 10^{-2})$
Group 1	n/a	n/a	n/a	n/a	n/a	n/a	1.30	1.00	4.42	0.67	1.11	48.81
Group 2	n/a	n/a	n/a	1.08	20.76	0.049	1.26	0.89	1.56	n/a	n/a	n/a
Group 3	n/a	n/a	n/a	0.73	1.27	36.62	n/a	n/a	n/a	n/a	n/a	n/a

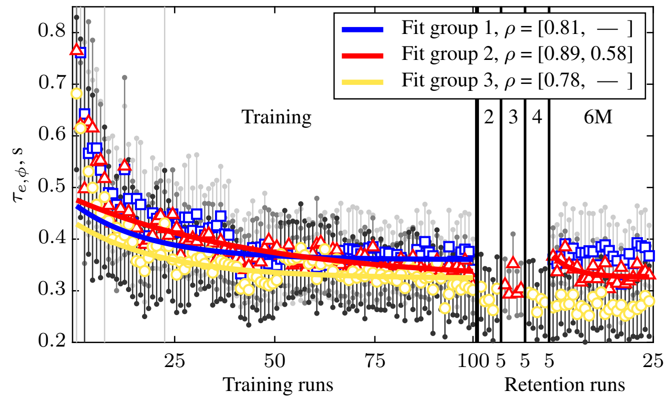
creased throughout training in both pitch and roll for all three experiment groups. Although ω_{nm} was somewhat higher in pitch than in roll, the between-group differences were also larger in pitch compared to roll. The neuromuscular frequencies ω_{nm} in the retention phase were similar to those at the end of training. The decreases observed during training were seen to continue throughout the retention phase. Figures 15(e) and 15(f) present the neuromuscular damping ratio ζ_{nm} in pitch and roll, respectively. No clear learning effects were observed. Throughout both the training and retention phases, extreme group averages higher than 1.0 were seen, especially for group 1, and even a little more extreme in the pitch axis compared to the roll axis. Still, the extreme values seemed to slightly decrease throughout training. Due to the large variances in the groups' averages throughout the training and retention phases, learning curves have been omitted due to the very low Pearson's correlation coefficients.

E. Crossover Frequencies and Phase Margins

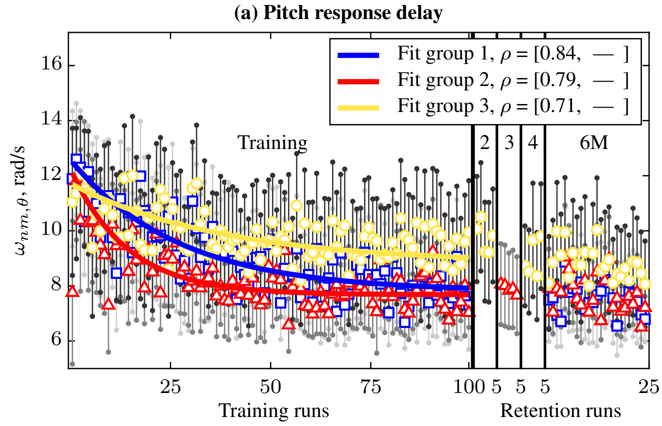
Crossover frequencies and phase margins, as estimated in the frequency domain, using the estimated open-loop dynamics in the separate axes, are provided in Fig. 16. The parameters of the corresponding learning curves for the crossover frequency are listed in Table 13. As shown in Figs. 16(a) and 16(b), crossover frequency increased throughout the training phase for all three groups from around 1.17 rad/s at the start of training to around 1.61 rad/s at the end of training in both the pitch and roll axes. This result is consistent with the improved tracking performance shown in Fig. 10. In the retention phase, the crossover frequencies of groups 1 and 2 in the pitch axis were very similar to their end-of-training values, whereas in the roll axis they decreased to around 1.52 rad/s. The spread in crossover frequency during the retention phase was also slightly larger than at the end of training. The crossover frequencies of group 3, on the other hand, increased in the retention phase



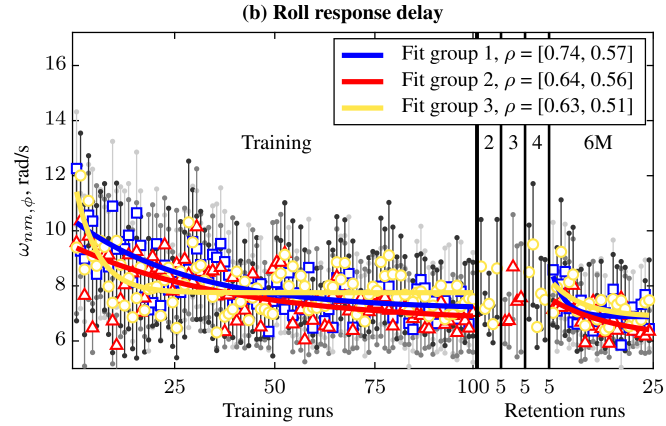
(a) Pitch response delay



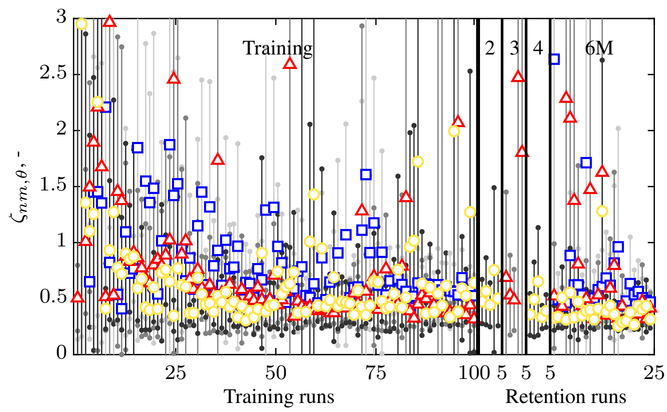
(b) Roll response delay



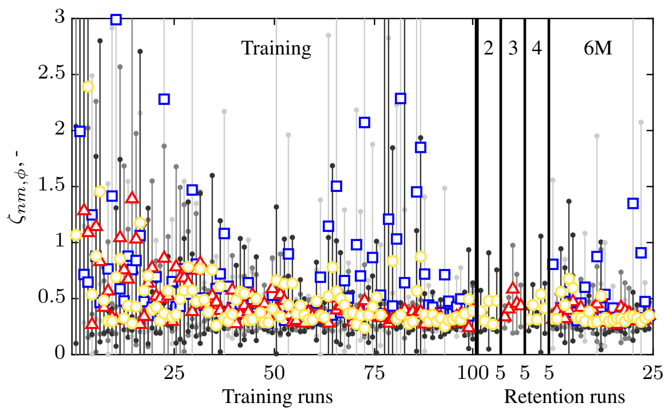
(c) Pitch neuromuscular frequency



(d) Roll neuromuscular frequency



(e) Pitch neuromuscular damping ratio



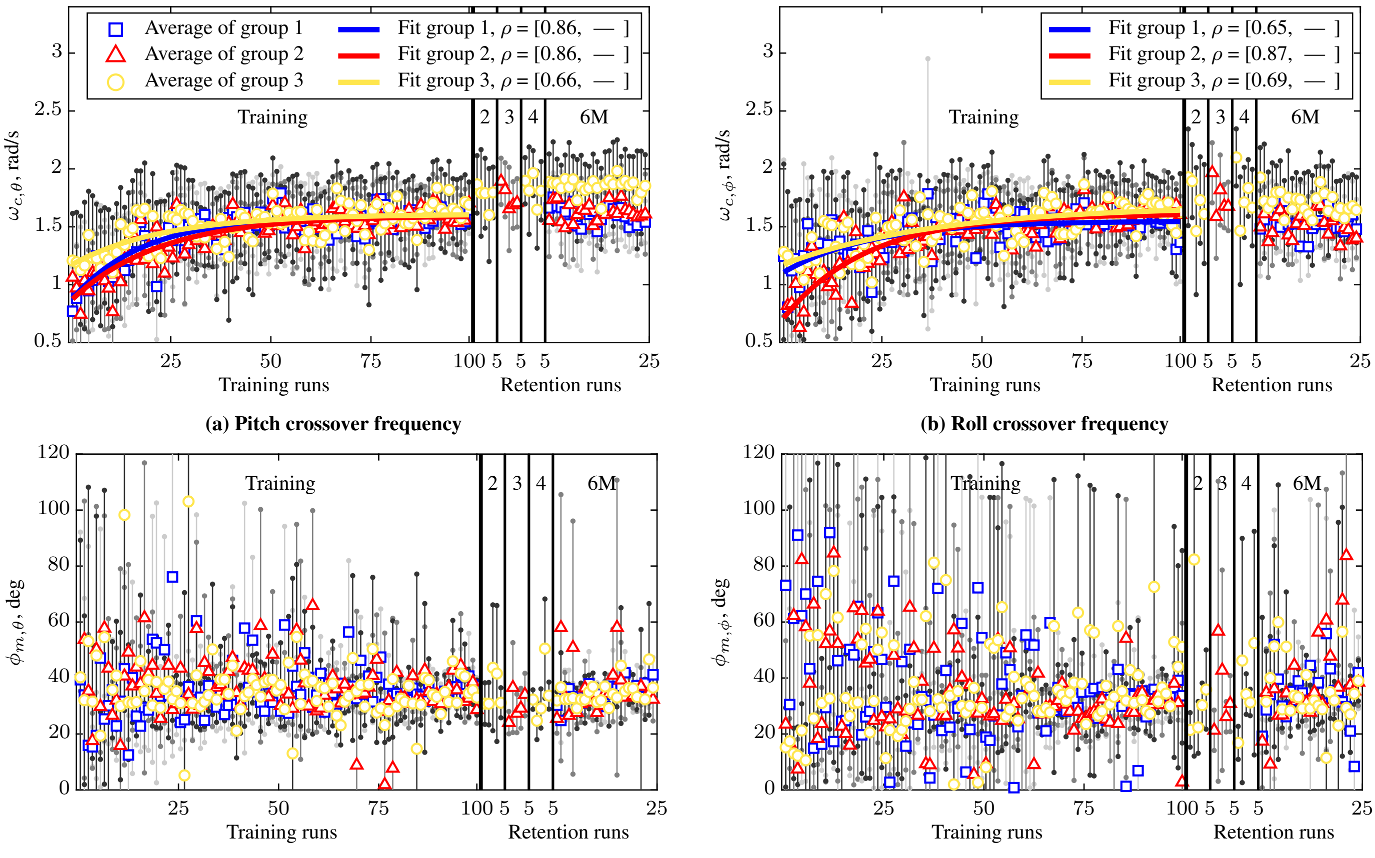
(f) Roll neuromuscular damping ratio

Fig. 15. Average pitch and roll limitation parameters with corresponding learning curves.

Table 12. Learning curve parameters for human operator response delay and neuromuscular frequency

τ_e, s	Pitch θ			Roll ϕ					
	Training phase			Training phase			Retention phase		
	p_0, s	p_a, s	$F(\times 10^{-2})$	p_0, s	p_a, s	$F(\times 10^{-2})$	p_0, s	p_a, s	$F(\times 10^{-2})$
Group 1	0.40	0.27	1.22	0.47	0.36	5.18	n/a	n/a	n/a
Group 2	0.44	0.32	4.59	0.48	0.32	2.08	0.38	0.33	17.72
Group 3	0.38	0.29	1.76	0.43	0.32	4.38	n/a	n/a	n/a

ω_{nm}, rad/s	Pitch θ			Roll ϕ					
	Training phase			Training phase			Retention phase		
	p_0, rad/s	p_a, rad/s	$F(\times 10^{-2})$	p_0, rad/s	p_a, rad/s	$F(\times 10^{-2})$	p_0, rad/s	p_a, rad/s	$F(\times 10^{-2})$
Group 1	12.67	7.80	3.63	10.41	7.13	3.35	8.86	6.88	25.81
Group 2	12.42	7.68	6.82	9.45	6.66	2.41	7.55	6.08	5.87
Group 3	11.76	8.82	2.61	12.13	7.75	16.28	8.24	6.87	11.63



(a) Pitch crossover frequency

(b) Roll crossover frequency

(c) Pitch phase margin

(d) Roll phase margin

Fig. 16. Average pitch and roll crossover frequencies and phase margins with corresponding learning curves.

Table 13. Learning curve parameters for crossover frequency

ω_c, rad/s	Pitch θ			Roll ϕ		
	Training phase			Training phase		
	p_0, rad/s	p_a, rad/s	$F(\times 10^{-2})$	p_0, rad/s	p_a, rad/s	$F(\times 10^{-2})$
Group 1	1.23	1.59	4.42	1.20	1.55	5.03
Group 2	1.13	1.62	3.25	0.88	1.64	3.62
Group 3	1.29	1.65	5.59	1.27	1.77	1.49

to values around 1.90 rad/s in pitch and 1.69 rad/s in roll. These relatively high crossover frequencies of group 3 compared to groups 1 and 2 are consistent with the higher control activity of group 3 during the retention phase, as shown in Fig. 11.

Although an increase in phase margin throughout the training phase was expected from earlier training studies [30, 45, 86], no clear learning trends were observed in the phase margin in the current experiment, as shown in Figs. 16(c) and 16(d). Due to the large spread in the data, learning curves have been omitted. However, it was observed that the phase margin during the retention phase was similar to its end-of-training values of around 35 deg in both pitch and roll.

F. Human Operator Remnant

In Fig. 17 an example of a human operator remnant spectrum from the current experiment is shown. It is demonstrated that the remnant model defined in Eq. 9 does indeed model the power spectra of the remnant obtained in the current experiment well.

The human operator remnant parameters (K_n, ω_n and ζ_n) together with their fitted learning curves are shown in Fig. 18. Figures 18(a) and 18(b) show that the human operator remnant gain decreased throughout training for both pitch and roll, respectively, which is consistent with the increase in VAF shown in Fig. 13. Early in the training phase, the remnant gain of group 1 was considerably higher than those of groups 2 and 3 in both the pitch and roll axes (around 0.058 for group 1 and 0.038 for groups 2 and 3 in the pitch axis, and around 0.061 for group 1 and 0.035 for groups 2 and 3 in the roll axis). This difference disappeared at the end of training, where very similar remnant gains were observed for all groups (around 0.016 in pitch and 0.012 in roll), as shown in Table 14. A possible explanation for this could be that during the early training runs, group 1 gave larger control inputs in both pitch and roll compared to groups 2 and 3, whereas the end-of-training control input values were very similar for all three groups (see Fig. 11). Remnant is known to scale with overall control activity [87]. Whereas the individual groups exhibited similar remnant gains in the pitch and roll axes at the start of training, towards the end of the training phase the remnant gains were higher in pitch than in roll, which was also the case for the control inputs. The instantaneous increase in remnant gain at the start of the retention tests, followed by a decrease throughout the retention runs in both the pith and roll axes, was also very similar to

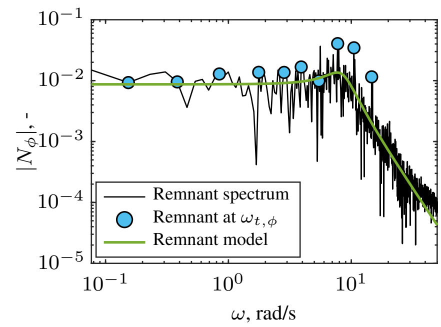
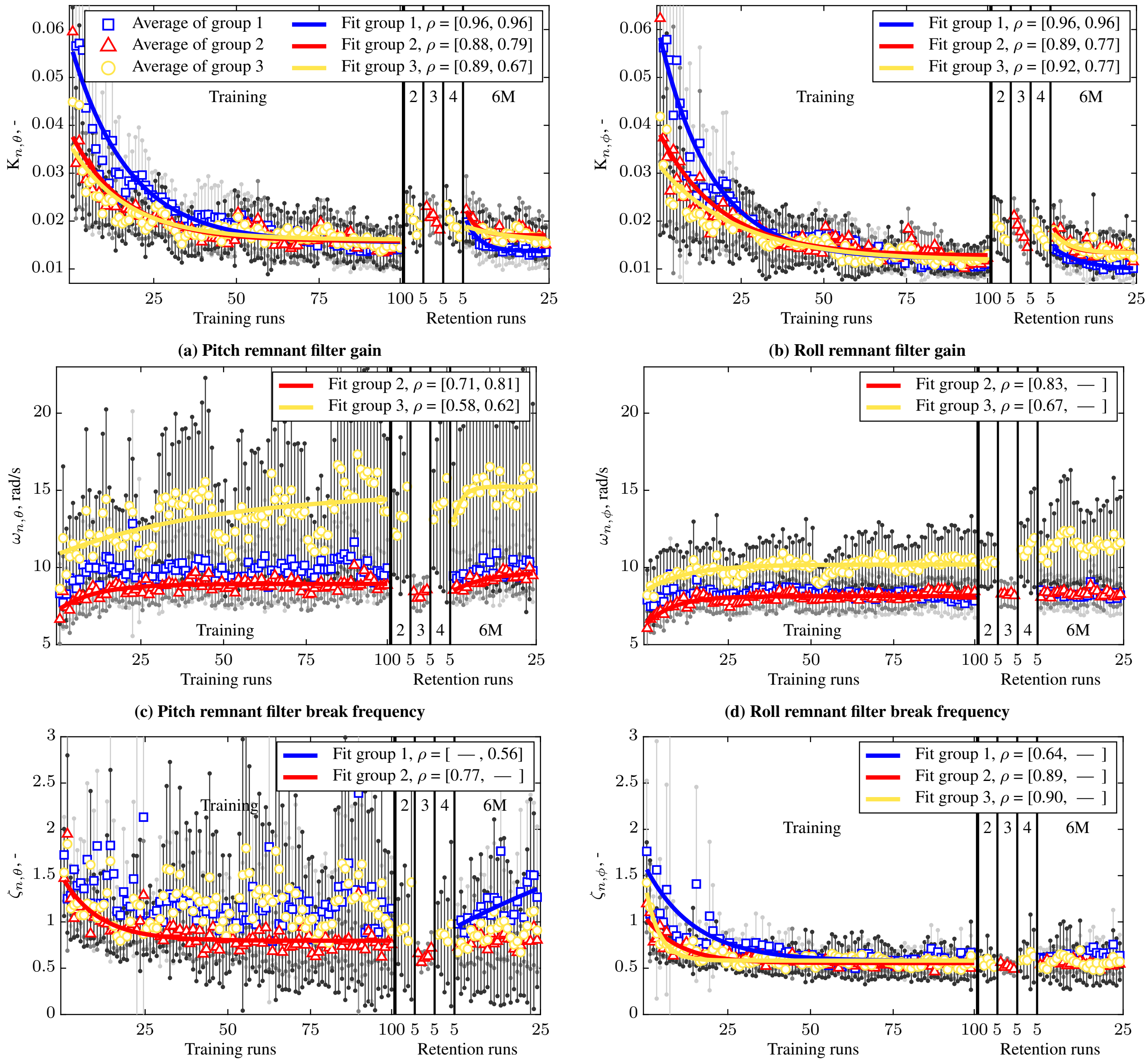


Fig. 17. Example human operator remnant spectrum (participant 17, training run 93, roll axis).



(a) Pitch remnant filter gain

(b) Roll remnant filter gain

(c) Pitch remnant filter break frequency

(d) Roll remnant filter break frequency

(e) Pitch remnant filter damping ratio

(f) Roll remnant filter damping ratio

Fig. 18. Average pitch and roll remnant filter parameters with corresponding learning curves.

the changes observed in the pitch and roll control inputs throughout the retention phase (see Fig. 11).

Figures 18(c) and 18(d) show a marked difference between the groups in their pitch and roll remnant filter break frequencies ω_n. Whereas the values were slightly lower for group 1 compared to group 2, group 3 showed considerably higher remnant filter break frequencies than the other two groups. The high ω_n values of group 3 were the result of four participants with considerably higher remnant filter break frequencies than the other participants in group 3 and the participants of groups 1 and 2. Closer inspection of the remnant spectra of these participants showed that these said spectra could perhaps better be modeled by a fourth-order low-pass filter with damping, as opposed to the remnant spectra of the other participants, which were generally accurately modeled by the selected third-order low-pass filter as illustrated in Fig. 17. This indicates how difficult it is to find a universal model that could accurately model the remnant of nearly all human operators. Whereas Fig. 18(f) shows that the roll remnant filter damping ratio approached an asymptotic value of around 0.57, which was retained throughout the retention phase, a considerable larger spread in the pitch remnant filter damping ratio is seen in Fig. 18(e). This result is consistent with the larger spread in pitch observed for the remnant filter break frequency.

G. Human Operator Crossfeed

1. *Crossfeed Contribution*

To gain insight into both the amount of crossfeed present in the dual-axis tracking task performed by task-naive participants, as well as the 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 Sec. II. These contributions are shown as a fraction of the total variance in Figs. 19 and 20 for the tracking error and control input, respectively. In these figures, the contribution from the target signal of the principal axis is

Table 14. Learning curve parameters for remnant filter parameters

$K_n, -$	Pitch θ						Roll ϕ					
	Training phase			Retention phase			Training phase			Retention phase		
	$p_0, -$ $(\times 10^{-2})$	$p_a, -$ $(\times 10^{-2})$	$F(\times 10^{-2})$	$p_0, -$ $(\times 10^{-2})$	$p_a, -$ $(\times 10^{-2})$	$F(\times 10^{-2})$	$p_0, -$ $(\times 10^{-2})$	$p_a, -$ $(\times 10^{-2})$	$F(\times 10^{-2})$	$p_0, -$ $(\times 10^{-2})$	$p_a, -$ $(\times 10^{-2})$	$F(\times 10^{-2})$
Group 1	5.77	1.55	5.58	21.76	2.19	1.37	6.13	1.21	6.13	1.49	0.98	10.57
Group 2	3.90	1.58	5.93	2.41	1.70	27.97	3.94	1.28	5.23	2.02	1.34	23.40
Group 3	3.61	1.61	5.70	1.91	1.58	7.13	3.22	1.18	4.13	1.82	1.33	16.82

ω_n, rad/s	Pitch θ						Roll ϕ		
	Training phase			Retention phase			Training phase		
	p_0, rad/s	p_a, rad/s	$F(\times 10^{-2})$	p_0, rad/s	p_a, rad/s	$F(\times 10^{-2})$	p_0, rad/s	p_a, rad/s	$F(\times 10^{-2})$
Group 2	7.17	8.93	8.49	8.38	9.74	8.73	6.17	8.14	11.34
Group 3	10.83	14.95	2.03	11.79	15.24	33.08	8.44	10.19	7.64

$\zeta_n, -$	Pitch θ						Roll ϕ		
	Training phase			Retention phase			Training phase		
	$p_0, -$	$p_a, -$	$F(\times 10^{-2})$	$p_0, -$	$p_a, -$	$F(\times 10^{-2})$	$p_0, -$	$p_a, -$	$F(\times 10^{-2})$
Group 1	n/a	n/a	n/a	0.94	2.72	1.06	1.64	0.58	6.51
Group 2	1.51	0.80	8.59	n/a	n/a	n/a	1.08	0.55	9.19
Group 3	n/a	n/a	n/a	n/a	n/a	n/a	1.52	0.58	21.45

shown in blue, the crossfeed contribution is shown in red and the remnant contribution is shown in green. Solid lines indicate the developments in the group medians over the course of the experiment, while transparent colored areas define the boundaries of the maximum and minimum contributions throughout the experiment.

Figure 19 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 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, the crossfeed contribution remained relatively constant, whereas the contribution from the remnant signal decreased and that from the target signal increased. These changes in the contributions from the target and remnant signals were consistent with the increase in human operator linearity, as illustrated by the increase in VAF shown in Fig. 13. At the end of training, all groups exhibited a slightly larger contribution from the target signal compared to the remnant signal in the pitch axis (around 52% opposed to around 43% for the 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 target signal (around 37% opposed to around 55% for the target and remnant signals, respectively), which was consistent with the lower VAF in the roll axis compared to the pitch axis as shown in Fig. 13. The values of the remnant contribution to the total error variance confirmed the generally accepted effect 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 contributions up to 40% for fully task-naive human operators and contributions around 20% for more experienced operators are generally reported [31, 41, 88].

An instantaneous increase in the remnant contribution was observed at the start of the retention phase, whereas the contribution of the target signal showed an instantaneous decrease, although no considerable decrease in VAF during the retention phase was observed in Fig. 13. The contribution of the off-axis target signal remained similar to its contribution during the training phase. Figure 19 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. 10). Throughout the retention phase, the contribution of the remnant signal decreased again while that of the target signal increased.

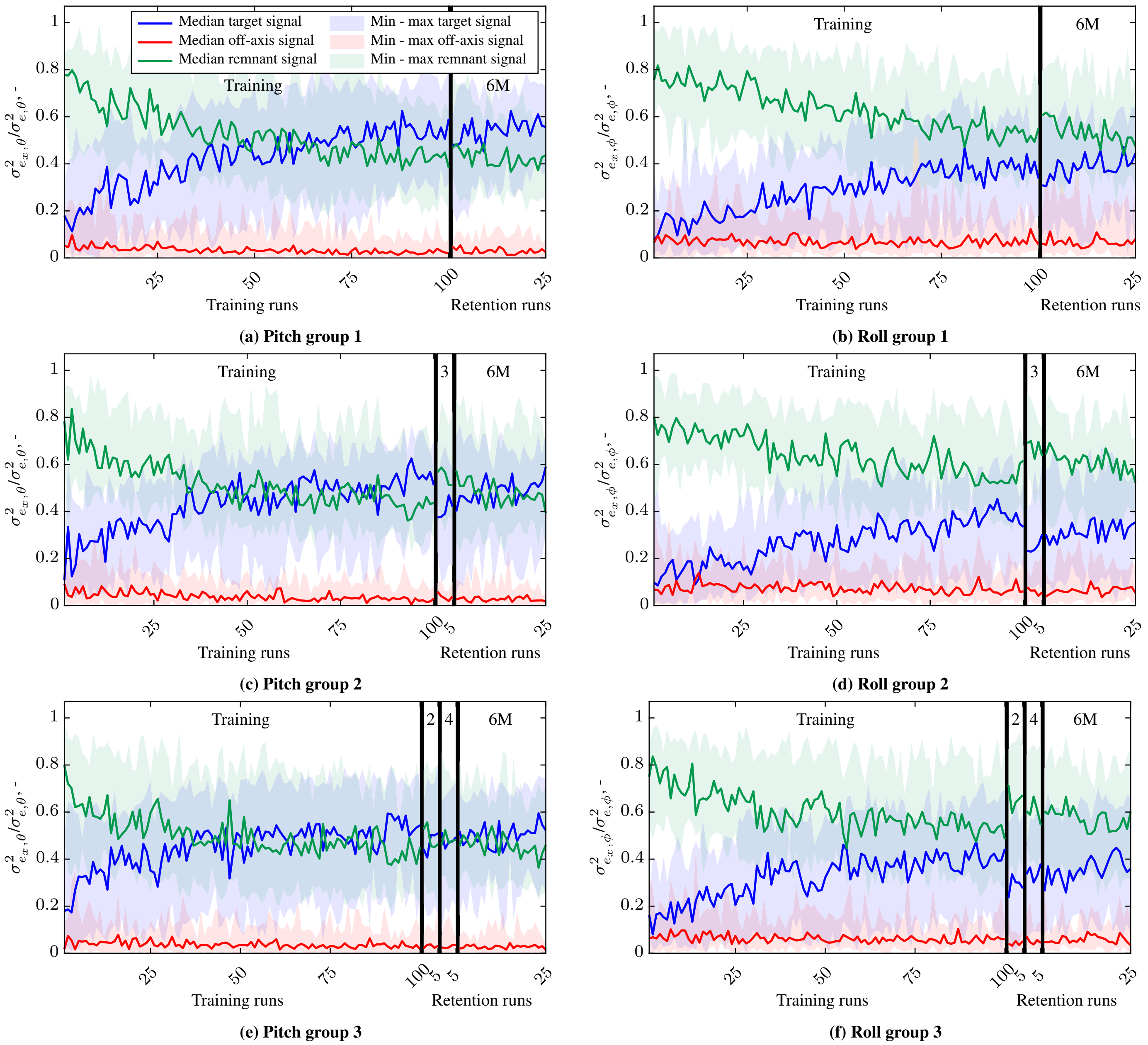
In the decomposition of the control input variance in Fig. 20, similar observations can be made as for the tracking error variance. Early on in the training phase, the human operator remnant was responsible for the largest contribution to the total control input variance as it composed around 80% of the total variance in the pitch axis and around 85% in the roll axis. The contributions of the principal 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 target contribution increased while the remnant contribution decreased. Similar to the error variance, the contribution of the off-axis target signal remained relatively constant throughout training. Towards the end of training, both the 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 target signal around 28%.

Whereas at the start of the retention phase, groups 2 and 3 showed an instantaneous increase and decrease in the remnant and 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 [87], as already mentioned in Sec. IIIF, 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 contribution of the off-axis target signal was again similar to its contribution in the training phase.

2. *Crossfeed Dynamics*

Using the FC method described in Sec. II, frequency response estimates were obtained for the human operator error response H_{p_e} and crossfeed response H_{p_c}. An example of these estimates is presented in Fig. 21. On the left, the blue asterisks present the FC estimates of the human operator error response H_{p_e} with the solid orange lines presenting the error models identified using the Genetic MLE method. On the right, the green asterisks present the FC estimates of the human operator crossfeed response H_{p_c}. The solid orange lines again illustrate the modeled error response.

When comparing the modeled error response with the FC estimates of H_{p_c}, it can be seen that the crossfeed dynamics seem very similar to those of the error response, but with a lower gain. The crossfeed dynamics can thus indeed be modeled well with the crossfeed model proposed in Eq. 11.



(a) **Pitch group 1**

(b) **Roll group 1**

(c) **Pitch group 2**

(d) **Roll group 2**

(e) **Pitch group 3**

(f) **Roll group 3**

Fig. 19. Normalized target, off-axis and remnant contributions to tracking error variance.

Unfortunately, the frequency response estimates for H_{p_c} were relatively inaccurate. Describing function estimates are in principle biased by the remnant still present at the excitation frequencies, but this bias can usually be neglected for signal-to-noise ratios above 5, for which accurate describing function identification can generally be achieved [43]. However, due to the small off-axis target signal contributions involved (see Figs. 19 and 20), the signal-to-noise ratio at the excitation frequencies was generally lower than 5. This resulted in unreliable identification of the describing functions, especially at lower frequencies, as sudden jumps in the frequency response estimates were observed. Therefore, most crossfeed response parameters were kept equal to their error response values and only the crossfeed gain was estimated using the Genetic MLE method, which was also used for the parameter estimation of the human operator error response in Sec. IIID.

Figure 22 presents the estimated crossfeed gains. The crossfeed gains were averaged between five consecutive experiment runs for illustration purposes. The crossfeed gains did not change much throughout the training or retention phase. Noteworthy is that the roll-to-pitch crossfeed gains were mostly positive, whereas the pitch-to-roll crossfeed gains were mostly negative, which is consistent with previous crossfeed research [31]. This means that human operators were unable to fully decouple the tracking tasks in the pitch and roll axes, as for a positive pitch input u_θ, the majority of participants gave a small coupled negative roll input u_ϕ. As a result, a -180° phase shift was observed in the crossfeed response (see Fig. 21). Furthermore, the pitch-to-roll crossfeed gains had higher absolute values than the roll-to-pitch crossfeed gains. This indicates that there was a stronger component of pitch in the roll axis than vice versa, which corresponds with the larger off-axis target signal contributions to the error and control variances in the roll axis compared to the pitch axis (see Figs. 19 and 20) and is also consistent with previous research [31].

To gain insight into the influence of the addition of the human operator crossfeed response on the ability to describe human operator control behavior by quasi-linear human operator models, the VAF obtained for the modeled human operator control input including the human operator error and crossfeed responses (see Fig. 8) was compared to the VAF obtained for the modeled human operator control input solely including the error response (see Fig. 1). The results of this comparison

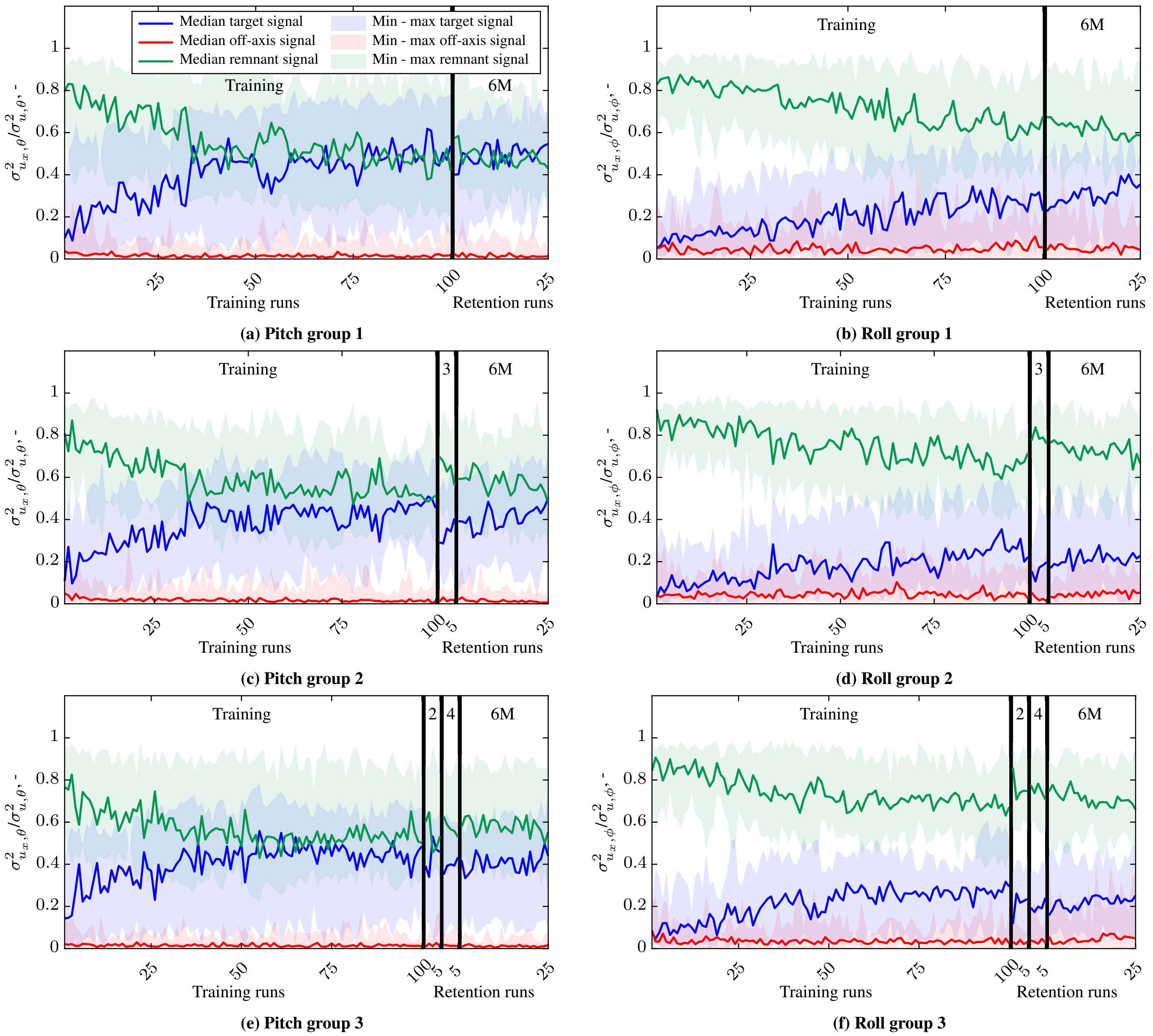


Fig. 20. Normalized target, off-axis and remnant contributions to control input variance.

are presented in Fig. 23. In this figure, the colored boxplots illustrate the VAF of the modeled control input including the crossfeed response and the gray boxplots show the VAF of the modeled control input without the crossfeed response. The VAF values were averaged between five consecutive experiment runs, as was also done for the crossfeed gains. For illustration purposes, Fig. 23 shows the VAF values of the human operator models at the start and end of the training phase, as well as at the start and end of the 6-month retention test and for the runs of the 2-month, 3-month and 4-month retention tests. The boxplots of the early training runs are not shown in their entirety to increase the visibility of the differences in VAF between the human operator models with and without crossfeed later on in the training phase and in the retention phase.

Overall, the human operator crossfeed response resulted in a small, but consistent increase in the VAF. The mean increase in VAF throughout the different experiment phases is listed in Table 15. This table shows that, except for at the start of the training phase, a larger increase in VAF was obtained in the roll axis compared to the pitch axis, which is consistent with the higher absolute values of the pitch-to-roll crossfeed gains in Fig. 22. Although the crossfeed contribution to the overall VAF may be small (around 0.5%-1.5% increase in the pitch axis and around 1.0%-3.0% increase in the roll axis), statistical analysis results in Table 16 show that all increases were statistically significant (dependent t tests and Wilcoxon signed-rank tests for normally and non-normally distributed data, respectively). This significant crossfeed contribution is, however, smaller than the 1%-5% increase observed in previous research [31]. Whereas in this previous study the identified crossfeed dynamics could explain up to 20% of the measured control inputs, in the current experiment a contribution up to 8% to the measured control inputs was found, as discussed earlier. A possible explanation for the smaller crossfeed contribution observed in the current experiment can be found in the experiment apparatus. Compared to the previous experiment investigating crossfeed, the position of the sidestick in the current experiment allowed for the placement of the participant's arm to be more in line with the sidestick's pitch axis. This resulted in a lower likelihood of introducing cross-coupled control inputs compared to in the previous experiment.

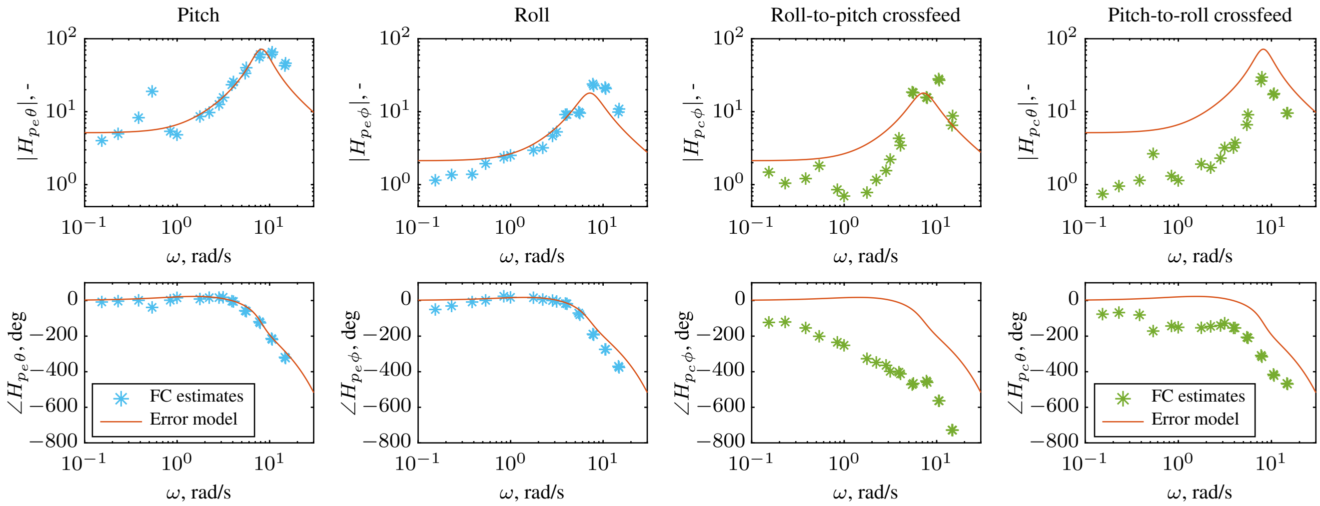
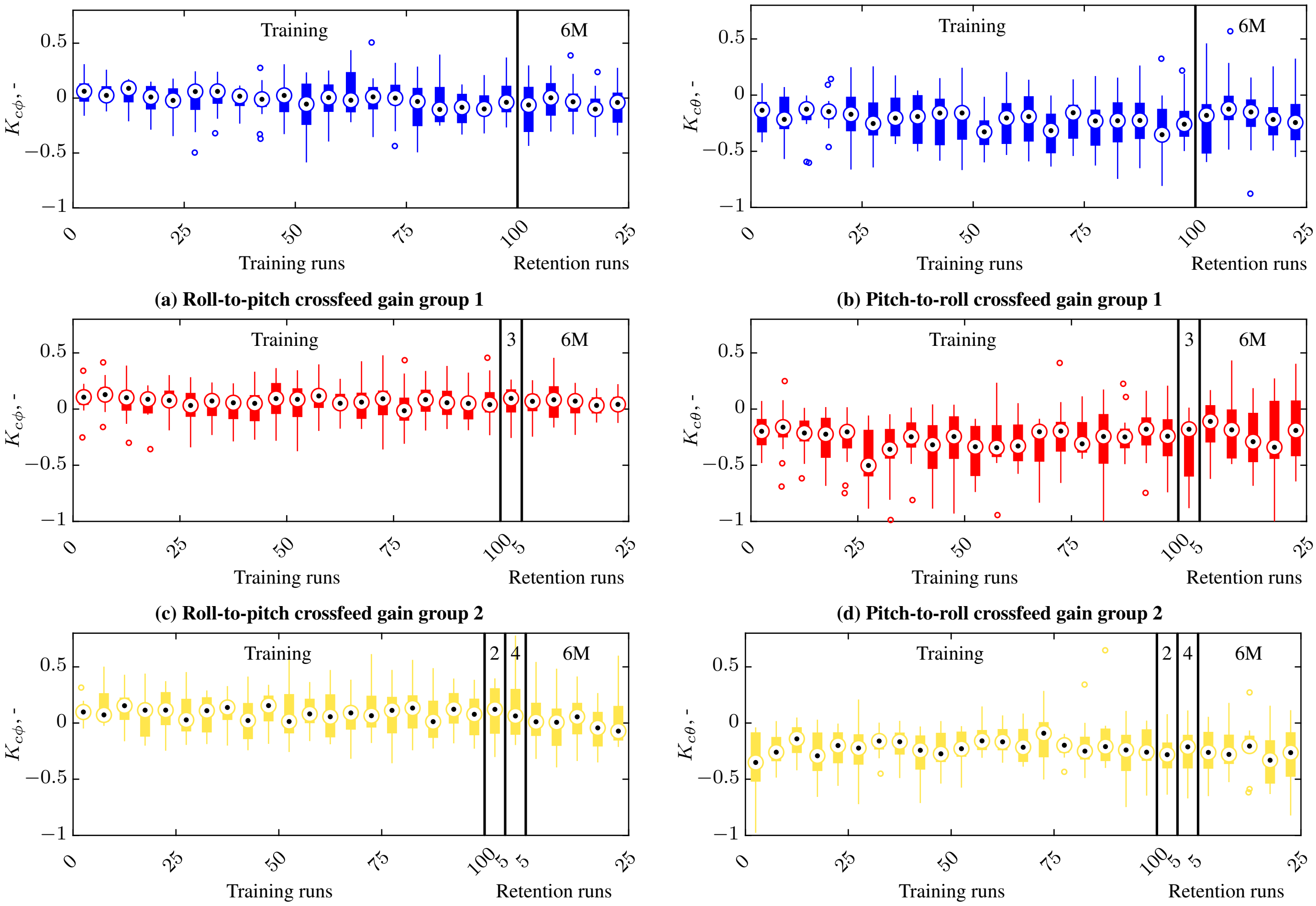


Fig. 21. Example of frequency response estimates for error and crossfeed response, together with estimated error model (participant 19, training run 83).



(a) Roll-to-pitch crossfeed gain group 1

(b) Pitch-to-roll crossfeed gain group 1

(c) Roll-to-pitch crossfeed gain group 2

(d) Pitch-to-roll crossfeed gain group 2

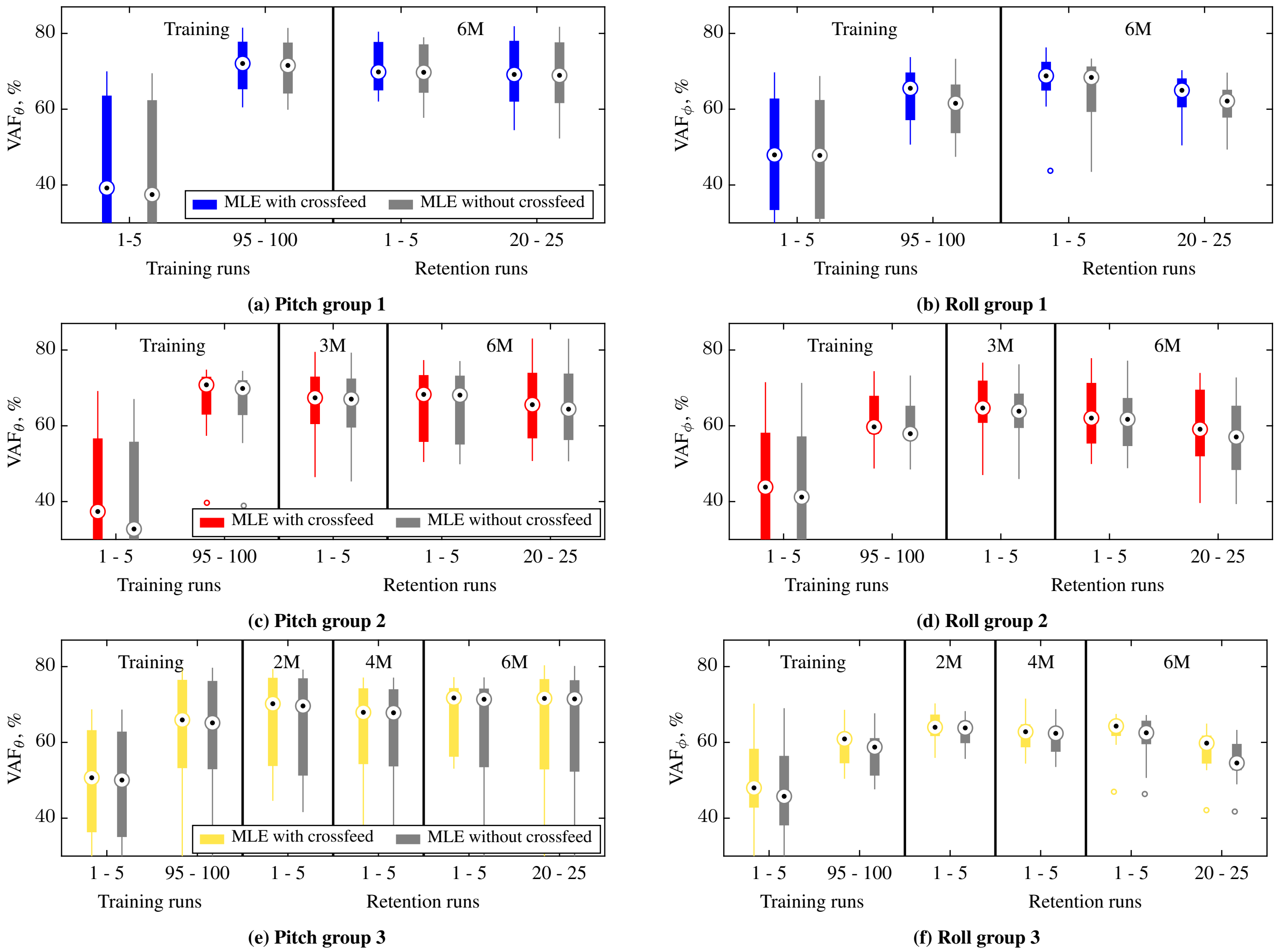
(e) Roll-to-pitch crossfeed gain group 3

(f) Pitch-to-roll crossfeed gain group 3

Fig. 22. Human operator crossfeed gain.

H. Time-Varying Control Behavior

As discussed in Sec. II, time-varying model identification has only been performed on the last five training runs and the retention runs of eleven participants. However, unfortunately the method was not successful for all evaluated runs. Of the 385 tracking runs evaluated in total, model identification of 78 runs was only successful in the pitch axis, 28 runs were only successfully identified in roll and 13 runs were not successfully identified in either of the two axes. Here, successful identification is defined as a condition in which the identified time-varying model leads to a VAF value equal to or higher than the VAF value of the identified time-invariant human operator model. This brings the success rate of the time-varying model identification to 89.4% in pitch and 76.4% in roll. In the tracking runs which could not be successfully identified, human operator control behavior was truly nonlinear, in addition to time-varying. This often occurs in extreme scenarios such as



(a) Pitch group 1 (b) Roll group 1

(c) Pitch group 2 (d) Roll group 2

(e) Pitch group 3 (f) Roll group 3

Fig. 23. Variance Accounted For with and without crossfeed.

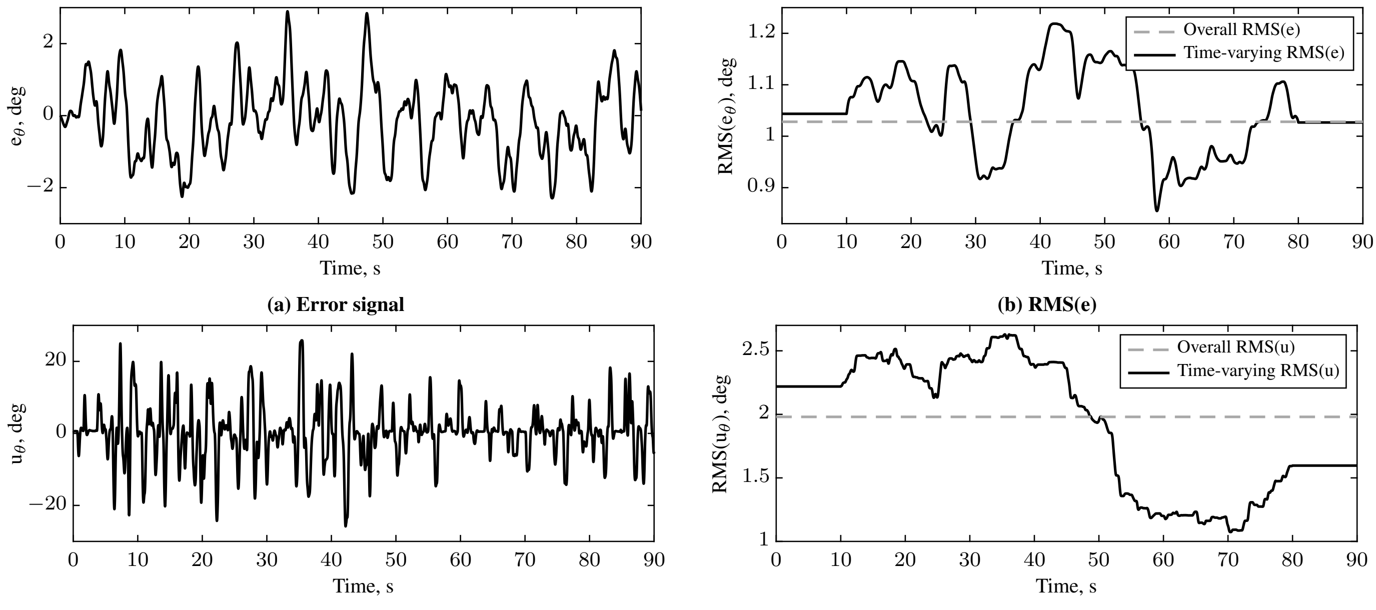
Table 15. Mean increase in VAF due to human operator crossfeed response

ΔVAF, %		Training phase		Retention phase				
				2M Test	3M Test	4M Test	6M Test	
		Runs 1 - 5	Runs 95 - 100	Runs 1 - 5	Runs 1 - 5	Runs 1 - 5	Runs 1 - 5	Runs 21 - 25
Pitch θ	Group 1	2.0%	1.1%				1.2%	0.9%
	Group 2	3.1%	0.6%		0.6%		0.5%	0.4%
	Group 3	1.2%	0.6%	1.1%		1.0%	0.9%	1.0%
Roll ϕ	Group 1	1.5%	3.1%				2.1%	2.6%
	Group 2	1.6%	1.8%		1.3%		1.3%	2.0%
	Group 3	2.2%	2.0%	1.1%		1.1%	1.7%	2.8%

Table 16. Statistical analysis results for VAF increase due to human operator crossfeed response

VAF		Training phase		Retention phase				
				2M Test	3M Test	4M Test	6M Test	
		Runs 1 - 5	Runs 95 - 100	Runs 1 - 5	Runs 1 - 5	Runs 1 - 5	Runs 1 - 5	Runs 21 - 25
Pitch θ	Group 1	**	*				**	*
	Group 2	**	**[a]		**		**	**
	Group 3	**	**	*		*	**[a]	**[a]
Roll ϕ	Group 1	**	**				**[a]	**
	Group 2	**	**		**		**	**
	Group 3	**	*	**		**	**[a]	*

[a] At least one sample not normally distributed, Wilcoxon signed-rank test applied instead of dependent t test.



(a) Error signal

(b) RMS(e)

(c) Control signal

(d) RMS(u)

Fig. 24. Example time-variance in error and control signal (participant 34, training run 98, pitch axis).

notable control mishaps.

Figure 24 shows an example of pitch error and control signals of a tracking run that was successfully identified using the time-varying MLE method. While Figs. 24(a) and 24(c) show the time traces of the pitch error and control signals, respectively, in Figs. 24(b) and 24(d) the RMS(e) and RMS(u) of these signals are presented as a function of time. These performance measures were calculated in a similar manner as the identification of the time-varying human operator model parameters as explained in Sec. II, i.e., using a sliding 20-second time window. As a result of this 20-second sliding time window approach, the RMS values were constant during the first and last 10 seconds of the run. In both Figs. 24(a) and 24(b), it can be seen that the worst tracking during the run was performed in the middle of the run, as that is when the largest error signal amplitudes, around 3 deg, are shown in Fig. 24(a) and the largest RMS(e) in Fig. 24(b). Also, both Figs. 24(c) and 24(d) indicate that the participant had a higher control activity during the first part of the run than during the second segment, as illustrated by the larger control signal amplitudes during the first 45 seconds of the run in Fig. 24(c) and by the higher RMS(u) in Fig. 24(d).

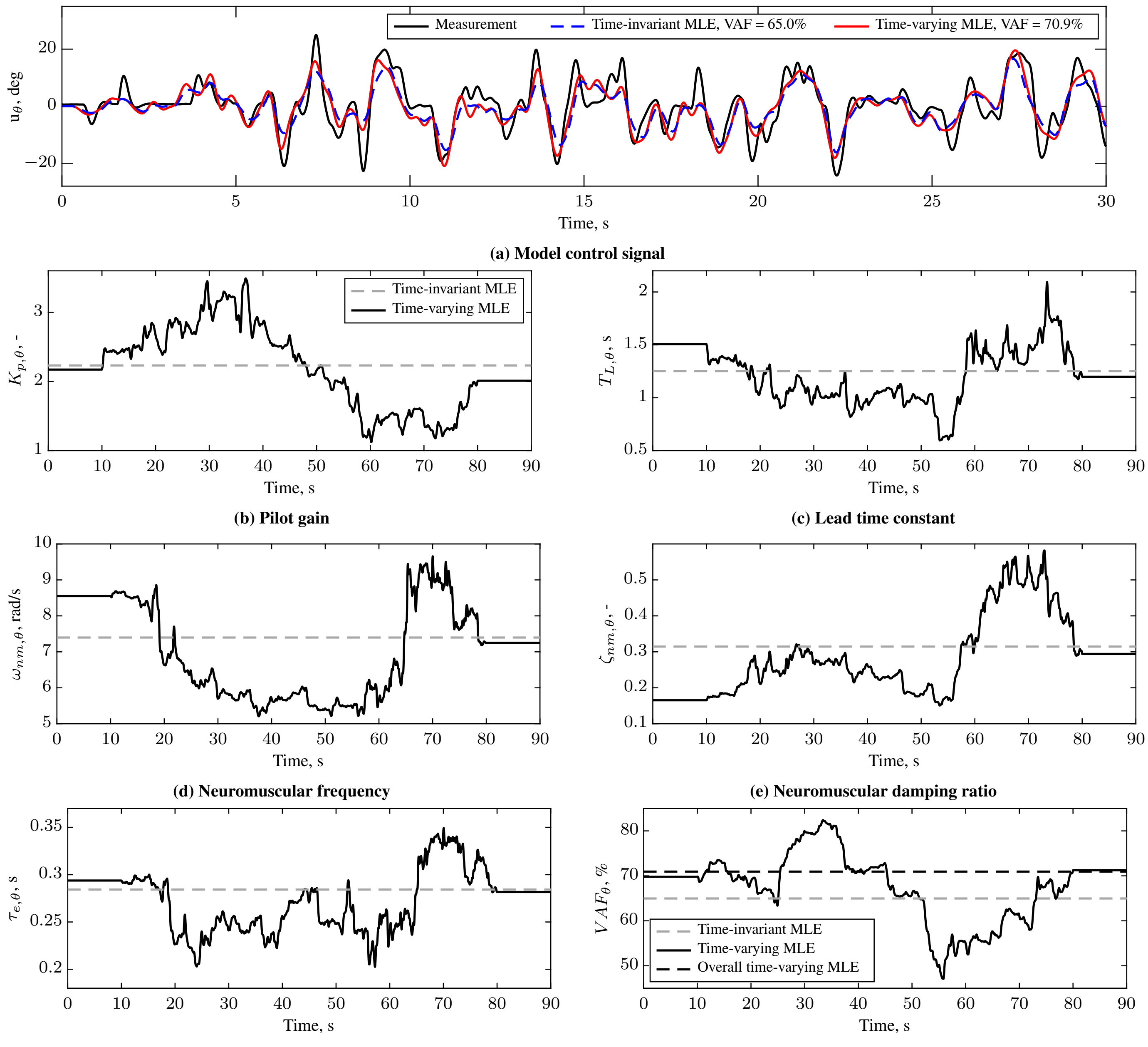
Figure 25 presents the results of the time-varying human operator model identification performed on the tracking run shown in Fig. 24. For illustration purposes, Fig. 25(a) shows the first 30 seconds of the pitch control signal together with the modeled control signals using the time-invariant and time-varying MLE methods. The VAF was higher for the control signal modeled using the time-varying modeling approach. Figure 25(b) presents the human operator gain, where the higher gain values in the first part of the run are consistent with the increased control activity as shown in Figs. 24(c) and 24(d). The time-varying lead time constant in Fig. 25(c) shows, for example, that the participant increased the lead he/she was generating between the 60- and 80-second time marks in an attempt to counteract the increasing tracking error shown in Fig. 24(b). Figures 25(d), 25(e) and 25(f) show that during this increase in tracking error, the neuromuscular frequency, neuromuscular damping ratio and human operator response delay, respectively, increased as well. This is consistent with the known observation that a lower neuromuscular damping ratio and human operator response delay are signs of enhanced task proficiency [26, 41, 45]. As shown in Fig. 25(g), the VAF was higher during the first part of the tracking run than during the second part, indicating that human operator control behavior was more linear early on during the run.

To gain insight into the ability of quasi-linear human operator models with time-varying parameters to describe human operator control behavior, the VAF values obtained for the human operator control input modeled using the time-varying and time-invariant MLE methods were compared. The results, which only consider the tracking runs which were successfully identified using the time-varying MLE method, are presented in Fig. 26. Similar to Fig. 23, the colored boxplots illustrate the VAF of the control input modeled using the time-varying approach, whereas the gray boxplots show the VAF of the control input modeled using the time-invariant method. Due to the small number of participants considered from each experiment group, the results from five consecutive experiment runs were not averaged, as was the case for the crossfeed results in Fig. 23, but simply grouped together.

Figure 26 and Table 17 show that time-varying model identification resulted in an increase in the VAF of around 1% to 3% on average, with occasional VAF increases up to 8%. Overall, these values are consistent with the increased VAF observed during an earlier study identifying time-varying human operator model parameters in a dual-axis tracking task [27]. Similar to the time-invariant identification results, higher VAF values were obtained in the pitch axis compared to the roll axis using the time-varying identification approach, as also consistent with earlier research [27]. Table 18 illustrates that the increase in VAF due to time-varying model identification was significant, but as indicated before, this only holds for around 89% of the pitch axis data and 76% of the roll axis data. Due to the relatively low success rate of the identification method, it was difficult to investigate the influence of a period of non-practice on the time-varying characteristics of manual control behavior.

IV. Discussion

The goal of this paper was to objectively and explicitly quantify the acquisition, decay and retention of skill-based manual control behavior using a cybernetic approach. This was done by analyzing changes in 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-naive participants. All participants were trained under the same conditions and subsequently divided into three groups based on their training performance. These groups differentiated from one another in their retention interval length and the number of retention tests they performed. The first group had their performance re-evaluated after a period of non-practice of six months, while 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 retention interval which ensures that manual control skills do not decay significantly, while simultaneously minimizing the amount of required refresher training,



(a) Model control signal

(b) Pilot gain

(c) Lead time constant

(d) Neuromuscular frequency

(e) Neuromuscular damping ratio

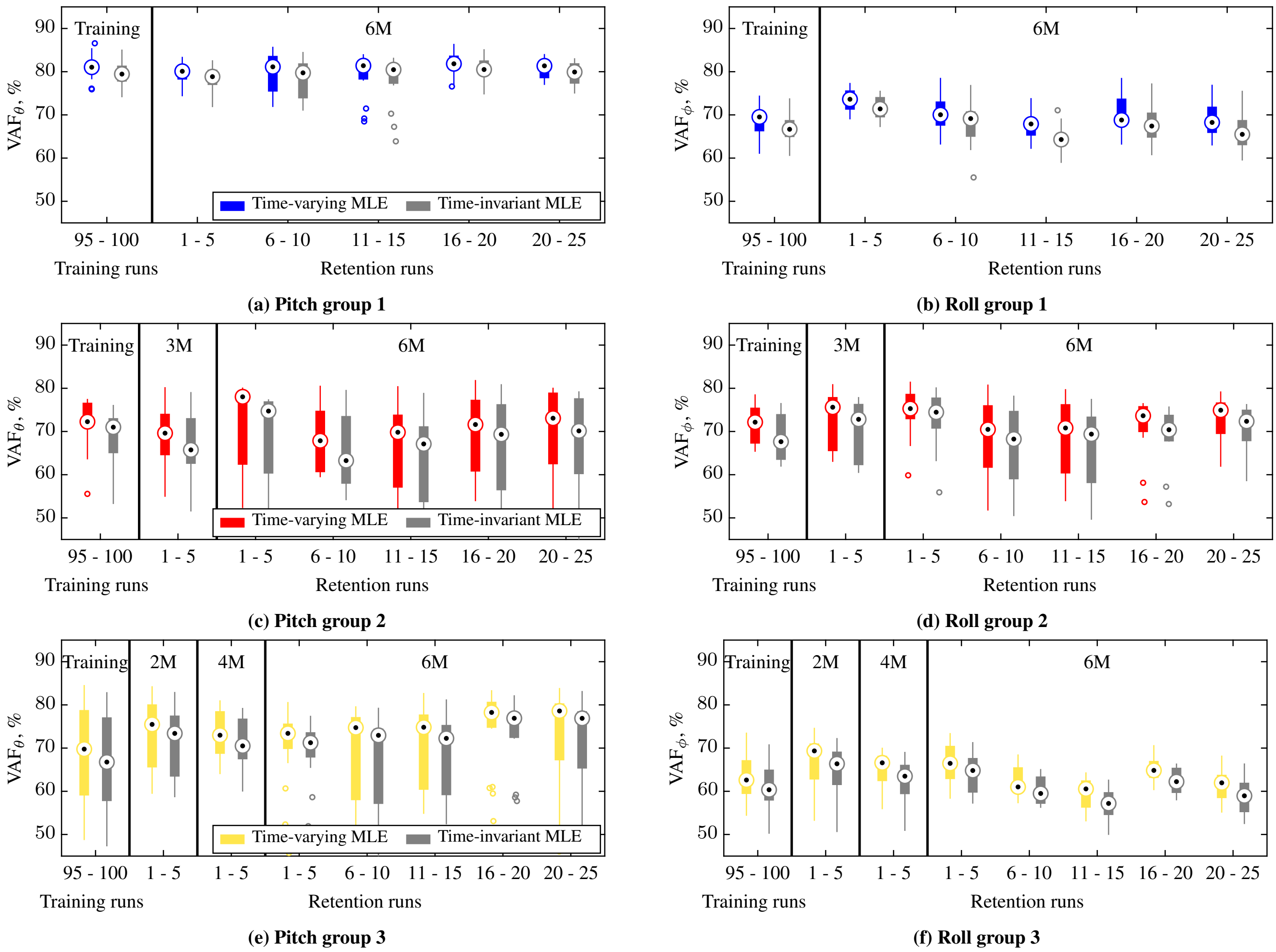
(f) Response delay

(g) Variance Accounted For

Fig. 25. Example time-varying human operator model (participant 34, training run 98, pitch axis).

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 control behavior, human operator modeling techniques were applied. In addition, to contribute to the investigation into more realistic control tasks than the tightly controlled and stationary conditions for which the current state-of-the-art cybernetics methods are only truly valid, the presence of crossfeed and its training and retention effects were investigated by attempting to model the human operator crossfeed response. Additionally, the time-varying characteristics of manual control behavior were studied by identifying linear, time-invariant human operator models, but with time-varying parameters.

Based on the findings of previous training experiments [26,41,45], 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. From the human operator modeling results, the human operator gain showed a significant increase with training, as expected, since a higher human operator gain is in general related to better performance [26, 41]. The human operator response delay showed a markedly lower value at the end of training, as it had reduced by around 120 ms throughout the training phase in both pitch and roll. Also, an increase in VAF indicated increased human operator linearity at the end of the training phase, while an increase in the crossover frequency showed that towards the end of training participants were able to track the pitch and roll target signals up to higher frequencies. However, no clear learning effects were observed in the lead time constant and phase margin, although they were expected to decrease and increase, respectively, with improved task performance [26, 41, 45]. The large variations in these parameters were due to difficulties with the identification of the human operator control dynamics caused by the relatively high levels of remnant present in the data. These high remnant levels were, in turn, due to a lack of linearity in human operator control behavior of the task-naive participants, especially during initial training [41, 45] (despite the increase in linearity indicated by the increased VAF throughout training), as well as the fact that participants had to divide their attention between two axes [67].



(a) Pitch group 1 **(b) Roll group 1**

(c) Pitch group 2 **(d) Roll group 2**

(e) Pitch group 3 **(f) Roll group 3**

Fig. 26. Variance Accounted For for time-varying and time-invariant MLE method.

Table 17. Mean increase in VAF due to time-varying human operator model identification

ΔVAF, %		Training phase	Retention phase							
			2M Test	3M Test	4M Test	6M Retention test				
		Runs 95 - 100	Runs 1 - 5	Runs 1 - 5	Runs 1 - 5	Runs 1 - 5	Runs 6 - 10	Runs 11 - 15	Runs 16 - 20	Runs 21 - 25
Pitch θ	Group 1	1.5%				1.3%	1.4%	1.3%	1.3%	1.4%
	Group 2	2.9%		2.0%		2.2%	2.5%	2.7%	2.7%	2.6%
	Group 3	1.8%	1.8%		1.8%	1.8%	1.9%	2.0%	1.8%	1.8%
Roll ϕ	Group 1	1.9%				1.6%	2.2%	3.0%	2.3%	2.5%
	Group 2	2.9%		2.3%		1.8%	1.9%	2.6%	2.0%	2.1%
	Group 3	2.3%	2.4%		2.3%	2.2%	2.2%	2.7%	2.8%	2.6%

Table 18. Statistical analysis results for VAF increase due to time-varying human operator model identification

VAF		Training phase	Retention phase							
			2M Test	3M Test	4M Test	6M Retention test				
		Runs 95 - 100	Runs 1 - 5	Runs 1 - 5	Runs 1 - 5	Runs 1 - 5	Runs 6 - 10	Runs 11 - 15	Runs 16 - 20	Runs 21 - 25
Pitch θ	Group 1	**				**	**	**[a]	**	**[a]
	Group 2	**		**		**[a]	**	**	**	**[a]
	Group 3	**	**		**	**	**[a]	**	**	**[a]
Roll ϕ	Group 1	**				**	**	**	**	**
	Group 2	**		**		**	**	**	**[a]	**[a]
	Group 3	**	**		**	**[a]	**	**	**	**

[a] At least one sample not normally distributed, Wilcoxon signed-rank test applied instead of dependent t test.

In accordance with previous dual-axis tracking task experiments [27–31], 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 perceivability of errors in roll due to a lower pixel resolution in roll compared to pitch. Additionally, higher human operator control gains were observed in pitch compared to roll, which was consistent with the known relationship between operator gain and task performance [26,41]. Also, lower human operator response delays and higher VAF values were observed in the pitch axis compared to the roll axis, as expected from previous dual-axis tracking task experiments [27–31]. 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 a task comparable to the current tracking task [23, 24]. The first retention test of each of the groups showed that all groups exhibited significantly decreased tracking performance compared to end of training in both pitch and roll. However, no significant differences were observed between these first retention tests (the 6-month retention test of group 1, the 3-month retention test of group 2 and the 2-month retention test of group 3). These findings actually suggest the opposite of the hypothesis, i.e., that skill-based manual control behavior actually decays according to a negatively accelerating decay curve, as tracking performance decreased quickly the first few months after training, after which the decrease 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, 22, 78, 79]; overlearning is known to enhance retention [16, 19, 80, 81].

For the 6-month retention tests 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 [23, 82–84]. In this case, the 2-month, 3-month, and 4-month retention tests of groups 2 and 3 can be regarded as additional practice moments scheduled during a 6-month retention interval. Surprisingly, the worst performance during the 6-month retention test was shown by group 2, while groups 1 and 3 exhibited very similar performance. However, statistical analysis 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 retention test, and the hypothesis could neither be accepted, nor rejected.

Based on previous retention studies [12, 79, 85], 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 retention test learning curves of groups 1 and 2 had higher learning rates than the learning curves fitted to their training phase data. The 6-month retention test 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 retention test, 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 worse performance of group 2 in the 6-month retention test 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 [89]. Additionally, the number of experiment groups should be increased to be able to evaluate retention intervals 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 retention test did not give an indication of what the 'optimal' retention interval is while at the same time minimizing the amount of refresher training, the 2-month retention test of group 3 and 3-month retention test of group 2 did give a slight indication. 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. 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 retention intervals.

Human operator model parameters showed that attempting to model the control behavior of fully task-naive human operators in a dual-axis compensatory tracking task is certainly pushing the limits of the current state-of-the art-manual control cybernetics methods. Human operator models with fairly low VAF values were observed throughout both the training and retention phases of the experiment. Also, extreme values in human operator parameters (especially in lead time constant, neuromuscular damping ratio and in the early stages of the training phase also in human operator response delay) were seen throughout the experiment. As current state-of-the-art cybernetics theory and methods only include accepted, universal, models for single-axis compensatory tracking tasks performed by experienced human operators [33, 58], attempting to apply this approach on a dual-axis compensatory tracking task performed by fully task-naive human controllers might be pushing the boundaries too far. Time-varying (nonlinear) human operator model identification methods are required to be able to accurately describe manual control behavior in other types of control tasks than the single-axis compensatory tracking task or to describe the behavior of human operators who are not highly experienced.

Exponential learning curves were heavily used to compare the initial development of control skills with their re-acquisition after decay resulting from a period of non-practice. Although learning curves with relatively high Pearson's correlation coefficients were obtained for the training phase data of several dependent measures, corresponding learning curves for the 6-month retention tests were often omitted due to their low correlation coefficients, as reduced consistency in the parameter values together with the length of the retention tests did not allow for reliable learning curve fitting. Therefore, increasing the number of tracking runs during the retention tests in order to accommodate the fitting of learning curve models could be considered for future retention experiments. Although the necessity for this might not seem immediately apparent from the current experiment results, as skills had not drastically decayed and as lost skills were retained rather quickly when compared to their training development, extending the retention tests might be of use in future experiments if longer retention intervals, such as intervals of several years, are involved, as the decay of skill might be considerable in that case.

As the human operator remnant was expected to be stronger than generally observed in compensatory (single-axis) tracking studies due to the dual-axis task selected for the experiment and due to the fact that participants were fully task-naive, an attempt was made to model the operator remnant power spectrum with a remnant model. The experiment results did indeed make clear that the operator remnant should not be neglected, as it accounted for a significant part of the operator control input. The marked differences in remnant filter parameters between the groups illustrated the challenges in defining a model that is applicable to nearly all human operators. A third-order low-pass filter with damping was selected for the current research, as the remnant spectra of the majority of the participants could be properly approximated by this model, as was consistent with earlier research [56]. However, to be able to develop human operator models for manual control behavior in more realistic control tasks than the tightly controlled and stationary conditions for which the current state-of-the-art cybernetics methods are only truly valid, a better quantitative understanding of the processes that underlie operator remnant is crucial. More research should be conducted into the effects of, among others, the controlled system dynamics, the forcing function characteristics and the relative position of the operator to the visuals on the human operator remnant [87].

The presence of crossfeed and its training and retention effects were investigated by detecting, identifying and modeling a human operator crossfeed response. The modeling results showed that the identified crossfeed dynamics could explain up to 8% of the measured control inputs and improve the accuracy of human operator modeling by 0.5% to 3%. 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 [31] it was found that in a similar task the identified crossfeed dynamics contributed up to 20% of the total human operator control response and improved the modeling accuracy by 1% to 5%. These differences demonstrate that the presence of crossfeed noticeably depends on the manipulator. Whereas the previous study [31] was performed in the SRS at the TU Delft, the current experiment was completed in the HMI Lab at the TU Delft. The position of the sidestick in the HMI Lab 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 HMI Lab resulted in a lower likelihood of introducing cross-coupled control inputs compared to in the previous experiment.

In an effort to study the time-varying characteristics of manual control behavior, an attempt was made to model operator control behavior with a linear time-invariant human operator model with time-varying parameters. Because of the high computational effort of the time-varying implementation of the MLE method, time-varying model identification was only performed on the last five training runs and the retention runs of eleven experiment participants, as the goal was to investigate the influence of a period of non-practice on the time-varying characteristics of manual control behavior. Time-varying model identification showed that the participants' control behavior was not always sufficiently linear for the parameter estimation method to be successful, where the method was considered successful if the modeled control input had a higher VAF than the control input modeled using the time-invariant modeling technique. The time-varying method showed a success rate of almost 90% in the pitch axis and of around 76% in the roll axis. For the successfully identified time-varying models, the VAF had increased around 1% to 3% on average, with occasional VAF increases up to 8%, which is consistent with earlier research on the identification of time-varying human operator control behavior in multi-axis tracking tasks [27]. However, as the time-varying parameter estimation method was not consistently successful in modeling human operator control behavior during the final training and retention runs, in which operator control behavior had already become notably more linear than during initial training, the time-varying implementation of the MLE method is clearly not suitable for modeling human operator control behavior in training studies. Therefore, future research should invest in developing different time-varying (nonlinear) model identification methods which can accurately and reliably identify (nonlinear) human operator control behavior of task-naive operators.

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 [90]. 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 retention research and the determination of optimal (recurrent) training procedures.

V. Conclusions

With this paper, an attempt was made 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-naive participants. Participants were trained in a compensatory dual-axis roll and pitch attitude tracking task and were divided into three groups, differentiating from one another in their retention interval length and the number of retention tests they performed. Performance of the first group was re-evaluated after a period of non-practice of six months, while 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 quantify skill development, decay and retention, participants' control behavior was modeled using quasi-linear human operator models for all individual experiment runs to ensure that changes in control behavior were captured in detail.

Participants showed consistent improvement in task performance during the training phase of the experiment. Also, tracking performance was consistently better in pitch than in roll. The results of the retention tests suggest that manual control skills decay following a negatively accelerating decay curve. The rate of decay was 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, as overlearning is known to enhance retention. Retention performance also suggests that, in order to reduce the likelihood of skill decay, refresher training can better be provided in the form of shorter refresher training sessions on a more regular basis. Lastly, during retention testing lost skills were reacquired at a higher rate than their initial development rate. 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 (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 (almost) the complete skill decay curve.

References

[1] Veillette, P. R., "Differences in Aircrew Manual Skills in Automated and Conventional Flight Decks," *Transportation Research Record*, Vol. 1480, July 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 [retrieved 11 April 2019].

[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%20Decek%20Automation_Final%20Report%5E1-7-16.pdf [retrieved 11 April 2019].

[4] Gillen, M. W., "Diminishing Skills?" *AeroSafety World*, Vol. 5, No. 6, July 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 [retrieved 11 April 2019].

[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, June 2013, pp. 477–485.

[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, Dec. 2014, pp. 1506–1516.

[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 [retrieved 11 April 2019].

[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 [retrieved 12 April 2019].

[10] Mars Architecture Steering Group, "Human Exploration of Mars: Design Reference Architecture 5.0," National Aeronautics and Space Administration Headquarters, Technical Rept. NASA/SP2009566, Washington, DC, July 2009, https://www.nasa.gov/pdf/373665main_NASA-SP-2009-566.pdf [retrieved 20 May 2019].

[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. NAVTRADEVCEN 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, Oct. 1971, pp. 397–405.

[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. 73-32, Alexandria, VA, Dec. 1973, https://apps.dtic.mil/dtic/tr/fulltext/u2/774853.pdf [retrieved 12 April 2019].

[15] Armstrong, M. B., Bleymaier, J. S., Hinkel, L. F., Levins, R., and Sheppard, R. R., "Flying Skill Retention and Proficiency Flying," Research Rept. No. 0095-75, Air Command and Staff College, Air University, Maxwell Air Force Base, 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 [retrieved 12 April 2019].

[17] Prophet, W. W., "Long-Term Retention of Flying Skills: A Review of the Literature," Final Rept. HumRRO-FR-ED(P)-76-35, Human Resources Research Organization, Alexandria, VA, Oct. 1976, https://apps.dtic.mil/dtic/tr/fulltext/u2/a036077.pdf [retrieved 12 April 2019].

[18] Prophet, W. W., "Long-Term Retention of Flying Skills: An Annotated Bibliography," Final Rept. HumRRO-FR-ED(P)-76-36, Human Resources Research Organization, Alexandria, VA, Oct. 1976, https://apps.dtic.mil/dtic/tr/fulltext/u2/a036114.pdf [retrieved 12 April 2019].

[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 [retrieved 15 April 2019].

[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, Oct. 1968, pp. 386–393.

[22] 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, Sept. 1962, pp. 215–226.

[23] Sitterley, T. E., and Berge, W. A., "Degradation of Learned Skills: Effectiveness of Practice Method 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 [retrieved 15 April 2019].

[24] 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, Oct. 1984, pp. 370–374.

[25] Youngling, E. W., Sharpe, E. N., Ricketson, B. S., and McGee, D. W., "Crew Skill Retention for Space Missions up to 200 Days," McDonnell-Douglas Astronautics Company, Technical Rept. F766, Berkeley, MO, Dec. 1968.

[26] 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.

[27] 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.

[28] 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.

[29] 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, May 2016, pp. 1046–1058.

[30] 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.

[31] 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, Jan. 2019, pp. 113–125.

[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 [retrieved 5 April 2019].

[33] 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, Oct. 2018, pp. 468–485.

[34] Levison, W. H., "Two-Dimensional Manual Control Systems," *Second Annual NASA-University Conference on Manual Control*, Cambridge, MA, Feb.-Mar. 1966, pp. 159–180.

[35] 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, June 1967, pp. 113–125.

[36] 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, Sept. 1967, pp. 231–249.

[37] 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, Mar.-Apr. 2009, pp. 366–377.

[38] 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, Jan.-Feb. 2010, pp. 29–41.

[39] 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, Sept.-Oct. 2012, pp. 1456–1471.

[40] Nieuwenhuizen, F. M., Mulder, M., Van Paassen, M. M., and Bülthoff, H. H., "Influences of Simulator Motion System Characteristics on Pilot Control Behavior," *Journal of Guidance, Control, and Dynamics*, Vol. 36, No. 3, May-June 2013, pp. 667–676.

[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, Apr. 2016, pp. 889–901.

[42] 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 [retrieved 5 April 2019].

[43] 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.

[44] Drop, F. M., De Vries, R., Mulder, M., and Bülthoff, 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.

[45] 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.

[46] 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, Sept.-Oct. 2009, pp. 1429–1445.

[47] 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, July-Aug. 2010, pp. 1064–1081.

[48] Robertson, E. M., Pascual-Leone, A., and Press, D. Z., "Awareness Modifies the Skill-Learning Benefits of Sleep," *Current Biology*, Vol. 14, No. 3, Feb. 2014, pp. 208–212.

[49] 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, Mar. 2012, pp. 219–231.

[50] 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, July 1952, pp. 19–26.

[51] 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, Jan. 1954, pp. 32–36.

[52] 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, Apr. 1956, pp. 591–599.

[53] Holding, D. H., *Principles of Training*, 1st ed., Pergamon Press, Oxford, England, 1965, pp. 92-94.

[54] Schmidt, R. A., *Motor Skills*, Harper & Row, New York, NY, 1975.

[55] Singer, R. N., *Motor Learning and Human Performance: Application in Physical Education Skills*, 2nd ed., Macmillan Publishers, London, England, 1975.

[56] Zaal, P. M. T., Pool, D. M., Chu, Q. P., Van Paassen, M. M., Mulder, M., and Mulder, J. A., "Modeling Human Multimodal Perception and Control Using Genetic Maximum Likelihood Estimation," *Journal of Guidance, Control, and Dynamics*, Vol. 32, No. 4, July-Aug. 2009, pp. 1089–1099.

[57] 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.

[58] 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.

[59] Zaal, P. M. T., and Sweet, B. T., "Estimation of Time-Varying Pilot Model Parameters," *Proceedings of AIAA Modeling and Simulation Technologies Conference*, Portland, OR, AIAA Paper 2011-6474, Aug. 2011.

[60] Wiener, N., *Cybernetics: Or Control and Communication in the Animal and the Machine*, 2nd ed., MIT Press, Cambridge, MA, 1961.

[61] McRuer, D. T., and Weir, D. H., "Theory of Manual Vehicular Control," *IEEE Transactions on Man-Machine Systems*, Vol. 10, No. 4, Dec. 1969, pp. 257–291.

[62] McRuer, D. T., and Krendel, E. S., "Mathematical Models of Human Pilot Behavior," AGARDograph AG 188, Advisory Group for Aerospace Research and Development, Paris, France, Jan. 1974, https://www.sto.nato.int/publications/AGARD/AGARD-AG-188/AGARD-AG-188.pdf [retrieved 10 April 2019].

[63] Jagacinski, R. J., and Flach, J. M., *Control Theory for Humans: Quantitative Approaches to Modeling Performance*, Lawrence Erlbaum Associates, Mahwah, NJ, 2003.

[64] Mulder, M., Van Paassen, M. M., and Boer, E. R., "Exploring the Roles of Information in the Manual Control of Vehicular Locomotion: From Kinematics and Dynamics to Cybernetics," *Presence: Teleoperators and Virtual Environments*, Vol. 13, No. 5, Oct. 2004, pp. 535–548.

[65] Pool, D. M., Zaal, P. M. T., Damveld, H. J., Van Paassen, M. M., Van der Vaart, J. C., and Mulder, M., "Modeling Wide-Frequency-Range Pilot Equalization for Control of Aircraft Pitch Dynamics," *Journal of Guidance, Control, and Dynamics*, Vol. 34, No. 5, Sept.-Oct. 2011, pp. 1529–1542.

[66] Nieuwenhuizen, F. M., Zaal, P. M. T., Mulder, M., Van Paassen, M. M., and Mulder, J. A., "Modeling Human Multi-Channel Perception and Control Using Linear Time-Invariant Models," *Journal of Guidance, Control, and Dynamics*, Vol. 32, No. 2, July-Aug. 2008, pp. 366–377.

[67] 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 [retrieved 23 April 2019].

[68] De Vries, R. J., "A Tracking Task for Quantifying Loss of Motor Skills due to Parkinson's Disease," unpublished M.Sc. Thesis, Faculty of Aerospace Engineering, Delft University of Technology, Delft, The Netherlands, May 2016.

[69] 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.

[70] Levison, W. H., Kleinman, D. L., Baron, S., "A Model for Human Controller Remnant: Final Report," George C. Marshall Space Flight Center, National Aeronautics and Space Administration, Technical Rept. 1731, Huntsville, AL, Oct. 1968, https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19700002139.pdf [retrieved 8 May 2019].

[71] 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, Sept. 1965, pp. 42–52.

[72] 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, June 1967, pp. 125–129.

[73] 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 [retrieved 30 April 2019].

[74] 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, Aug.-Sept. 2016, pp. 189–194.

[75] 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.

[76] 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.

[77] 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 [retrieved 11 April 2019].

[78] 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 [retrieved 15 April 2019].

[79] 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, Apr. 1958, pp. 318–328.

[80] Hammerton, M., "Retention of Learning in a Difficult Tracking Task," *Journal of Experimental Psychology*, Vol. 66, No. 1, July 1963, pp. 108–110.

[81] Melnick, M. J., "Effects of Overlearning on the Retention of a Gross Motor Skill," *Research Quarterly*, Vol. 42, No. 1, Mar. 1971, pp. 60–69.

[82] Sitterley, T. E., Zaitzeff, L. P., and Berge, W. A., "Degradation of Learned Skills: Effectiveness of Practice Method 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 [retrieved 15 April 2019].

[83] 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 [retrieved 15 April 2019].

[84] 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 [retrieved 15 April 2019].

[85] Hill, D. S., "Minor Studies in Learning and Relearning," *Journal of Educational Psychology*, Vol. 5, No. 7, Sept. 1914, pp. 375–386.

[86] Zaal, P. M. T., Popovici, A., and Zavala, M. A., "Effects of False Tilt Cues on the Training of Manual Control Skills," *Proceedings of the AIAA Modeling and Simulation Technologies Conference*, Kissimmee, FL, AIAA Paper 2015-0655, Jan. 2015.

[87] 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, Dec. 1969, pp. 101–108.

[88] 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.

[89] Hinton, P. R., *Statistics Explained: A Guide for Social Science Students*, 2nd ed., Taylor & Francis, London, England, 2004.

[90] 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, May-June 1983, pp. 257–266.

Part II

Preliminary Report

NOTE:
This part has already been graded under AE4020

Chapter 2

Preliminary Research Questions

In this chapter, the preliminary research questions are defined. To ensure that an extensive literature study would be performed, it was decided that during the literature review the focus would be on the long-term retention of skill-based control behavior in general, meaning not only on the retention of tracking tasks, but also of other skill-based tasks. Additionally, the literature review not only focuses on skill retention of novices, but also of the more experienced population.

The main research question to be answered during this thesis research was stated in Chapter 1 and is restated below for completeness.

Main research question

"To what extent do manual control skills of novices decay during periods of non-practice?"

To give more guidance to the literature study, the research question is broken down into several sub-questions. Answering these questions might lead to discovering a research gap in the current state of scientific knowledge regarding the retention of manual control skills.

Preliminary research questions

PT sub-question 1: "Which variables influence skill retention?"

PT sub-question 2: "After what time do manual control skills start to deteriorate significantly?"

PT sub-question 3: "What kind of trend does skill decay follow?"

PT sub-question 4: "What is the required amount of retraining after a period of non-practice to regain previously achieved performance levels?"

After the literature review, the sub-questions are adapted to reflect the knowledge gained during the literature study. The updated questions are presented in the final experiment design in Chapter 6.

Chapter 3

Retention

Skill retention can be defined as the ability to perform a certain skill after a period of non-practice [15]. This chapter will provide an overview of previous research concerning the retention of manual control skills. Because the scope of the current research has been limited to skill-based control behavior, this will also be the focus of this literature review. This chapter aims to summarize the most important findings in literature concerning the retention of tracking skills in order to form a solid basis for the remainder of this research. First, the variables known or suspected to affect retention will be introduced in Section 3-1. Then, previous findings on for how long different types of control skills are retained will be discussed in Section 3-2. This will be followed by a review of previous attempts to model the shape of the skill decay curve in Section 3-3. Next, the amount of retraining required to regain after a period of non-practice the performance levels achieved during training are discussed in Section 3-4. Finally, conclusions that can be drawn from this literature review and which are of paramount importance to the remainder of this research are presented in Section 3-5.

3-1 Variables Affecting Retention

Between 1960 and 2000 several studies have been performed to identify the variables which might affect skill retention [15,23,30–35]. All these studies adopt different ways to categorize these variables. For consistency and clarity, in this research the variables known or suspected to influence retention will be divided according to the categories established by McRuer and Jex [27] to classify the variables that influence a pilot's characteristics as a controller. These categories are as follows:

- *task variables* define the nature and characteristics of the manual control task itself and are comprised of all the system inputs and control system elements external to the human operator which affect directly and explicitly the operator's control task;
- *environmental variables* define the environment in which the control task is performed;
- *operator-centered variables* are directly related to the individual pilots themselves; and
- *procedural variables* are mainly concerned with how an experiment is conducted.

An extensive overview of the variables known or suspected to affect skill retention is provided in the subsections below. Additionally, some remarks are made concerning variables of which the effect upon skill retention is unknown.

3-1-1 Task Variables

Four task variables have been found to have an influence on skill retention.

Task variables

- Nature of response required
- Task organization
- Task difficulty
- Compatibility of display-control relationships

Nature of response required
Motor responses are often defined as being either continuous or discrete, although responses are usually a combination of both. A response can be regarded continuous if it concerns a repetitive movement pattern without an apparent beginning or end, whereas a response with a definite beginning and end is considered discrete. A discrete response often takes only a short period of time. Tracking tasks are one of the most often used continuous tasks in studies concerning control skills [30]. An example of a continuous response in controlling an aircraft would be using the yoke or sidestick, whereas moving the landing gear up or down would be an example of a discrete response. Instead of making a distinction in the nature of the response required, sometimes a distinction is made between closed-loop and open-loop tasks. Closed-loop tasks are the ones requiring discrete responses, whereas open-loop tasks involve continuous responses [32]. It has been demonstrated that open-loop tasks (continuous responses) are better retained than closed-loop tasks (discrete responses) [18, 36–39]. The nature of the response is frequently believed to be the cause of this difference, however, the exact cause remains unknown. A possible explanation is that continuous tasks are overlearned and thus retained better (see variable '*Level of original learning*'), because it is unclear what constitutes an individual trial, which may allow for repeated practice. Another hypothesis is that open-loop tasks are simply retained better because their continuous responses are more integrated or coherent than the discrete responses of closed-loop tasks. Because of this hypothesis, some researchers believe that the nature of the response cannot actually be pointed out as a variable influencing skill retention and the retention can be attributed to task organization (see variable '*Task organization*') [31]. Lastly, it is hypothesized that the methods used to measure the retention of discrete responses are more sensitive to slight performance deviations than those used to measure the retention of continuous responses [40].

Task organization
Organization refers to the process by which the elements that define a task can be organized or ordered by establishing consistent relations among the elements [41]. Highly structured tasks are believed to be learned at a faster rate than less structured tasks. Depending on the amount of learning, highly structured tasks might also be retained at a higher level than less structured tasks. This mainly holds for conditions of moderate learning. However, once the advanced learning stage has been reached, less structured tasks can be retained as proficiently as highly structured ones. The reason for this is that even unstructured tasks can be organized by individuals, however, organizing takes time and practice. Seemingly unstructured tasks are thus better structured after a longer practice period, and hence, are retained better after more training [42–45].

Task difficulty
The more difficult a task, the worse it is retained [32,34]. As a result, for example, a tracking task with acceleration control is retained worse than a tracking task with position or velocity control, because human operators have difficulty with accelerated motion [46].

Display-control compatibility
A display-control relationship ensures that if a manipulator is moved in a specific direction, the controlled 'element' moves in a corresponding direction. Certain display-control relationships are believed to be more 'natural' and 'expected' by human operators than others because of previous experiences. This belief is based upon the observation that when multiple display-control relationships are possible for a given perceptual-motor task, one relationship will lead to substantially better initial performance than the others [47]. An example of a natural display-control compatibility in aircraft control would be that if the sidestick is moved forward, the nose of the aircraft is expected to go down. A nose-up movement with a forward movement of the sidestick would be seen as 'unnatural'. The compatibility of display-control relationships not only influences the ease with which a skill is learned and its transfer to a different environment [46–51], but also its performance after a retention interval [52]. This means that training with high-compatibility equipment leads to less required training to achieve and maintain an adequate performance level than training with equipment having incompatible display-control relationships.

Specificity of task displays
A factor found not to influence skill retention is the specificity of task displays. This can be explained in the following manner. Many control tasks, such as tracking tasks, depend heavily upon the processing of visual information from a task display, especially during the initial stages of learning, when the learner is forced to rely upon visual cues to guide their performance. Later on, the learner might rely more heavily on proprioceptive or other internal cues and depend less on external (visual) cues [53, 54]. This means that additional visual cues designed to supplement the information provided by the task display facilitate the early acquisition of a skill. However, once the learner relies more upon internal cues, these additional visual cues become unnecessary, because they do not provide extra information anymore beyond the information provided by internal cues. As a consequence, a higher display specificity does not influence the final level of performance attained, nor the skill retention. However, there must of course be a base level in the specificity of a task display to be able to perform the task accurately.

3-1-2 Environmental Variables

Little research has been performed into the influence of environmental variables on skill retention. Therefore, only one environmental variable has been identified as having an influence on skill retention.

Environmental variables
• Fidelity of training devices

Fidelity of training devices
The fidelity of a training device is the similarity between training environment and the operational setting in which the trained skills will be used. There are several types of fidelity, including, but not limited to, physical fidelity and functional fidelity of training devices. Physical fidelity refers to the physical resemblance between the displays and controls on a training device and those on the operational equipment. Functional similarity, on the other hand, refers to the 'degree of representativeness' or 'psychological realism' of a training device compared to the operational equipment [30]. An investigation into all different types of fidelity is considered outside the scope of this research. Functional fidelity is often considered the most important type of fidelity in training [30, 34]. Although several studies have been performed on the influence of training device fidelity on the retention of procedural skills [55–57], little research has been conducted on its influence on the retention of continuous movement tasks. It has, however, been shown that functional fidelity is a necessary and sufficient condition for learning procedural tasks [30].

3-1-3 Operator-Centered Variables

Four operator-centered variables have been found to have an influence on skill retention.

Operator-centered variables

- Individual ability levels
- Attention
- Presence of facilitating and interfering activities
- Transfer of training

Individual ability levels
Some individuals have a higher initial or 'natural' ability in performance of a task without prior practice than others. Individuals having a higher initial ability generally require less time to reach a specific performance level than individuals with less initial ability [58]. Also, individuals of higher initial ability often reach higher proficiency levels and retain skill at a higher level than individuals of lower initial ability [59]. However, the rate at which skill proficiency is lost does not depend on initial ability levels [58–60]. This is illustrated in Figure 3-1, which presents hypothetical skill decay curves for three groups of varying initial ability. Note that at the end of training, the highest task performance is attained by individuals with the highest initial ability, whereas individuals with the lowest initial ability achieve the lowest task performance. The same observation can be made regarding task performance after the retention interval. As a consequence, individuals with higher initial ability levels require refresher training less often than those with lower initial ability levels. Furthermore, refresher training can be briefer for individuals with higher initial ability because they retrain to required levels faster than those with lower initial ability levels [60]. Also note that the skill decay curves are parallel to one another, meaning that the actual rate of skill decay is independent of initial ability level. This would mean that if individuals of higher initial ability and those of lower initial ability would be trained to the same level of proficiency (with individuals of higher initial ability receiving less initial training than those of lower initial ability), both would demonstrate the same task performance after a certain retention interval [34]. Also note that the shape of the skill decay curves in Figure 3-1 is purely hypothetical. The shape of the skill decay curve is the subject of Section 3-3.

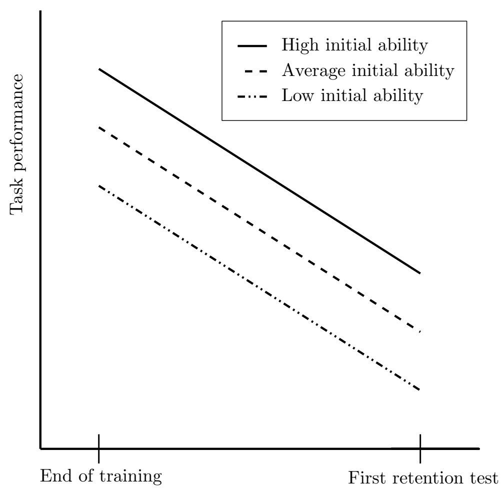


Figure 3-1: Hypothetical skill decay curves for three groups of varying initial ability (adapted from [30]).

Attention

The more attention is given to training, the better training results are. Since better training results lead to better retention, retention is benefited by increased attention during training [61]. Additionally, skills are better retained when learners adopt an external focus of attention (i.e., attention to the results of a movement) instead of an internal focus of attention (i.e., attention to the body movements themselves) [62–65]. The explanation for this can be found in the constrained action hypothesis, according to which conscious attention impairs the bodily mechanisms in which well-developed motor skills are represented [66].

Presence of facilitating and interfering activities

Facilitating activities are activities positively influencing the retention of another task, whereas interfering activities have an inhibitory effect on the retention of another task. Facilitating and interfering activities can be performed either before acquisition of the task of which they influence the retention, or after, during the retention interval. Assume there are two tasks, task A and task B, and task A was practiced before the acquisition of task B. If practicing task A enhances the retention of task B, it is called proactive facilitation, whereas if it degrades its retention, proactive interference is said to have occurred. On the other hand, if the acquisition of the later acquired task B has a positive effect on the retention of the earlier practiced task A, it is known as retroactive facilitation. The converse of retroactive facilitation is called retroactive interference [30]. A quick overview of these facilitating and interfering effects is given in Figure 3-2. The earlier acquired task A is shown in a navy color, whereas the subsequently acquired task B is shown in dark cyan. The horizontal arrows demonstrate the direction of the effect. Upward, green arrows represent a positive effect on retention, whereas downward, red arrows depict an adverse effect. The vertical arrows are positioned next to the task of which the retention is being influenced.

A ⟶ B ↑	proactive facilitation
A ⟶ B ↓	proactive interference
↑ A ⟵ B	retroactive facilitation
↓ A ⟵ B	retroactive interference

Figure 3-2: Facilitating and interfering effects of the acquisition of one task on the retention of another task. (Earlier acquired tasks 'A' are shown in navy and later acquired tasks 'B' are shown in dark cyan, horizontal arrows demonstrate the direction of the effect, upward, green arrows represent a positive effect on retention, whereas downward, red arrows depict an adverse effect. The vertical arrows are positioned next to the task of which the retention is being influenced.)

It is believed that interference effects are restricted to situations in which two tasks use identical stimuli, but require antagonistic responses from the learner [67], and usually do not last longer than a handful of trials [68, 69]. In other situations, facilitation effects can be anticipated.

'Facilitating and interfering activities' are categorized as operator-centered variables here, because the practicing of tasks other than the task of which the retention is being assessed can depend on numerous different factors, which are unique for every individual. Admittedly, the inclusion of this variable in the category of operator-centered variables might come across as curious and inclusion of 'facilitating and interfering activities' in the list of variables influencing retention might promote the use of different categories to divide the variables.

Transfer of training

Transfer of training refers to the influence of past learning on new learning. Learning one task may facilitate learning or performing another task, called positive transfer, but it may also interfere with learning or performing a new task, negative transfer. Transfer of training depends on multiple variables, including the similarity between the stimuli and required responses in both tasks [70–72], the individual's level of learning [48, 51, 70, 73–75] and the task difficulties [70, 76, 77]. Transferring between motor tasks is usually accompanied by positive transfer, however, the strength of the transfer effects is often small because of the task differences and because of the effects of forgetting [48]. Transfer of training has large overlaps with the variable 'presence of facilitating and interfering activities', although in transfer of training the influencing factor is always from before the acquisition of the task whose retention is being assessed, while facilitating and interfering activities can also occur during the retention interval. Since transfer of training influences the rate of learning of a new task and its performance level obtained during training, it subsequently influences the retention of the new task [30]. Transfer of training can also refer to the same task, but used under different circumstances. Because transfer of training is regarded as a particularly important factor in training studies, it is mentioned here separately.

Aging
Aging has not (yet) been identified as a variable influencing skill retention. Although aging is accompanied by slow, progressive declines in physical strength as well as a decline in learning efficiency, aging has not (yet) been demonstrated to have an effect on skill retention probably because previous experience tends to compensate for losses caused by aging [15, 34].

3-1-4 Procedural Variables

Most of the research performed on skill retention concerned the influence of different procedures. As a result, many procedural variables have been found to affect skill retention.

Procedural variables

- Retention interval
- Level of original learning
- Knowledge of results during initial training
- Response-produced feedback during initial training
- Refresher training
- Schedules of practice
- Use of whole-task versus part-task training methods
- Presence of a secondary task
- Performance measures

Retention interval
The retention interval is the period of no practice between the acquisition and subsequent testing of a skill. The longer the duration of the retention interval, the larger is the decline in skill performance [15, 30–34]. However, because the retention interval also interacts with many other factors discussed in this section, the detailed time line of skill decay and the exact shape of the skill decay curve as a function of retention interval length are discussed more elaborately in Sections 3-2 and 3-3, respectively.

Level of original learning
The level of original learning is the degree of proficiency attained during initial training. This variable is often seen as the most important determinent of the retention of control skills [14, 23, 36, 78, 79]. Knowledge of results and response-produced feedback (see below) have been identified as contributing most to original learning. Overlearning is known to enhance retention [18, 31, 46, 80]. Furthermore, overlearning also causes learners to be able to focus on other tasks or different aspects of their main task, because a skilled learner is able to devote less of their total attentional capacity to an ongoing

task than a novice [81]. It is also suggested that overlearning makes learners more resistent to stressful environments [82].

Knowledge of results during initial training

Knowledge of results refers to externally provided error information about the difference between the actual response given by the learner and the intended response. Especially the early acquisition of a skill relies heavily on knowledge of results. Performance generally improves with knowledge of results, whereas it shows no further improvement or even deteriorates when this knowledge is withdrawn [83, 84]. Further along in the learning process, when the learner relies more upon their own knowledge of correctness, knowledge of results can be reduced or withdrawn without causing a decline in performance [84–86]. Effectiveness of knowledge of results increases with its availability and precision. In general, learners who receive more precise knowledge of results require less training and achieve higher performance at the end of training than those who receive less knowledge of results. This higher performance benefits retention. However, if the amount of information provided by knowledge of results becomes too much for the learner to process within the time allowed, a decline in performance will actually result [87]. It has even been shown that if individuals have a good sense of how well they performed, trial-by-trial knowledge of results might be a distraction that results in worse performance on retention tests. Instead, periodic summary knowledge of results, given only on some proportion of training trials, often promotes long-term retention better than trial-by-trial knowledge of results [88].

Response-produced feedback during initial training

Examples of response-produced feedback are proprioceptive, visual and auditory cues. The more response-produced feedback an individual receives, the more accurate and confident they become in their performance [89–91]. Retention is also facilitated by increasing the number of available feedback channels [92].

Refresher training

Time to retrain individuals to original performance levels is usually less than 50% of the original training time [36, 79, 93]. The required amount of retraining increases, however, with an increase in the length of the retention interval [79, 94] and an increase in task difficulty [95]. Also, previously highly trained individuals tend to require more retraining time to regain their old performance levels than less-trained individuals require to regain their own (lower) old performance levels [36, 79]. This can be explained by the fact that previously highly trained individuals have more to relearn, because they had more to forget in an absolute sense. If learners are provided with some form of practice during the retention interval, they perform better at retention testing than those without practice [37, 96–98]. Refresher training may also lend new opportunities for learning, where performance levels at the end of refresher training might be superior to proficiency levels at the end of original training [99].

Schedules of practice

Learning appears to go slower when practice sessions are longer and more heavily massed than when spaced practice is implemented [100]. Various researchers [67, 70, 101] believe, however, that spaced practice is not actually more favorable for learning and retention than massed practice. Instead, learners are more susceptible to the effects of boredom and fatigue during massed practice. This causes learners to show worse performance during massed practice than the performance level that would reflect their actual learning. Learners who are given rest after learning under massed practice can show performance levels equal to those demonstrated by learners using a spaced practice schedule [102–104]. However, an important exception to this observation must be made. The acquisition of highly fatiguing or dangerous tasks may be hindered during massed practice [67]. In such a case, individuals may start showing involuntary lapses of attention and when these accumulate, these lapses may impair the early learning of individuals under massed practice conditions compared to those under spaced practice conditions [70]. However, note that these observations were made for continuous tasks, the subject of this research. For discrete tasks, however, the distribution of practice does not seem to effect task acquisition nor performance [67, 105]. More important than the manner in which practice sessions are scheduled is the amount of practice time offered. Therefore, in limited-duration training programs, retention is positively influenced by massed practice, because massing allows more trials per unit time and thus more opportunity for initial learning than spaced practice [67, 106].

Use of whole-task versus part-task training methods
In whole-task training a learner practices a task entirely from the beginning until the end, whereas in part-task training the task is divided into a number of subparts, which are practiced separately and integrated later to form the whole task. Several variables are believed to influence the effectiveness of whole-task versus part-task training methods. These variables can be divided into the same general categories again:

- Task variables, for example, task difficulty, task organization;
- Operator-centered variables, for example, experience;
- Procedural variables, for example, amount of practice.

Task variables: The effectiveness of part-task training methods as opposed to whole-task training depends on the difficulty of the independent subtasks of a larger task as well as on the degree to which the subtasks are interrelated [67, 70, 101, 107]. Several remarks can be made about this. Firstly, simple to moderately difficult tasks are easier to learn using whole-task training, whereas difficult tasks can better be learned using part-task training methods [101]. Further, tasks that can be divided into meaningful independent subtasks can better be learned using part-task training methods, whereas tasks that require high coordination and timing of the 'subtasks' are learned at a faster rate using whole-task training. Lastly, a relation can be seen between task difficulty and task organization impacting the relative effectiveness of whole-task and part-task training. As task difficulty increases, training for tasks of high organization can best be performed using whole-task training. However, training for tasks of low organization is benefited by part-task training methods when task difficulty increases [108, 109].

Operator-centered variables: Older individuals with more task-related experience often learn better and at a faster rate using whole-task training rather than part-task training methods. This might be explained by the fact that to more experienced individuals a task seems less difficult and less difficult tasks can better be learned using whole-task training [108].

Procedural variables: The later in training, the more learning is benefited by whole-task training. This can also be explained by the fact that task difficulty 'decreases' with training time, and easier tasks can better be learned using whole-task training [108].

The training method yielding the highest level of performance during original training should lead to better retention. However, tracking tasks require continuous responses, thus favoring whole-task training.

Presence of a secondary task
Practicing a motor task can benefit from the presence of a secondary task when the secondary task is difficult [110], or when it engages similar processes as the primary motor task [111]. However, Goh [112] found that the primary task retention mainly benefited from a secondary task which engaged similar processes as the primary task. Secondary task difficulty did not seem to influence primary task retention. Another experiment involving a compensatory tracking task as the primary task also found that retention performance of the tracking task was not influenced by the difficulty of the secondary task [113]. However, these results hold for short retention intervals. More research is warranted on the effects of secondary tasks on long-term skill retention of primary tasks.

Performance measures
Different types of performance measures can lead to different retention results. For example, absolute error scores can lead to different skill decay curves compared to measured time-on-target. However, the effect of different performance measures on skill retention seems to be artificial instead of having a real influence on retention performance [30]. Hence, this variable could also be excluded from the influencing variables list, if desired. However, it definitely is something to be noted.

Next to the procedural variables that have been shown to affect skill retention, there are also several variables of which the effect on skill retention is still unclear, including *extra test trials prior to final testing*, *augmented feedback* and *mental rehearsal*.

Extra test trials prior to final testing
Additional test trials prior to final testing, without knowledge of results, facilitate memory retention. However, there is no evidence concerning the effects of additional test trials, without knowledge of results, upon the retention of motor skills [30].

Augmented feedback
Information about a response may be supplemented by additional visual or auditory cues. For example, if an individual can see each time they hit a target, feedback can be augmented by saying 'hit' or flashing a light. The effects of augmented feedback vary for different types of tasks, both while the additional feedback is present during training as well as in tests after it has been removed [83], although augmented feedback often, but not always, facilitates task performance during training [49]. However, there is no definitive conclusion on the underlying reason for this. This facilitation effect is sometimes attributed to a change in motivation [114], learning [115], or both [116]. However, the effect of augmented feedback on performance after it has been removed differs considerably. It could lead to lasting benefits [115, 116], but also to no performance difference when compared to individuals that performed training without augmented feedback [114, 116] and even to performance deficits [117]. Although considerable research has been performed on the effects of augmented feedback on training and transfer, little research has been done to assess the effects of it on the long-term retention of motor tasks. An experiment conducted to assess the effect of supplementary auditory cues on the retention of a compensatory attitude control task did not identify any effects [118]. However, more research is warranted before drawing any conclusions. If augmented feedback is found to enhance retention, new questions arise, such as: *"What type of augmented feedback leads to the best retention? When to provide augmented feedback? How often should it be provided?"*

Mental rehearsal
Instead of physically training a skill-based control task, it can also be practiced mentally by merely imagining the required responses. The benefits of mental rehearsal on skill retention are still unclear. Whereas some studies have reported that mental rehearsal does not have any beneficial effects on retention [119], others have concluded that it only enhances retention of largely cognitive tasks, but not of mainly motoric tasks [120, 121]. Again others have shown benefits even for motoric tasks [122]. Therefore, additional research is warranted before drawing a definitive conclusion.

3-2 Duration of Retention

To determine for how long different kinds of manual control skills are retained, 16 retention experiments concerning motor skill retention were reviewed. The objective was to assess whether an investigation into the retention of manual control skills would be feasible within the time set for a M.Sc. thesis and if so, what kind of tracking task could best be used. Only experiments relevant to the retention of tracking or manual flying skills for extended periods of time were included. The control tasks reviewed ranged from a simple rotary pursuit task, to a single-axis tracking task with acceleration control, to a bidimensional compensatory tracking task, to complex flight maneuvers, to spacecraft landing tasks [36–38, 44, 46, 78, 79, 98, 99, 113, 123–128]. Summaries of the 16 retention experiments including task description, experiment procedures, amount of training, length of retention interval and retention performance are provided in Appendix A. It was hoped that more experiments could be reviewed, however, many of the earlier skill retention experiments concerned procedural tasks with either short or long retention intervals [55, 57], or tracking tasks with very short retention intervals ranging from less than an hour to a few days [129]. This can probably be attributed to the challenging nature of the research due to the fact that subjects have to remain available for extended periods of time [129, 130].

Furthermore, experiments concerning the retention of manual flying skills are difficult to perform with pilots, because it would require them to refrain from flying for the duration of the entire experiment.

Experiments concerning the retention of rotary pursuit tasks [123–125] have demonstrated that simple tracking tasks are retained over extensive periods of time, far exceeding intervals of one year, two years or even a decade. More complex tracking tasks, such as bidimensional compensatory tracking tasks can also be well-retained over fairly long periods of time of at least almost a year, however, this requires extremely exhaustive training [126]. An experiment concerning a single-axis tracking task with acceleration control showed that it is possible to observe significant skill loss in a retention experiment with a retention interval of only six months while subjects had reached asymptotic performance at the end of training, as long as the tracking task is sufficiently difficult [46]. The most difficult flight maneuvers, such as landing an aircraft, might also be retained for less than six months [98]. However, it is still uncertain for how long manual flying skills are generally retained due to the small number of retention studies directly relating to flight skills [15, 130]. A retention duration up to 30 months has been observed [79, 128], but also durations of even less than four months [37, 98]. Moreover, the fact that significant performance losses are observed in simulation environments does not necessarily mean that the losses are operationally significant [36, 38].

A quick overview of for how long motor tasks are retained as a function of their difficulty is depicted in Figure 3-3. Note that the retention intervals are not provided on scale. Intermediate training can roughly be described as training to just before or just after asymptotic performance has been reached, whereas high training constitutes a considerable amount of overlearning far past the point at which asymptotic performance had been attained. Due to the large variety in control tasks and amount of training used in the small number of experiments conducted, it is extremely challenging up to this day to narrow down the definitions of low, intermediate and high training. Because of the uncertainty in retention of manual flying skills, these skills could lie anywhere on the retention interval-axis of Figure 3-3, as indicated by the blue filling.

Because the retention interval in the current research cannot last longer than six months, it can be derived from Figure 3-3 that the tracking task used must be sufficiently difficult. Furthermore, an intermediate amount of training to just before or just after asymptotic performance has been reached, must be provided.

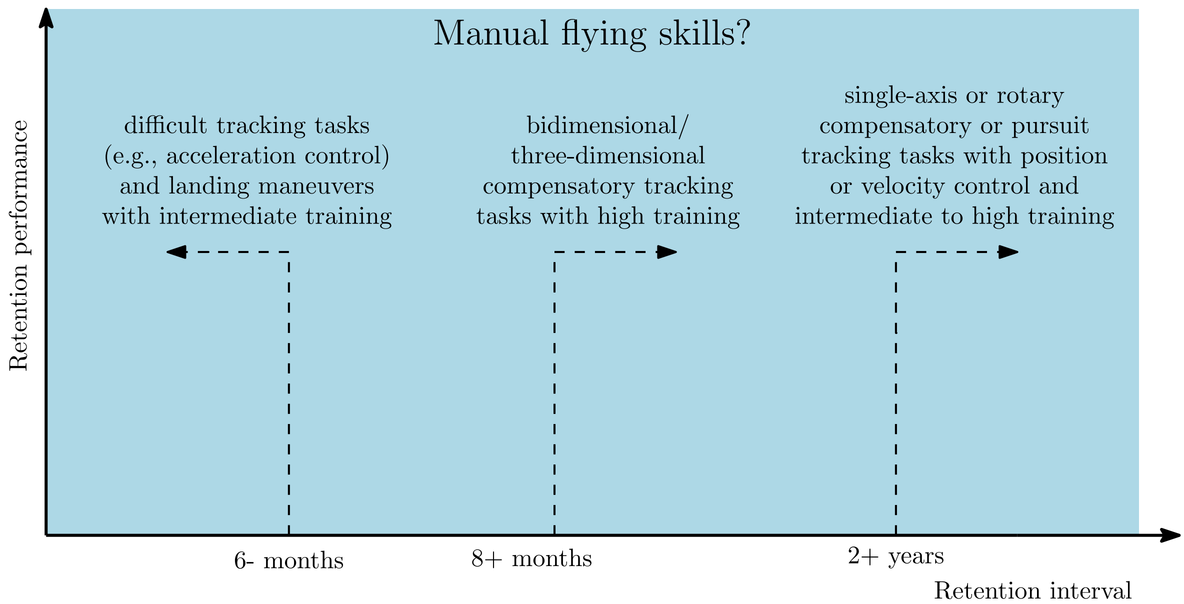


Figure 3-3: Duration of retention for various kinds of tracking and manual flying tasks.

3-3 Skill Decay Curve

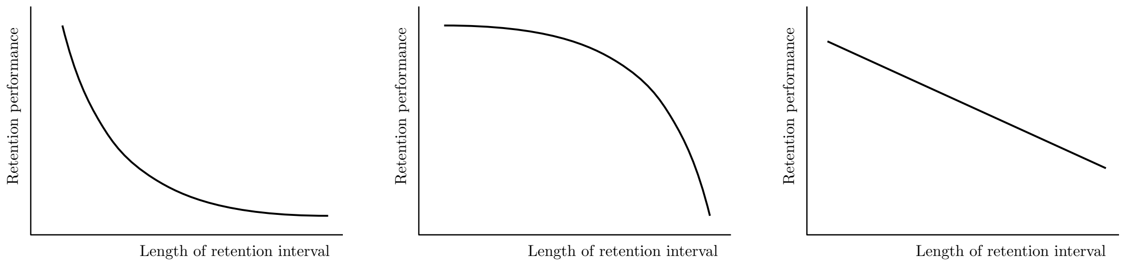
An overview of the general retention duration of various types of tracking and manual flying tasks has been provided in Section 3-2. However, as has been made clear in Section 3-1, skill decay is much more than simply a function of time. Many researchers in the past have tried to model the decay of skill, however, much uncertainty remains.

Ebbinghaus was the first person to quantitatively describe memory decay in 1885 [131]. He noticed that memory retention follows a negatively accelerating decay curve if no attempts are made to retain it, meaning that memory declines fastest immediately after initial learning. This decay curve became known as the classical curve of forgetting. This negatively accelerating decay curve is depicted in Figure 3-4a. The classical curve of forgetting has often been assumed by researchers to also apply to motor skills, including manual flying skills [15, 30].

One research effort on the retention of manual flying skills of a pilot population that observed a negatively accelerating decay curve was Wright's 1973 research [13]. Wright found that 90% of the loss in flying ability of military aviators occurred within the first 12 months of no flying. After 12 months, flying ability remained practically constant. However, it must be noted that these results were not obtained from in-flight proficiency measurements, but were based on ratings of own proficiency provided by 525 aviators.

On the contrary, Sitterley et al. [37] found a completely different skill decay curve for flight control performance of a spacecraft. Manual flying skills were retained fairly well during the first few months of the retention interval, but after that started to deteriorate rapidly. According to these observations, flight control skill decay follows a positively accelerating decay curve. This positively accelerating decay curve is shown in Figure 3-4b. A very similar observation was made in research on the retention of helicopter flying skills. Ruffner et al. [132] found that the decay of helicopter flying skills started to accelerate after a non-utilization period of six months. Fleishman et al. [78] also found an indication in their radar intercept mission tracking task that skill decay could possibly follow a positively accelerating decay curve, since retention performance remained constant during the first 14 months of the retention interval and after that started to deteriorate.

Youngling et al. [127] found again a different result on the retention of an image motion compensation task used in space missions. It was concluded that retention performance decreases linearly as a function of the duration of the retention interval. This linear decay curve can be seen in Figure 3-4c.



(a) Classical curve of forgetting as constructed by Ebbinghaus [131].

(b) Positively accelerating decay curve as found by Sitterley et al. [37] and Ruffner et al. [132].

(c) Linear decay curve as found by Youngling et al. [127].

Figure 3-4: Shapes of skill decay curves found in previous skill retention experiments.

Many other researchers have only felt comfortable stating that retention losses appear to be positively correlated with retention interval length, meaning that retention losses become larger as the duration of the retention interval increases. However, they do not state anything about the shape of this correlation [32, 44, 79].

It remains difficult, however, to draw any conclusions from the different shapes of skill decay curves found in previous experiments due to the fact that different performance measures were used in the different experiments, which could influence the shape of the skill decay curves [31]. For example, in the experiment of Sitterley et al. [37], in which a positively accelerating skill decay curve was identified, performance was measured in terms of absolute errors between target state and current state, whereas Youngling et al. [127], who identified a linear skill decay curve, measured performance in terms of time-on-target instead of absolute errors. Wright [13], who found a negatively accelerating curve, found so on the basis of pilots' self ratings. There has been, however, one study that concluded that retention curves as a function of different methods of assessing retention, are generally comparable, but this study primarily dealt with verbal learning [133].

3-4 Retraining Time

After having identified which variables influence skill retention, for how long control skills are retained and what trend skill decay follows, the last question to be answered is how much retraining time is required to regain old performance levels.

As already discussed in Section 3-1, time to retrain individuals to original performance levels is often less than 50% of the original training time [36, 79, 93]. However, length of retraining time increases with an increase in the duration of the retention interval [79, 94] and an increase in task difficulty [95].

In a rotary pursuit experiment conducted by Bell [123], it was found that after a one-year retention interval retraining up to the performance level achieved at the end of initial training took 40% of the initial training time for this rather simple tracking task with relatively little initial training of only 20 minutes and a required amount of retraining of 8 minutes. A very similar observation was made in a rotary pursuit experiment with a 15-year retention interval [125]. Initial training took a total of 24 minutes and after 9 minutes of retention testing 15 years later, subjects had regained 99.5% of their end-of-training performance, meaning that retraining to old performance levels took about 38% of the initial training time. It must be emphasized here that although one of the retention intervals was fifteen times as large as the other, retraining took almost the same absolute amount of time. Assuming that the subjects in both experiments followed approximately the same performance improvement during initial training because of the similar amount of training, these results would contradict, at least for very simple tracking tasks, the statement above, an often observed result, that the longer the duration of the retention interval, the longer is the required training time. In this case, a moderate amount of initial training led to an equal amount of retraining required after one year and after 15 years.

In a second rotary pursuit experiment with a one-year retention interval conducted by Eysenck [124], it was observed that in case of extensive training of 12.5 hours in total and spread out over approximately 2 months, meaning 37.5 times as much initial training as in the experiment of Bell [123], only 2% (15 minutes) of the initial training time was required for retraining up to performance levels achieved at the end of initial training. However, it must be noted that, although subjects received much more training than in Bell's experiment, retraining to end-of-training performance levels did not take less time in an absolute sense. It even took them more time than Bell's participants. However, this could be due to the fact that Eysenck's participants had possibly reached much better initial training levels than Bell's subjects. This would be in line with the observation made in Section 3-1 that previously highly-trained individuals tend to require more retraining time to regain their old performance levels than do less-trained individuals, because they had more to lose in an absolute sense. However, since the end-of-training performance levels in the two experiments could not be found, this can only be speculated.

During his experiment on the retention of a single-axis tracking task with acceleration control, Hammerton [46] found that it took 83% of his subjects, of whom two-thirds was task-naive, at most 25 retraining trials to reach their old performance levels again after a retention interval of six months, while during initial training reaching these performance levels took them 40 to 110 trials. This means that it took them at most 63% of their initial training time to reach their old performance levels again. However, it would be reasonable to think that the ones with less initial training time also took less retraining trials, whereas the subjects with more initial training trials were also the ones with more retraining trials. This would cause the number of 63% to decrease to be more in the range of possibly 50% or even less. However, since Hammerton was not more specific on the retention results, this can only be speculated.

In their bidimensional compensatory tracking task, Battig et al. [126] found that almost no retraining time was required after a retention interval of eight months. This remarkable observation might be due to the extensive amount of initial training of 19 hours in total spread out over four months. Since a small amount of training was performed each day, degradation of skills between training sessions was almost impossible.

A similar observation was made in a retention experiment using a simulated radar intercept mission tracking task [78]. It was found that almost no retraining time was required after retention intervals up to 24 months. Up to 14 months no retraining time was required at all, whereas after 24 months only a few minutes were required. This experiment included a training phase of almost six hours in total, where training was spread out over six weeks with approximately 3 training sessions per week. Again, due to the distribution of practice, there was little opportunity for skill loss between the extensive number of practice sessions.

A curious result was found in an image motion compensation task for space missions [127]. Although it is unknown how much retraining time individual participants required to regain their old performance levels, on average the required amount of retraining was the same for subjects with one- and three-month retention intervals, while it was remarkably more for subjects with a 6.5-month retention interval. This is remarkable since the group with a three-month retention interval did exhibit more skill loss than the group with a one-month retention interval, as can be read in Appendix A. This experiment again contradicts the often drawn conclusion that the required amount of retraining increases with the length of the retention interval. In this case, this only seems to hold from a certain retention interval length onward.

In a compensatory three-axis control task for aircraft control [79] it was found that after eight hours of original training and a retention interval of one to two years up to 75% of skill loss was regained during the first 5 minutes of retraining. However, very slight performance gains were still realized during the first 48 minutes of retraining, meaning for a retraining time equal to 10% of the initial training time.

During their experiment on the retention of a complex simulated flight task including take-off, climbing and descending turns and landing, Mengelkoch et al. [36] found that after a retention interval of four months it took task-naive subjects on average about three retraining trials to get the primary flight control parameters at the same performance level as at the end of training. Since this observation was the same for the groups with 5, respectively, 10 initial training trials, this means that it took them, 60%, respectively, 30% of their initial training time to reach their old performance levels again. This result confirms the observation that individuals with more initial training do not necessarily require less retraining time to reach their old performance levels again than individuals with less training, because the former group had more to forget in an absolute sense and thus also has to regain more than the latter.

Due to the limited number of skill retention experiments performed and the wide variety of results between those experiments, the required amount of retraining cannot be narrowed down much further than stating that retraining up to end-of-initial-training performance levels occurs rather quickly. The required amount almost never exceeds 50% of the initial training time when low amounts of initial training are received, and does usually not exceed 10% when moderate to high amounts of initial training are performed. In contrast to the required amount of retraining relative to the amount of initial training, it can be concluded that in absolute measures, the required amount of retraining might still be larger when larger amounts of initial training were performed, because the higher performance levels achieved at the end of initial training can lead to more skill loss in an absolute sense, which must all be regained again. However, the exact definitions of low, moderate and high amounts of training depend on the task, since the amount of training per grade increases with an increase in task difficulty. Furthermore, the required amount of retraining often increases with an increase in the length of the retention interval. However, this does not always hold for simple tracking tasks.

3-5 Conclusions

Based on literature, the following conclusions regarding the retention of manual control skills can be drawn:

- The retention of manual control skills is influenced by numerous variables, of which the most important ones are the level of original learning, the length of the retention interval and the task difficulty. Overlearning is known to enhance retention (**PT sub-question 1**).
- Simple tracking tasks, e.g. single-axis or rotary compensatory tasks or pursuit tasks, with position or velocity control can be retained for over two years when an intermediate to high amount of initial training is performed. Difficult tracking tasks, e.g. tasks with acceleration control, and landing maneuvers are retained for less than six months when an intermediate amount of training is performed (**PT sub-question 2**).
- Significant skill loss can be observed in a tracking task experiment with a retention interval of only six months while the subjects had reached asymptotic performance at the end of initial training, as long as the tracking task is sufficiently difficult (**PT sub-question 2**).
- It is difficult to draw any conclusions from the different trends of skill decay observed in earlier retention experiments due to the extremely limited amount of research performed. On top of that, in the little amount of research performed different performance measures were used to measure retention, which could influence the shape of the skill decay curves (**PT sub-question 3**).
- Retraining after the retention interval up to performance levels achieved at the end of initial training occurs rather quickly. However, due to the limited amount of literature on this issue, this observation cannot be narrowed down much further than stating that the required amount of retraining is usually not more than 50% of the initial training time when low amounts of initial training were received and not more than 10% of the initial training time when moderate to high amounts of initial training were performed (**PT sub-question 4**).
- The required amount of retraining to achieve end-of-training performance levels usually increases with an increase in length of retention interval and/or in task difficulty. However, this does not always seem to hold for very simple tracking tasks (**PT sub-question 4**).

Chapter 4

Cybernetic Approach

To objectively and explicitly quantify skill development, decay and retention a cybernetic approach will be used. This cybernetic approach has been used before in training studies [25, 134, 135], but not yet in retention studies. The advantage of the cybernetic approach over more conventional performance measures is that the cybernetic approach utilizes multi-channel pilot models that separate pilots' responses to multiple stimuli. This allows for a quantitative analysis of pilots' use of multiple stimuli/cues for manual control, as well as the development of these control skills during training, after transfer and, for example, after a period of non-practice [25]. The cybernetic approach itself is described in Section 4-1. Section 4-2 describes the training effects on the human operator model defined using the cybernetic approach, after which Section 4-3 discusses the limitations of the cybernetic approach. Lastly, Section 4-4 summarizes the key takeaways about human operator modeling.

4-1 Pilot Model Identification

The cybernetic approach is a system-theoretical, model-based approach to mathematically describe how humans perform many different skill-based manual control tasks [25, 27, 136–141]. To facilitate this approach, research often focuses on skill-based behavior in manual tracking tasks, since it has been demonstrated that for such continuous and stationary control tasks, the adopted control behavior of the human operator can be accurately modeled and determined objectively using system identification and parameter estimation methods [27, 138, 142–144]. The cybernetic approach uses multi-channel models to separately model pilots' use of each perceived stimulus, due to which the approach can give insight into how pilots use different types of cues when developing manual control skills and how this use changes during training and after transfer, or even after a period of non-practice [25]. Using the cybernetic approach, pilots' control behavior adopted during tracking tasks is modeled in terms of distinct contributions that are physically interpretable.

In controlled and slightly simplified manual control tasks, the non-linear and time-varying skill-based control behavior of human operators can be modeled using linear transfer functions and a remnant signal that accounts for non-linear behavior and measurement noise [27]. The inputs to the linear transfer functions describing the pilot's control behavior are the stimuli perceived by the pilot (e.g., visual, somatosensory, vestibular cues), while the sum of the outputs is the human operator's control action [25,27,136–141]. The top part of Figure 4-1 depicts the schematic representation of a single-axis compensatory tracking task with only visual cues. Since the current research will be carried out in a fixed-base simulator setup, a pilot's motion response has not been modeled in Figure 4-1. Such a block diagram as in Figure 4-1 is often applied to model manual control behavior. In a compensatory tracking task, the pilot is required to follow a target signal f_t by giving control inputs u to the controlled aircraft dynamics H_c using some form of controls. The change in aircraft attitude θ resulting from these control inputs can be observed on a visual display in the form of the tracking error e between the target attitude and the current aircraft attitude.

The human pilot is modeled using a linear error response function H_{p_e} and a remnant n as depicted in the lower part of Figure 4-1. In case of visual feedback, the first part of the linear response function typically are the dynamics of the visual perception sensors, meaning the eyes that perceive the tracking error on the visual display. These dynamics are often modeled by a unity gain. The other parts of the linear response function are the equalization dynamics and limitation dynamics, where the latter account for some of the physical limitations of human manual control behavior that have been found to affect pilot control behavior. The limitation dynamics are split up into two parts. The first part is a time delay, modeled as an exponential function containing the visual response delay τ_e, to account for the time delays incurred in the perception and processing of the visual information. The second part are the neuromuscular actuation dynamics H_{nm}, which are modeled as a second-order mass-spring damper system with a neuromuscular frequency ω_{nm} and a neuromuscular damping ratio ζ_{nm} [145]. The equalization dynamics, on the other hand, symbolize the pilot's interpretation and usage of the visually perceived information in the formulation of an appropriate control input, meaning that the human operator's control strategy is captured in the equalization dynamics. These dynamics depend on the controlled aircraft dynamics and other task variables [25]. Whereas in Figure 4-1 the equalization dynamics consist of the pilot visual response gain K_p, the lead time constant T_L, and the lag time constant T_I, these dynamics are adjusted by the pilot such that the combined pilot-aircraft system, $H_{p_e}(j\omega)H_c(j\omega)$, approximates single-integrator dynamics for a wide frequency range, including the cross-over frequency [145]. Depending on what type of equalization is required to satisfy this condition for a given controlled element dynamics, the lead-lag equalization term could reduce to, for example, a pure lead, a pure lag or a pure gain [144].

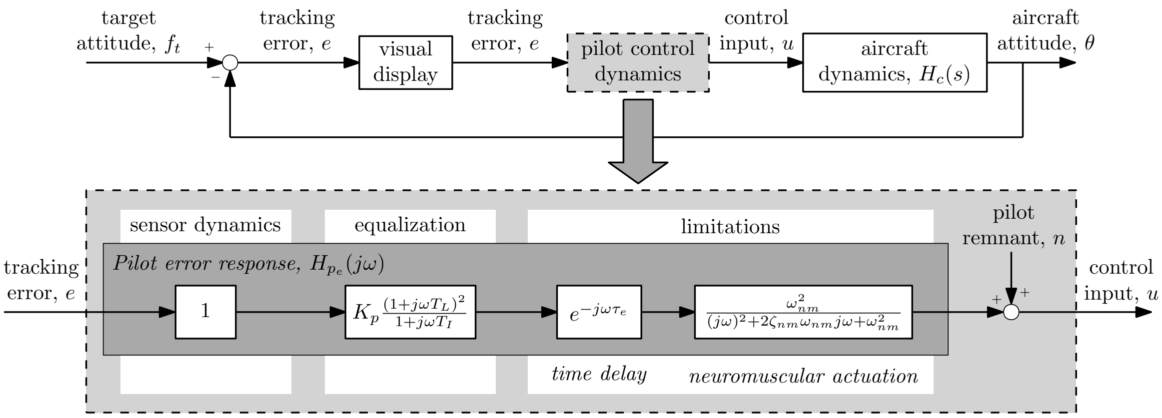


Figure 4-1: Schematic representation of the cybernetic approach (adjusted from [25]).

The pilot model parameters can be estimated using different techniques, for example, using the Fourier Coefficient (FC) Method [146] or using linear time-invariant models, such as Auto-Regressive models with an eXogeneous input (ARX) [147]. These are two frequency-domain identification techniques, of which the latter one yields superior performance. Another method would be Maximum Likelihood Estimation (MLE), a time-domain identification technique. Frequency-domain identification techniques consist of two steps, since an additional step, the first step, is required to obtain nonparametric pilot describing functions in the frequency domain. Then, in the second step, a parametric model is fit to the obtained frequency-domain describing functions to obtain the estimates of the pilot model parameters. A disadvantage of these two-step frequency-domain identification techniques is that estimation errors originating from the determination of the pilot describing function estimates in the first step influence the reliability of the model parameter estimates in the second step. This leads to biases originating from both identification steps, compared to a single identification step in time-domain identification [148], where pilot model parameters can be estimated directly from time-domain measurements. Additionally, for frequency-domain identification, more stringent requirements on the adopted forcing functions exist [147]. On the other hand, a disadvantage of one-step time-domain identification techniques is their sensitivity to the nonlinearity and many local minima of the cost function. However, to cope with these nonlinearities and local minima, a genetic algorithm can be added to the MLE method [148]. This extended method has been shown to yield more accurate and reliable results than obtained with two-step frequency-domain identification techniques. Therefore, Genetic MLE is used in the current research to estimate the parameters of the pilot models applied to the measured data, at least for an initial data analysis. Details on the Genetic MLE technique can be found in [148].

The human operator model accuracy in describing the pilot's control behavior can be assessed using the Variance Accounted For (VAF), which is calculated using Equation (4-1). The VAF is a time-domain validation metric which compares a modeled signal with a measured signal. The higher the VAF, the better the model is able to capture the human operator dynamics. A VAF of 100% means that 100% of the measured signal is explained by the model.

$$\text{VAF} = \left(1 - \frac{\sum_{i=1}^{N} |u(i) - \hat{u}(i)|^2}{\sum_{i=1}^{N} u^2(i)}\right) \cdot 100\% \tag{4-1}$$

where u is the measured signal, $\hat{u}$ the modeled signal and N the number of data points.

To quantitatively describe how pilots' control behavior changes during training, after transfer and even after a period of non-practice, exponential learning curves can be fitted to any pilot model parameter, as well as to task performance and control activity measures and the VAF. The exponential learning curve model is given by Equation (4-2). This model has successfully been applied in earlier training studies [134, 135, 149, 150]. The fit of the learning curves can be assessed using Pearson's correlation coefficient ρ.

$$y_{lc}(n) = p_a + (1 - F)^n (p_0 - p_a) \tag{4-2}$$

where y_{lc} is the vertical coordinate of the learning curve, p_0 the initial value, p_a the asymptotic value, F the learning rate and n the tracking run number. The parameters p_0, p_a and F are determined using a non-linear optimization method to minimize the summed squared error between the experimental data and the learning curve.

When a human operator makes a notable control mishap, pilot behavior is not sufficiently stationary and linear for the human operator modeling methods to be reliable. This is often expressed in terms of a very low VAF, or inconsistent and unrealistic pilot model parameter values. Therefore, unreliable pilot model results should be excluded when analyzing training effects. Unreliable human operator model fits lead to (relatively) extreme model parameter values, which strongly bias the analysis of training effects with learning curve models.

4-2 Training Effects on Pilot Model Identification

In almost all training studies, clear effects of training can be observed [134, 135, 149, 151]. In tracking tasks, training often becomes evident from improved performance in terms of lower tracking errors [25, 134, 135, 149, 151]. Additionally, for predominantly disturbance-rejection tracking tasks, control activity usually increases throughout training when motion feedback is present [25, 134]. Also, task proficiency is often considerably better when training is performed in a motion-base setting instead of a fixed-base setting [152, 153]. However, the effects of motion feedback are considered to be outside the scope of this research.

Next to training effects on the tracking performance and control activity, effects can also be seen on the human operator model parameters. The pilot gain K_p often increases throughout training, whereas the lead and lag time constants, T_L and T_I, respectively, as well as the response delay τ_e usually decrease throughout training [25, 134]. In manual control skill acquisition studies with task-naive participants, the neuromuscular damping ratio ζ_{nm} typically decreases [134, 135]. This is a sign of task proficiency, since phase lag is slightly lower with decreased damping ratios in the frequencies at which the pilot is actively controlling (the crossover region). Better performance in skill-based manual control is often accompanied by a higher crossover frequency and phase margin [27].

An increase in human operator consistency and linearity is often expected to occur throughout training [27]. To confirm this, the coherence of the pilot's control inputs with respect to the applied forcing functions [134] can be analyzed. The validity of the obtained human operator modeling results is often assessed explicitly, not only by comparing the human operator models to independently estimated describing functions, but also by evaluating the fit of the pilot models in terms of the VAF as described in Section 4-1. An increase in human operator linearity is expressed through an increase in coherence and/or an increase in the VAF. When the human operator linearity is high, a good match between the estimated describing functions and the pilot models should be seen. If the estimated describing functions would exhibit considerable scatter and large offsets from the pilot models, the pilot's manual control behavior is not sufficiently consistent or linear. This phenomenon often occurs during the early runs of training due to more pilot remnant. In that case, also the coherence will be lower, and when using a method such as MLE more local minima will occur. A reduction in pilot consistency and linearity causes the reliability of time-invariant human operator modeling methods to reduce. Issues in manual control cybernetics are more elaborately discussed in Section 4-3.

Pilot model parameters that are often estimated with a comparatively low accuracy are the neuromuscular system parameters [134, 154]. A possible explanation for this is that added noise and reduced coherence mainly affect the higher frequencies in the measurement range commonly used in tracking studies. The neuromuscular dynamics influence the human operator response at those higher frequencies. Ultimately, around 2% of the pilot models are usually excluded from analysis in single-axis tracking studies with task-naive participants due to the low reliability of results [134, 135].

4-3 Issues in Manual Control Cybernetics

Most of current-day cybernetics theory was developed in the 1960s, based on technology and analysis methods from that time [27,141,155]. This theory has shown to have severe limitations in its capability to capture the full breadth of human cognition and control [139,141,155]. Below, the discussion on the issues in manual control cybernetics will, similar to the rest of this chapter, be limited to compensatory control with merely visual cues. Issues concerning the two higher levels of the Successive Organization of Perception (SOP) hierarchy [143], namely pursuit and precognitive control, are extensively discussed in [141, 155]. No universally accepted pilot models exist yet for these higher levels of control in the SOP. The issues with the cybernetic approach relevant to the current research can be split up into two categories: neuromuscular adaptations and the learning, adaptive controller.

First addressing neuromuscular adaptations: during learning, the human operator dynamics not only change due to cognitive adaptations, but also due to physical adaptations in the neuromuscular system [156]. These physical adaptations, such as increased stiffness from cocontraction or reflective activity, often occur subconsciously and much faster than adaptations due to higher-level learning [141,155]. Current human operator models provide a "lumped" insight into all effects of human operator learning due to the fact that identification techniques are based on the operator's 'inputs' and 'outputs'. However, in order to get a better understanding of learning, the effects of the 'lower'-level physical adaptations occurring in the neuromuscular system must be separated from the effects of the 'higher'-level cognitive adaptations. Measurement techniques must be improved in order to obtain more accurate and less intrusive approximations of the possibly time-varying neuromuscular dynamics and then isolate them from the overall, “lumped” response. Better estimates can, for example, be obtained by taking nonintrusive grip force measurements using pressure gloves, since grip force is often related to neuromuscular system admittance settings [141, 155, 157]. However, this will not be brought up in the remainder of this Preliminary Thesis, since this is considered outside the scope of this research.

The second limitation in manual control cybernetics concerns the time-varying control of learning human operators. The (quasi-)linear, time-invariant (LTI) feedback systems which are currently used to describe human control behavior can accurately model a human operator’s behavior in highly-constrained single-loop compensatory tracking tasks, without any preview on future control constraints, only allowing the operator to react on the current situation, from the moment the operator is done learning [144]. The assumption is made that the human operator has no preview on future control constraints, which in reality is the exception in manual control, and not the rule: the majority of relevant control tasks have some sort of preview of the future task constraints, as more elaborately discussed in [141, 155]. However, while the crossover model together with the verbal adjustment rules can be used to model compensatory tracking relatively accurately [27, 137, 138, 144], some restrictions apply to the compensatory tracking task in order for the crossover model to be valid. The target (or disturbance) forcing functions cannot contain a ‘recognizable’ pattern. If the operator would be able to detect a repeating pattern, he/she would introduce a feedforward control loop to improve performance and stability and reduce control effort, in which case the crossover model would not be valid anymore [141, 155]. Moreover, the restriction on the LTI model of only being able to model a fully-trained human operator can to a large extent be attributed to the time-invariance assumption of the model. It is this assumption that prevents the accurate modeling of human learning and adaptation, preventing a true understanding of human behavior. When the operator has not yet finished learning, the remainder of the human operator response, called ‘remnant’, which is the part that cannot be explained by the LTI model and which is usually neglected, is often relatively large, meaning that a significant part of the human operator response is ignored. State-of-the-art cybernetics theory must be extended to also include learning effects in the human operator modeling theory [141, 155].

Besides the lack of understanding of realistic human control behavior, rather crude experimental techniques are used to identify manual control behavior. Only the overall, “lumped” response of a fully-trained human based on prolonged measurements can be identified [141, 144, 147, 152, 155, 158]. This “lumped” response fuses all cognitive and physiological adaptations and averages-out all adaptation effects, which prevents an accurate understanding of human adaptation and learning. To accurately model the learning human operator, the LTI models must be replaced by time-varying identification techniques, perhaps even by methods that can be employed in real-time [141, 155]. In order to do so, Mulder et al. [155] have suggested to start with developing *“recursive, 5-to-20 seconds sliding-window (Extended) Kalman Filter techniques that estimate the linear time-invariant (LTI) manual control parameters.”* In this manner, the extent to which these parameters vary in time, both within-subjects as well as between-subjects, can be investigated. This could reveal the consequences of the “averaging effect” of current techniques, which use the entire measurement run to approximate the human operator’s control behavior. However, in order to truly move forward, the concept of LTI systems must be abandoned altogether, and time-varying manual control models must be identified, for example, by applying closed-loop identification methods for linear parameter-varying systems [159, 160]. These time-varying identification techniques can be used to model the human operator during the full learning curve, from novice to expert. Identifying time-varying manual control models also provides the opportunity to determine to what extent the universal “time-invariance” assumption of cybernetics is valid [141, 155].

Lastly, multiloop control tasks often result in overdetermined human operator models, which require extended identification and modeling methods to separate the different operator responses [146, 147, 152, 161]. No universally accepted models exist yet for these multiloop control tasks.

4-4 Key Takeaways on Pilot Model Identification

The most important takeaways of the literature study on human operator modeling are the following:

- To objectively and explicitly quantify skill development, decay and retention a cybernetic approach can be used, which is a system-theoretical, model-based approach to mathematically describe how humans perform skill-based manual control tasks. The advantage of this cybernetic approach over more conventional performance measures is that it uses multi-channel pilot models to allow for a quantitative analysis of pilots' use of multiple stimuli/cues for manual control, as well as the development of these control skills during training, after transfer, and even after a period of non-practice.
- In controlled and slightly simplified skill-based control tasks, human operators are modeled using a linear, time-invariant transfer function and a remnant signal that accounts for non-linear behavior and measurement noise. This remnant signal is usually neglected. The linear transfer function typically consists of sensor dynamics, which in case of visual feedback are modeled by a unity gain, equalization dynamics and limitation dynamics. These limitation dynamics are split up into a time delay to, in the case of visual feedback, account for the time delays incurred in the perception and processing of the visual information, and neuromuscular actuation dynamics. The pilot's control strategy, captured in the equalization dynamics, heavily depends on the controlled aircraft dynamics. The equalization dynamics are adjusted by the pilot such that the combined pilot-aircraft system, $H_{p_e}(j\omega)H_c(j\omega)$, approximates single-integrator dynamics for a wide frequency range, including the cross-over frequency.
- In the current research, Genetic Maximum Likelihood Estimation is used to estimate the parameters of the pilot models applied to the experiment data, at least for an initial data analysis. This one-step time-domain identification method yields more accurate and reliable results than two-step frequency-domain identification techniques.
- In compensatory tracking tasks with task-naive participants and only visual cues, training becomes evident through the following changes:
  - Improved performance in terms of lower tracking errors;
  - Increased pilot gain K_p;
  - Decreased lead and lag time constants, T_L and T_I, respectively;
  - Decreased response delay τ_e;
  - Decreased neuromuscular damping ratio ζ_{nm};
  - Increased crossover frequency ω_c and phase margin ϕ_m;
  - Increased human operator consistency and linearity expressed through an increase in coherence and/or VAF.
- Current-day cybernetics theory can only accurately model a human operator's control behavior in highly-constrained single-loop compensatory tracking tasks, without any preview on future control constraints, from the moment the operator is done learning.
- To be able to accurately model the learning human operator, the LTI models must be replaced by time-varying identification techniques, perhaps even by methods that can be employed in real-time. These time-varying identification techniques can be used to model the human operator during the full learning curve, from novice to expert. Identifying time-varying manual control models also provides the opportunity to determine to what extent the universal "time-invariance" assumption of cybernetics is valid.

Chapter 5

Preliminary Experiment

Based on the findings of the literature review, a preliminary experiment was designed and performed to be able to refine the design of the final experiment. This chapter describes how this preliminary experiment was designed and executed, and discusses its results. The objective of this preliminary experiment is given in Section 5-1. The design of this experiment is described in Section 5-2, after which its results are presented in Section 5-3. A discussion of these results is provided in Section 5-4. This chapter is concluded with recommendations for the final experiment in Section 5-5.

5-1 Preliminary Experiment Objective

During the literature review, it was found that significant skill losses can be observed in a tracking task experiment with a retention interval of only six months while the subjects had reached asymptotic performance at the end of initial training, as long as the tracking task is sufficiently difficult. Based on this finding, it was decided to use a dual-axis compensatory tracking task with challenging controlled aircraft dynamics for the current research. It is believed that such a task is challenging enough to be able to observe skill decay within a retention interval of six months, as long as participants do not overlearn extensively.

A preliminary experiment with a dual-axis compensatory tracking task was set up. The objective of this preliminary experiment (PE) was twofold. If both objectives were met, a research proposal to investigate skill retention could be made.

Preliminary experiment objectives

PE objective 1: Determine how many experiment runs should generally be performed by task-naive participants for them to just reach asymptotic task performance.

PE objective 2: Determine whether the dual-axis aircraft roll and pitch tracking task provides data suitable for pilot model identification.

Determining how long it takes for participants to reach asymptotic task performance is important for two reasons. First, if participants would already reach asymptotic performance within a few experiment runs, it would mean that the control task is too straightforward and it would be highly likely that (almost) no skill decay would be seen after a retention interval appropriate for a M.Sc. thesis. Secondly, if it would take participants too many experiment runs to reach asymptotic performance, it would mean that training a considerable number of participants would take more time than available to be able to finish the research within the allocated time. In that case, a slightly easier control task might be more suitable for the final experiment. Furthermore, the experiment data needs to be suitable for pilot model identification to be able to objectively and quantitatively evaluate skill acquisition, decay and retention.

5-2 Preliminary Experiment Design

The preliminary experiment design consists of the control task, controlled aircraft dynamics, forcing functions, control and independent variables, apparatus, participants, experimental procedures, dependent measures and hypotheses. Two preliminary experiments were performed, each by one individual. The two experiments differed from one another in the forcing functions applied, as explained in Section 5-2-4. The preliminary experiments were similar in setup to a training experiment, in which participants are extensively trained on a specific skill-based tracking task over the course of multiple, in this case four, days, where on each day they perform a specific number of training runs. Although it was the intention that the test subjects in the two preliminary experiments would perform the same number of tracking runs, due to some unfortunate events there was a small difference between the two experiments in the total number of runs performed by the test subjects as well as in the division of the tracking runs over the four training days. This will be more elaborately discussed in Section 5-2-9. Besides the forcing functions and the number of tracking runs, all other aspects of the experiment design were the same for both preliminary experiments. Therefore, when in this chapter the preliminary experiment is mentioned in general, without an explicit mentioning of the first or second preliminary experiment, it means that the information holds for both preliminary experiments.

5-2-1 Dual-Axis Control Task

The task to be controlled in the preliminary experiment was a compensatory dual-axis aircraft roll and pitch tracking task, which has successfully been applied in previous research [150, 162–165]. A schematic representation of this task is shown in Figure 5-1. The participant's objective was to simultaneously minimize the roll and pitch errors, e_ϕ and e_θ, at all times. The roll and pitch errors are the differences between the roll and pitch target forcing functions, $f_{t\phi}$ and $f_{t\theta}$, and the roll and pitch attitudes, ϕ and θ, controlled by the participant using a sidestick with roll and pitch gains $K_{s\phi}$ and $K_{s\theta}$. These errors were presented on a two-axis compensatory display, similar to an attitude indicator, as, respectively, the angle and vertical distance between a reference line (artificial horizon) and a static aircraft symbol. This display is depicted in Figure 5-2. Note that the arrows indicating the magnitude of the roll and pitch errors as well as the error symbols themselves were not depicted on the display during the experiment.

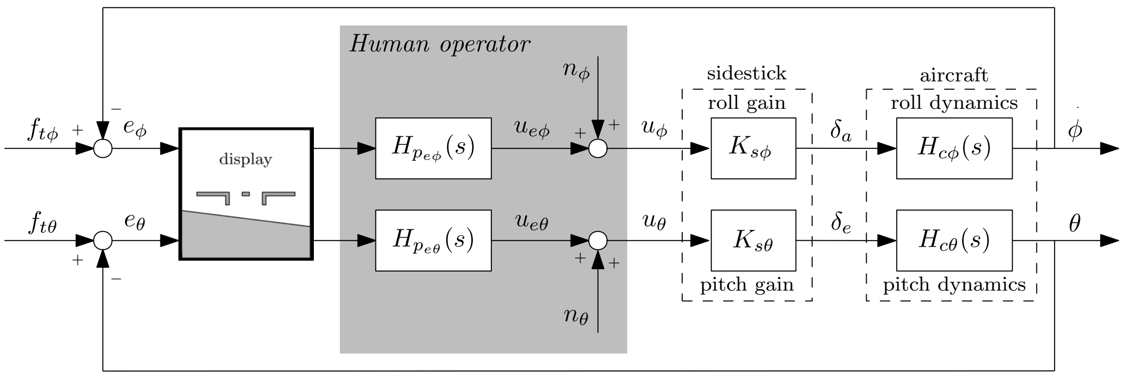


Figure 5-1: Schematic representation of the compensatory dual-axis aircraft roll and pitch tracking task.

The human operator is modeled using two quasi-linear models, one for his/her roll response and one for his/her pitch response. The roll and pitch control inputs, u_ϕ and u_θ, respectively, both consist of a linear error response, $u_{e\phi}$ and $u_{e\theta}$, and a remnant, n_ϕ and n_θ, accounting for non-linear behavior and measurement noise. The linear human operator response functions in roll and pitch are represented by $H_{p_{e\phi}}$ and $H_{p_{e\theta}}$, respectively. To make the task feel more realistic, different roll and pitch dynamics, $H_{c\phi}$ and $H_{c\theta}$, were used.

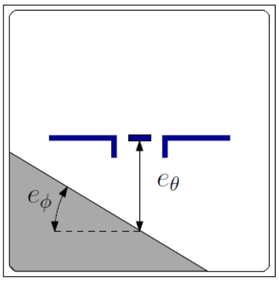


Figure 5-2: Two-axis compensatory display.

5-2-2 Human Operator Modeling

Due to the controlled aircraft dynamics described in Section 5-2-3, the equalization dynamics in the pilot error response H_{p_e} of Figure 4-1 can be somewhat simplified. The pilot model considered for the preliminary experiment is given by Equation (5-1).

$$H_{p_e}(s) = K_p(T_L s + 1)e^{-\tau_e s} H_{nm}(s) \tag{5-1}$$

where H_{nm} are the neuromuscular dynamics modeled by Equation (5-2) [145], as also shown in Figure 4-1.

$$H_{nm}(s) = \frac{\omega_{nm}^2}{s^2 + 2\zeta_{nm}\omega_{nm}s + \omega_{nm}^2} \tag{5-2}$$

5-2-3 Controlled Aircraft Dynamics

The linearized aircraft roll and pitch dynamics are defined by Equations (5-3) and (5-4), 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 is set to 185,800 lbs. The aircraft dynamics are 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 [150].

$$H_{c\phi}(s) = \frac{\phi}{\delta_a} = \frac{0.76773(s^2 + 0.2195s + 0.5931)}{(s + 0.7363)(s - 0.01984)(s^2 + 0.1455s + 0.6602)} \tag{5-3}$$

$$H_{c\theta}(s) = \frac{\theta}{\delta_e} = \frac{0.33282(s^2 + 0.09244s + 0.002886)}{(s^2 - 0.01388s + 0.004072)(s^2 + 0.446s + 0.4751)} \tag{5-4}$$

As can be seen in Equation (5-3), the linearized roll dynamics have a mildly unstable pole (spiral) at this flight condition. In Figure 5-3, it is shown that 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.

The linearized pitch dynamics of Equation (5-4) have an unstable phugoid. The pitch dynamics approximate a double integrator ($\frac{1}{s^2}$) at frequencies higher than 0.6 rad/s, as can be seen in Figure 5-4.

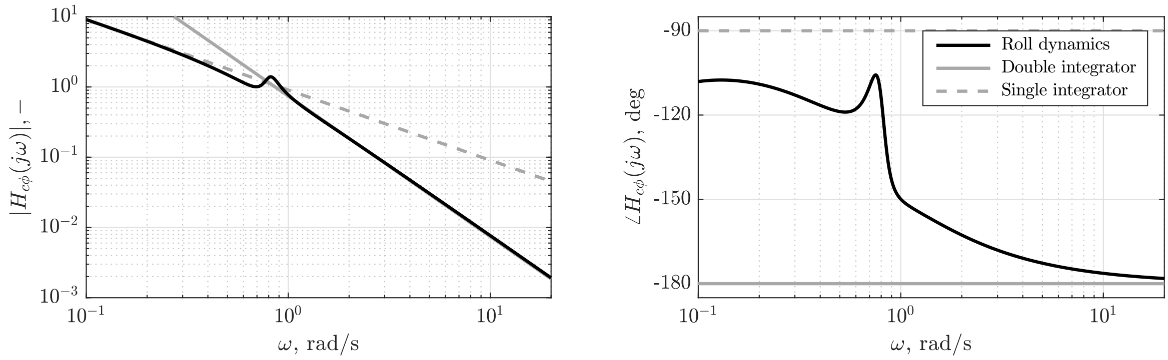


Figure 5-3: Frequency response of linearized aircraft roll dynamics.

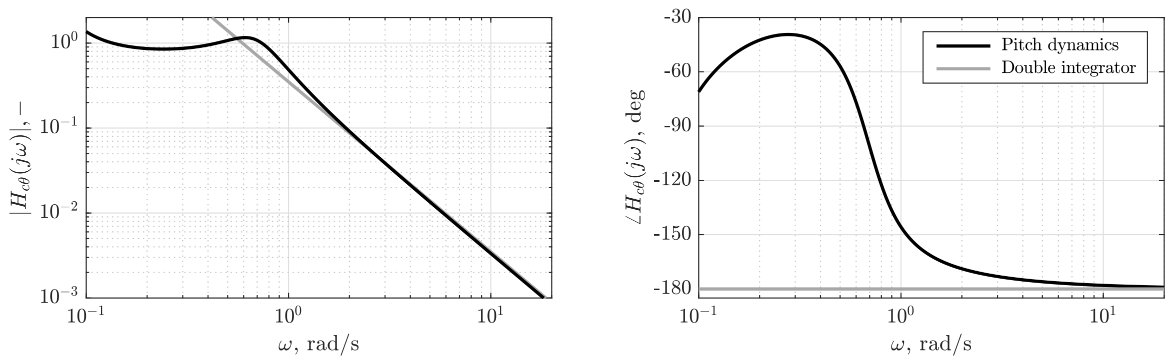


Figure 5-4: Frequency response of linearized aircraft pitch dynamics.

5-2-4 Forcing Functions

To identify human operator control behavior, forcing functions with a specific frequency content can be applied to excite the operator's control system. Designing appropriate forcing functions to be applied is challenging, as they can affect the operator's control behavior and as such, affect the experiment results. To obtain accurate pilot models, forcing functions must meet the following five requirements:

- The forcing function must be unpredictable for the human operator in order to prevent the operator from detecting patterns in and being able to anticipate the signal, in which case the operator would introduce feed-forward behavior and thereby change the control structure from a feedback system to a system including an additional feedforward path [144, 166];
- The forcing function must be difficult enough for the human operator in order to prevent boredom and ensure relatively constant control behavior with high levels of control behavior linearity and thereby maximizing describing function accuracy [144, 167–169];
- The forcing function must also not be too difficult to avoid fatigue from an excessive workload as that could lead to crossover regression, a phenomenon where the human operator adopts a lower tracking bandwidth to improve performance [144, 167, 169];
- The forcing function must have a high signal-to-noise ratio at frequencies of interest to maximize identification accuracy [167]; and
- The forcing function must have a Gaussian magnitude distribution to ensure that describing functions resemble real-life control behavior as closely as possible [166, 169].

To be able to perform pilot model identification in the frequency domain, two additional requirements must be met [144, 166, 167]:

- The forcing function must have a limited number of excitation frequencies to ensure that signal power does not spread out over too many frequencies, causing a lower signal-to-noise ratio; and
- The excitation frequencies should preferably be equally spaced on a logarithmic scale over approximately two decades in order to identify the describing function.

Forcing functions are often composed of a sum-of-sinusoids [144]. In this experiment, the roll and pitch target forcing functions, $f_{t\phi}$ and $f_{t\theta}$, respectively, were also independent sum-of-sines signals defined by Equation (5-5), as used successfully in numerous earlier tracking studies [135, 148, 150, 158, 162–165, 170, 171].

$$f_{t\phi,\theta}(t) = \sum_{k=1}^{N_{t\phi,\theta}} A_{t\phi,\theta}[k]\sin(\omega_{t\phi,\theta}[k]t + \phi_{t\phi,\theta}[k]) \tag{5-5}$$

In Equation (5-5), $A_{t\phi,\theta}[k]$, $\omega_{t\phi,\theta}[k]$ and $\phi_{t\phi,\theta}[k]$ represent the amplitude, frequency and phase of the k^{th} sine in $f_{t\phi}$ or $f_{t\theta}$, respectively. $N_{t\phi,\theta}$ is the number of sines used to build the forcing function.

Due to time pressure, the forcing functions used in the first preliminary experiment were equal to the roll and pitch target forcing functions applied in an earlier experiment investigating human crossfeed in dual-axis manual control with motion feedback [162]. Due to the motion feedback, additional disturbance forcing functions were required in both axes in the earlier experiment. However, because the present research is performed in a fixed-base simulator, as described in Section 5-2-7, disturbance forcing functions are not required. Therefore, only the target forcing functions from the experiment described in [162] were used in this preliminary experiment. The roll and pitch target forcing function parameters can be found in Table 5-1.

Table 5-1: Roll and pitch target forcing function parameters in the first preliminary experiment.

	Roll target signal, $f_{t\phi}$				Pitch target signal, $f_{t\theta}$			
k, -	$n_{t\phi}$, -	$\omega_{t\phi}$, rad/s	$A_{t\phi}$, deg	$\phi_{t\phi}$, rad	$n_{t\theta}$, -	$\omega_{t\theta}$, rad/s	$A_{t\theta}$, deg	$\phi_{t\theta}$, rad
1	9	0.690	1.681	3.075	6	0.460	1.657	3.489
2	16	1.227	1.129	5.049	13	0.997	1.159	0.656
3	31	2.378	0.499	0.760	27	2.071	0.523	6.169
4	45	3.451	0.283	3.956	41	3.145	0.282	4.723
5	56	4.295	0.202	3.475	53	4.065	0.189	0.405
6	76	5.829	0.129	5.546	73	5.599	0.117	6.201
7	106	8.130	0.084	6.222	103	7.900	0.074	2.662
8	142	10.891	0.062	0.217	139	10.661	0.054	0.183
9	195	14.956	0.049	2.639	194	14.880	0.042	0.607
10	233	17.871	0.045	2.373	229	17.564	0.039	2.072

The frequencies of the individual sinusoids, $\omega_{t\phi,\theta}[k]$, are 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 equals $\omega_m = 2\pi/T_m = 0.0767$ rad/s and the measurement time equals $T_m = 2^{13} = 8192$ ms. The measurement time is taken as the last 81.92 seconds of a 90-second run, where the first 8.08 seconds are considered the run-in time, as also seen in many previous tracking studies [134, 153, 163–165]. This run-in time is 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. Using a data sampling frequency of 100 Hz, the measurement time contains the highest power-of-two measurements in the total length of an experiment run.

Details on how the exact target forcing function parameters for the experiment in [162] were determined can be found in [172].

For the second preliminary experiment, new target forcing functions were defined. The parameters of the new roll and pitch target forcing functions can be found in Tables 5-2 and 5-3, respectively.

Table 5-2: Roll target forcing function parameters in the second preliminary experiment.

Roll target signal, $f_{t\phi}$								
k_ϕ, -	$n_{t\phi}$, -	$\omega_{t\phi}$, rad/s	$A_{t\phi}$, deg	$\phi_{t\phi,1}$, rad	$\phi_{t\phi,2}$, rad	$\phi_{t\phi,3}$, rad	$\phi_{t\phi,4}$, rad	$\phi_{t\phi,5}$, rad
1	2	0.153	1.334	0.300	2.381	4.068	4.619	6.002
2	5	0.384	1.239	0.779	3.931	2.995	4.273	1.254
3	11	0.844	0.937	2.880	4.957	6.065	4.753	1.007
4	23	1.764	0.467	2.367	3.478	5.460	1.650	3.055
5	37	2.838	0.238	4.319	0.335	5.556	0.730	2.074
6	51	3.912	0.145	4.056	2.990	0.593	0.550	2.652
7	71	5.446	0.088	1.421	5.516	1.169	4.398	5.213
8	101	7.747	0.055	5.717	1.195	3.397	3.815	3.439
9	137	10.508	0.040	3.634	2.205	2.811	2.204	5.957
10	191	14.650	0.031	3.431	0.527	4.760	6.161	2.335

Table 5-3: Pitch target forcing function parameters in the second preliminary experiment.

Pitch target signal, $f_{t\theta}$								
k_θ, -	$n_{t\theta}$, -	$\omega_{t\theta}$, rad/s	$A_{t\theta}$, deg	$\phi_{t\theta,1}$, rad	$\phi_{t\theta,2}$, rad	$\phi_{t\theta,3}$, rad	$\phi_{t\theta,4}$, rad	$\phi_{t\theta,5}$, rad
1	3	0.230	1.404	6.137	3.088	6.118	2.355	3.703
2	7	0.537	1.229	2.041	5.551	5.407	4.129	0.244
3	13	0.997	0.896	3.634	0.901	3.296	1.360	3.050
4	29	2.224	0.366	2.536	0.616	4.078	2.272	2.251
5	41	3.145	0.218	0.866	0.978	2.904	0.833	5.150
6	53	4.065	0.146	4.636	1.245	2.919	2.333	3.509
7	73	5.599	0.091	4.345	2.019	0.920	5.331	4.573
8	103	7.900	0.058	2.748	4.612	1.687	3.547	4.034
9	139	10.661	0.042	5.681	2.675	4.146	4.951	1.065
10	194	14.880	0.033	3.803	5.144	5.621	3.641	5.280

To meet the requirement of a limited number of excitation frequencies, both the roll and pitch target forcing functions are the sum of $N_{t\phi,\theta} = 10$ individual sinusoids, the same number as used in the forcing functions for the experiment in [162], each with a different amplitude, frequency and phase. According to McRuer et al. [144], at least five sinusoids are needed to form a sufficiently unpredictable forcing function with a close approximation to a Gaussian amplitude distribution. However, it is desired to have a little more sinusoids to be able to capture all the pilot dynamics over the frequency range of interest. The frequencies of the individual sinusoids were changed compared to the ones used for the forcing functions in the first preliminary experiment. The selected integer multiples for the target forcing functions were chosen in such a way that the ten sinusoid frequencies cover the frequency range of human control at regular intervals on a logarithmic scale and thereby satisfying the forcing function requirements for pilot model identification in the frequency domain. Moreover, the integer multiples were selected such that they were not multiples of one another. It was decided to add sinusoids at frequencies lower than those used in the first preliminary experiment in order to make the pilot model describing functions more reliable at lower frequencies.

The amplitude spectrum of a signal determines its power. A second-order low-pass filter was used to determine the amplitudes of the individual sines in both the roll and pitch target forcing functions. The low-pass filter is described by Equation (5-6) and was used in many previous tracking studies [134, 135, 152, 153, 158, 162, 165], where $T_{A_1} = 0.1$ s and $T_{A_2} = 0.8$ s. This led to a forcing function input bandwidth of $\omega_B = 1.26$ rad/s. The amplitude distributions $A_{t\phi,\theta}[k]$ were scaled to attain variances for $f_{t\phi,\theta}$ of $\sigma^2_{t\phi,\theta} = 1.5$ deg^2.

$$A_{t\phi,\theta}[k] = \left| \frac{1 + T_{A_1} j\omega_{t\phi,\theta}}{1 + T_{A_2} j\omega_{t\phi,\theta}} \right|^2 \tag{5-6}$$

To make the target forcing functions randomly appearing to the human operator as well as homogeneous in order to prevent peaks which cause sudden moments of high workload, signals with a Gaussian-like distribution and an average Crest Factor (CF) were desired [173]. The CF is dependent on the choice of the respective phases $\phi_{t\phi,\theta}$ of the individual sinusoids. The CF is determined using Equation (5-7), and is defined as the maximum amplitude of the target signal divided by the Root Mean Square (RMS) of this signal.

$$CF\left(f_t\left(t\right)\right) = \frac{max\left(f_t\left(t\right)\right)}{rms\left(f_t\left(t\right)\right)} \tag{5-7}$$

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 [173]. For both the roll and pitch target forcing functions, five different realizations were used, differing only by the phases $\phi_{t\phi,\theta}$ of the individual sinusoids. The amplitudes and frequencies of the individual sinuosoids were the same for the five different realizations. These five different forcing function realizations in roll and pitch yielded five different forcing function settings, because the m^{th} forcing function realization in roll was always paired with the m^{th} forcing function realization in pitch. The different forcing function realizations were used to assure that it was virtually impossible for participants to memorize the signals.

5-2-5 Control Variables

The control variables for the preliminary experiment are stated below.

Control variables

- Control task
  - Display
  - Controlled aircraft dynamics
  - Sidestick
- Training procedures
  - Duration of single tracking run

5-2-6 Independent Variables

The preliminary experiment was designed not to have any independent variables. Although the first and second preliminary experiments did apply different forcing functions, the focus of the experiment was not on identifying any performance differences caused by the different forcing functions. The reason for changing the forcing functions in the second preliminary experiment compared to the first one was to make the tracking task less predictable for subjects as well as to improve the identification of the pilot describing functions. Additionally, although the total number of tracking runs and the division of the tracking runs over the four training days were different for the two preliminary experiments, these differences had arisen unintentionally.

5-2-7 Apparatus

The preliminary experiment was performed in the fixed-base simulator setup in the HMI Laboratory at the Faculty of Aerospace Engineering at Delft University of Technology, as shown in Figure 5-5. To make roll and pitch control inputs, participants used a control-loaded hydraulic sidestick with $\pm$ 30° excursion in roll and $\pm$ 22° excursion in pitch. Besides using the sidestick, no other control inputs had to be given. The sidestick was installed on the right-side of the seat, which was a fully adjustable aircraft seat. Each participant 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 similar to the size of an attitude indicator in an aircraft. Besides the compensatory display, no other visuals were provided.

The control-relationships of the sidestick were the same as in an aircraft. Moving the sidestick forward pitches down the nose of the aircraft, whereas pulling the stick backwards pitches up the nose. Moving the sidestick left or right causes the aircraft to roll to the left or right, respectively.

Figure 5-5: Fixed-base simulator setup in the Human-Machine Interaction Laboratory with the sidestick installed on the right side of the pilot seat.

5-2-8 Participants

The preliminary experiment was performed by two students of the Delft University of Technology. Both participants were right-handed and had no previous flying experience and little to no previous tracking task experience. The first participant performed the experiment with the first set of forcing functions, whereas the second student performed it with the newly designed forcing functions, as described in Section 5-2-4.

5-2-9 Experiment Procedures

Before the start of the experiment, participants were verbally briefed on the goal of the experiment as well as on the experiment procedures. Additionally, participants were made familiar with the compensatory display and the direction in which they had to move the sidestick in order for a certain error correction to occur. The main instruction to subjects was to simultaneously minimize their roll and pitch tracking errors, presented on the compensatory display, within their capabilities. Any additional questions that subjects had about the experiment and which were not believed to be able to influence the experiment results were also answered during the verbal briefing before the first experiment run.

The aim was to perform 25 tracking runs per day for four days, meaning 100 tracking runs in total. However, due to some problems with starting up the simulator, not every day the same number of runs was performed. The number of runs performed per day for both test subjects can be seen in Figure 5-6. Subject 1 performed 100 tracking runs spread out over four consecutive days, whereas subject 2 performed 95 tracking runs spread out over four experiment days, but with a rest day between experiment days 3 and 4. After each 90-second tracking run, the subject's performance scores in roll and pitch were displayed on the PFD in order to motivate the subject to perform to the best of their abilities. The scores were expressed as the root mean square of the tracking error signals. 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 an affirmative answer, the next run was started. Otherwise, participants were offered to take their time until they felt ready to perform the next run in order to ensure that participants' concentration levels were high and as constant as possible throughout the training session. Additionally, after any of the runs a short break could be taken to alleviate any discomfort that might have occurred from controlling the sidestick or sitting in a fixed position for a prolonged period of time.

	Day 1 *9:00 - 10:00*	*Day 2* *9:00 - 10:00*	*Day 3* *8:00 - 9:00*	*Day 4* *12:30 - 13:30*
Test subject 1	25 runs	20 runs	30 runs	25 runs

	Day 1 *9:00 - 10:00*	*Day 2* *9:00 - 10:00*	*Day 3* *9:00 - 10:00*	*Day 4* *-*	*Day 5* *9:00 - 10:00*
Test subject 2	25 runs	25 runs	25 runs	-	20 runs

Figure 5-6: Preliminary experiment training schedules of test subjects 1 and 2.

5-2-10 Dependent Measures

To quantify the acquisition of manual control skills, the error between the target and attitude signals, and control stick inputs were logged every 0.01 seconds. These error signals and stick inputs were used to analyze task performance and control activity as well as to identify human operator models. An overview of all dependent measures is provided below.

Dependent measures

- Root mean square of tracking error (RMS(e))
- Root mean square of control input (RMS(u))
- Human operator model parameters
  - Pilot gain K_p
  - Lead time constant T_L
  - Response delay τ_e
  - Neuromuscular frequency ω_{nm}
  - Neuromuscular damping ratio ζ_{nm}
- Variance Accounted For (VAF)

The VAF is a measure of the pilot model accuracy in describing the measured human operator data. The eight dependent measures were determined for roll and pitch separately, meaning that a total of 16 objective dependent measures were analyzed. Additionally, the learning curve parameters p_0, p_a and F, as depicted in Equation (4-2), were determined for all dependent measures.

5-2-11 Hypotheses

Based on previous experimental research in which (dual-axis) tracking tasks were employed, the following results were envisioned for the preliminary experiment:

H1: Training causes an improvement in performance and task proficiency [134, 135].
H2: Participants perform better in pitch than in roll [150, 162–165].

These hypotheses were tested using the performance measures defined in Section 5-2-10. An improvement in performance and task proficiency is expressed in terms of a lower RMS(e), increased pilot gain K_p, decreased lead time constant T_L and response delay τ_e and an increased VAF [25, 134, 135].

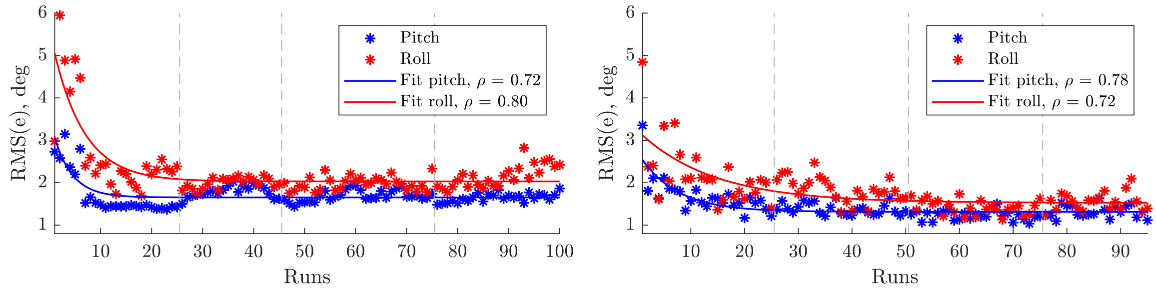
5-3 Preliminary Experiment Results

This section presents the results of the preliminary experiment, which was conducted as described in Section 5-2. The objectives of the preliminary experiment were described in Section 5-1. To accomplish the first objective of determining how many runs it takes for task-naive participants to reach asymptotic task performance, participants' tracking performance and control activity were analyzed. To accomplish the second objective of determining whether the dual-axis aircraft roll and pitch tracking task provides data suitable for pilot model identification, pilot models were fitted for each individual tracking run.

All measured data were analyzed in roll and pitch separately. In figures, the dashed vertical grey lines represent a transition to the next training day.

5-3-1 Tracking Performance

Tracking performance was measured in terms of the root mean square of the tracking error, i.e. the error presented to the pilot on the PFD. The lower the value of RMS(e), the better was the task performance. Figures 5-7a and 5-7b show the root mean square of the tracking error per experiment run, together with fitted learning curves, for test subjects 1 and 2, respectively. The learning curve parameters are presented in Table 5-4. Several conclusions can be drawn from the tracking error results presented in Figure 5-7. The tracking error indeed decreased, i.e. tracking performance improved, throughout training, as was hypothesized in Section 5-2-11. Furthermore, performance in pitch was constantly better than performance in roll, as was also expected from previous studies [150, 162–165]. As can be seen from the learning curves in Figure 5-7a, test subject 1 showed a steep improvement in performance on the first training day, reaching asymptotic performance rather quickly. Test subject 2, on the other hand, showed a more gradual improvement in performance, as presented in Figure 5-7b and confirmed by the lower learning rate F in Table 5-4, but he/she also continued to improve over a longer period of time. A slight learning curve in roll was still observed on day 3. Compared to test subject 1, test subject 2 did reach slightly better performance at the end of training, but he/she also started off with better performance at the beginning of training. These results were confirmed by the initial and asymptotic values of the RMS(e) learning curves in Table 5-4.



(a) Test subject 1.

(b) Test subject 2.

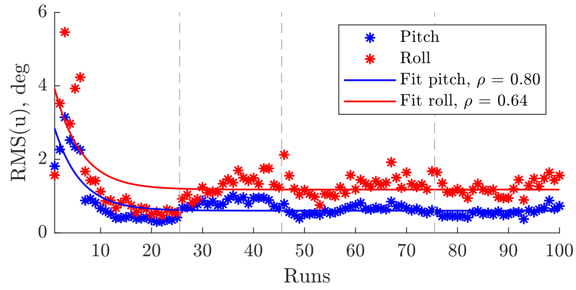
Figure 5-7: Root mean square of tracking error of test subjects 1 and 2.

Table 5-4: Learning curve parameters for root mean square of tracking error of test subjects 1 and 2 (TS = Test Subject).

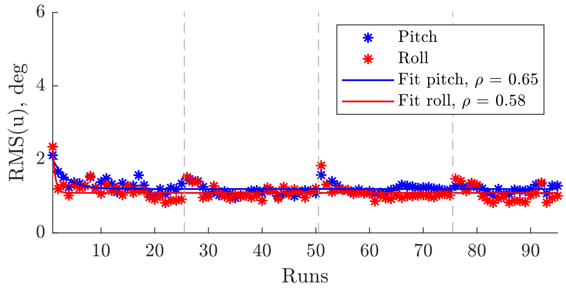
	Learning Curve Parameters					
	RMS(e_θ)			RMS(e_ϕ)		
	p_0, deg	p_a, deg	F	p_0, deg	p_a, deg	F
TS 1	3.520	1.653	0.244	5.599	2.030	0.153
TS 2	2.731	1.314	0.135	3.238	1.526	0.069

5-3-2 Control Activity

Control activity was measured in terms of the root mean square of the control input, RMS(u). Figures 5-8a and 5-8b show the root mean square of the control input per experiment run, together with fitted learning curves, for test subjects 1 and 2, respectively. The learning curve parameters are presented in Table 5-5. The steep performance improvement of test subject 1 in Figure 5-7a was accompanied by a steep decrease in control activity, as presented in Figure 5-8a. Test subject 2 exhibited rather constant control activity throughout training, although his/her control inputs were also relatively low, as presented in Figure 5-8b.



(a) Test subject 1.



(b) Test subject 2.

Figure 5-8: Root mean square of control input of test subjects 1 and 2.

Table 5-5: Learning curve parameters for root mean square of control input of test subjects 1 and 2 (TS = Test Subject).

	Learning Curve Parameters					
	RMS(u_θ)			RMS(u_ϕ)		
	p_0, deg	p_a, deg	F	p_0, deg	p_a, deg	F
TS 1	3.329	0.604	0.178	4.502	1.179	0.174
TS 2	2.403	1.194	0.326	11.45	1.089	0.879

5-3-3 Human Operator Modeling Results

The development of skill-based control behavior throughout training was further analyzed by identifying human operator models in roll and pitch for every run performed by the two test subjects using Genetic MLE [148]. An assessment of the quality of the fitted models was performed in terms of the VAF. Pilot model parameters as well as VAF values of the pilot models of test subjects 1 and 2 are shown in Figures 5-9 and 5-10, respectively. Learning curves are shown when Pearson's correlation coefficient was larger than 0.5. The learning curve parameters are provided in Table 5-6.

Figure 5-9 reveals that the pilot gains K_p of test subject 1 were extremely low. No exponentially increasing curve, as expected from previous training studies [25, 134], could be identified. On the other hand, the lead time constants T_L, neuromuscular frequencies ω_{nm} and damping ratios ζ_{nm} were extremely large. There were no exponential decay curves identified for the human operator lead time constant T_L and response delay τ_e, which were expected from previous training studies [25, 134]. Although the evolution of the VAF throughout training did show somewhat of a trend, the trend was not entirely as expected. VAF values were rather low, especially for pitch, where it even followed an exponential decay curve, instead of an increasing curve as was expected.

Pilot model parameter results of test subject 2 in Figure 5-10 were slightly better than those of test subject 1. The pilot gain K_p in roll showed a slightly exponentially increasing curve as was expected. Overall, the values of the pilot gain K_p were slightly higher and the values of the lead time constant T_L slightly lower than those of test subject 1. However, the lead time constant T_L and response delay τ_e did, again, not show an exponentially decay curve. The neuromuscular parameters were even higher than those of test subject 1. Positively, the VAF values in pitch and roll were larger than those of test subject 1 and showed an increasing trend.

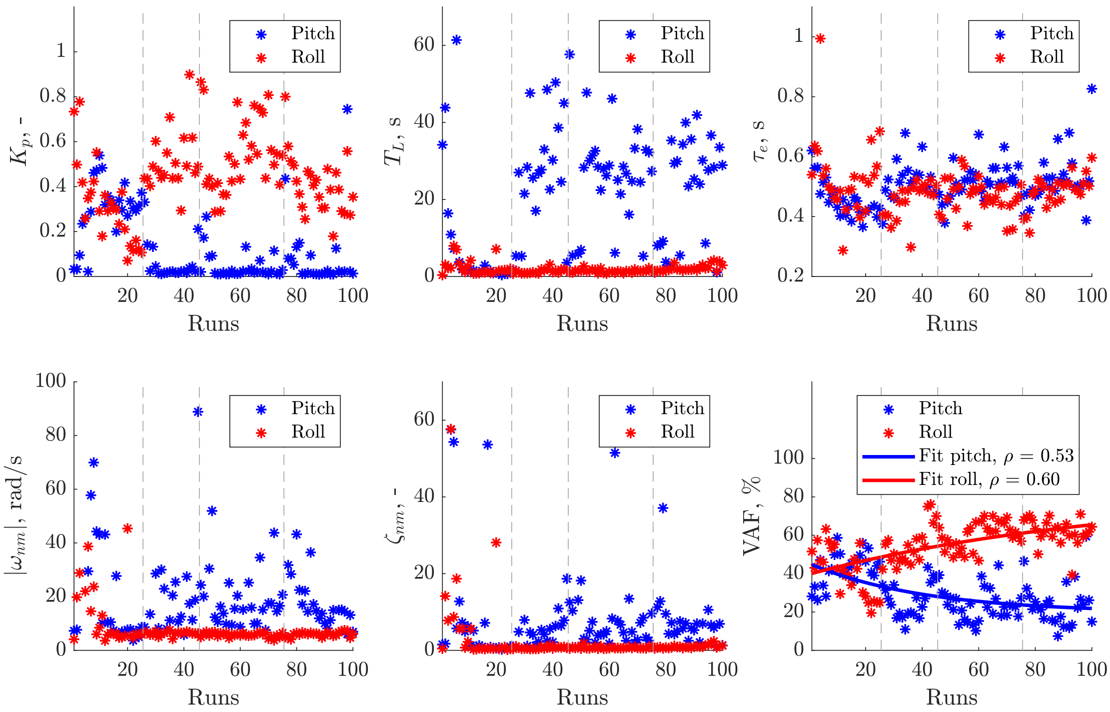


Figure 5-9: Estimated human operator model parameters and Variance Accounted For of test subject 1.

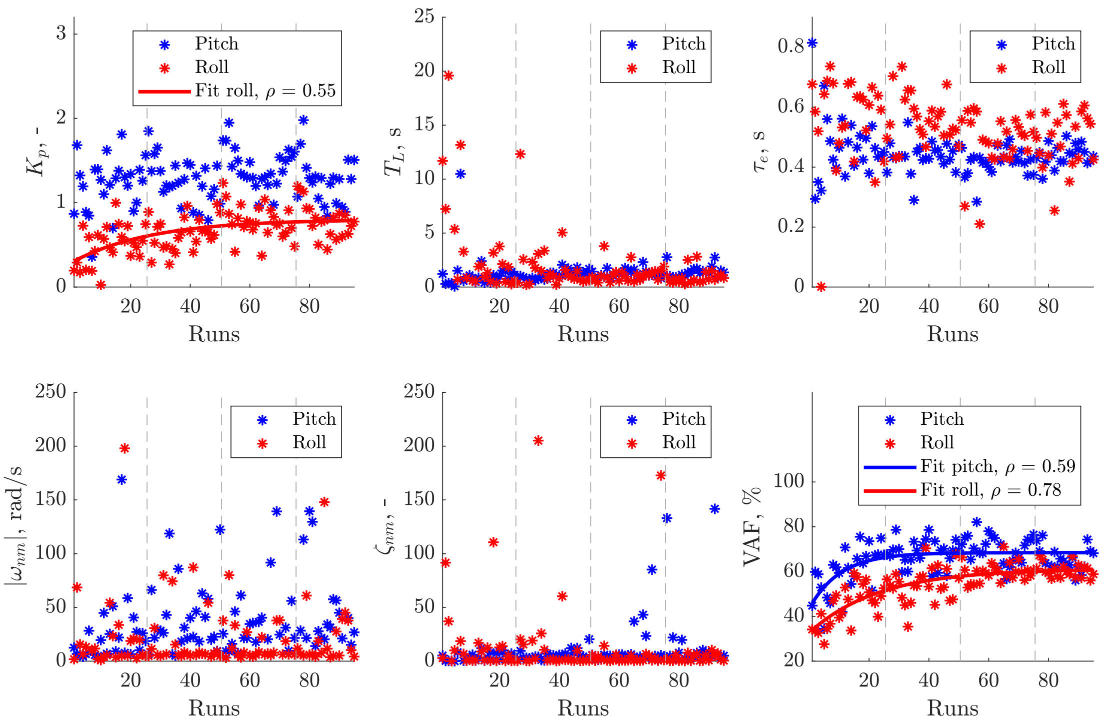


Figure 5-10: Estimated human operator model parameters and Variance Accounted For of test subject 2.

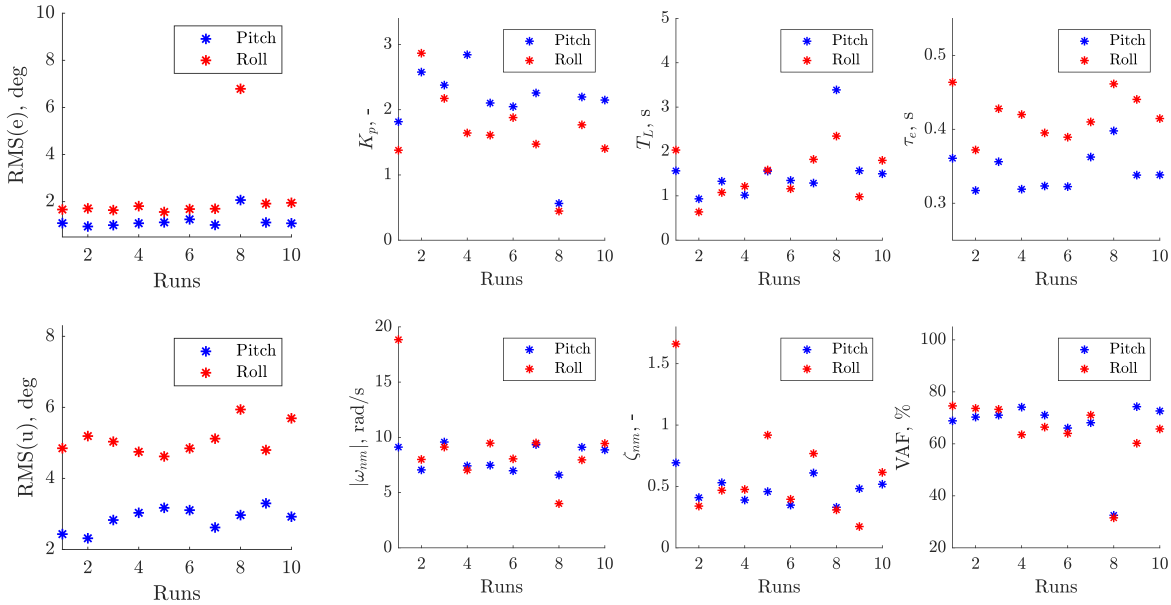
Table 5-6: Learning curve parameters for estimated pilot model parameters and Variance Accounted For of test subjects 1 and 2 (TS = Test Subject).

	Learning Curve Parameters								
	$K_{p\phi}$			VAF_θ			VAF_ϕ		
	p_0, deg	p_a, deg	F	p_0, deg	p_a, deg	F	p_0, deg	p_a, deg	F
TS 1	n/a	n/a	n/a	45.32	19.65	0.024	39.74	83.18	0.009
TS 2	0.288	0.803	0.037	43.55	68.45	0.093	32.68	61.15	0.042

Unfortunately, the pilot model parameter results of test subjects 1 and 2 were not considered good enough to be able to draw a conclusion on whether the dual-axis aircraft roll and pitch tracking task provides data suitable for pilot model identification. Therefore, some additional tracking runs were performed by an experienced subject to be able to draw a definitive conclusion. The results of these additional tracking runs can be found in the next subsection.

5-3-4 Results of Experienced Subject

To be able to draw a more reliable conclusion on the suitability of the dual-axis aircraft roll and pitch tracking task for pilot model identification, ten tracking runs were performed by an experienced subject. The results of these tracking runs are displayed in Figure 5-11, presenting the root mean square of the tracking error and control input as well as the human operator modeling results. Learning curves were not fitted to these results, since only ten tracking runs were performed.



(a) Root mean square of tracking error and control input.

(b) Estimated human operator model parameters and Variance Accounted For.

Figure 5-11: Preliminary experiment results of experienced test subject.

Figure 5-11a shows that task performance was slightly better than the best task performance of test subject 2. Furthermore, control activity was much higher than that of test subjects 1 and 2. Figure 5-11b shows realistic values for the pilot gain K_p, lead time constant T_L and response delay τ_e. All but one neuromuscular frequency were smaller than the largest target forcing function input frequency and all but one neuromuscular damping ratio were smaller than one. Eighteen out of twenty pilot models had a VAF larger than 60%.

5-4 Discussion of Preliminary Experiment Results

The goal of the preliminary experiment was to be able to refine the design of the final experiment by determining how many tracking runs should generally be performed by task-naive participants for them to reach asymptotic task performance. The second goal was to determine whether the dual-axis compensatory aircraft roll and pitch tracking task provides data suitable for pilot model identification. These goals were achieved by measuring tracking behavior in this dual-axis tracking task in a preliminary training experiment with two task-naive participants. Tracking performance, control activity and human operator modeling results in each run and for each participant were used to quantify the development of control skills during extensive training.

Based on findings from previous training experiments [134,135], clear positive developments of control skills were expected to occur (Hypothesis **H1**). Indeed, throughout training, improved tracking performance was observed, with a decrease in the root mean square of the tracking error. Secondly, based on previous dual-axis tracking task experiments [150,162–165], it was hypothesized that consistently better performance would be visible in pitch than in roll (Hypothesis **H2**). Throughout the entire training phase, both test subjects showed lower tracking errors in pitch than in roll.

Some additional remarks concerning tracking performance must be made. Due to the fact that the tracking performance of test subject 1 was slightly worse on day 3 compared to day 2 and even worse on day 4, it is believed that he/she did not perform to the best of his/her abilities, perhaps due to a lack of sleep or a lack of motivation. The low control inputs of test subject 1 throughout training might also be an indication of fatigue or lack of motivation. Additionally, because the tracking performance of test subject 2 was slightly worse on day 4 compared to day 3, it is also suggested that this subject did not perform to the best of his/her abilities on the last training day.

Although test subject 1 did seem to be less involved in the tracking task during the third and fourth training day as is suggested from his/her tracking performance, on the last day of training he/she did realize that in all tracking runs the same target forcing functions were applied. Therefore, it was decided that for test subject 2 multiple forcing function realizations would be used, as described in Section 5-2-4, in order to avoid that this test subject would have the same realization. Test subject 2 was not able to recognize any of the five forcing function realizations during the entire training phase.

From the tracking errors of test subject 2, it could be concluded that four training days of 25 tracking runs each would be suitable for the final experiment. Although test subject 2 was still learning on the third training day, due to the worse performance on day 4, it could not be determined whether test subject 2 would have shown even better performance on day 4 compared to day 3 if he/she had performed to the best of his/her abilities. Nonetheless, given the fact that test subject 2 already showed relatively good performance on the first training run, it can be expected that subjects who perform less well on their first training day due to a less 'natural' ability in performing the task will still be showing a learning curve on the fourth training day. Since the use of more than four training days is not possible due to logistical reasons, the use of four training days is recommended for the final experiment. No more than 25 training runs should be performed per day to avoid influences of fatigue or boredom. Additionally, a short break in the middle of a training session would benefit concentration levels. It was expected that a total of 100 training runs would lead to a good balance between "overlearning" and "underlearning" between experiment subjects.

When looking at the pilot model parameters, surprisingly, none of these parameters showed a clear learning trend during training, except for a slight increase in the pilot gain in roll of test subject 2.

Although unexpected, this is consistent with the research results described in [135], in which none of the pilot model parameters showed a clear learning trend in the training of a compensatory roll-axis tracking task with visual cues. Furthermore, the extremely large neuromuscular parameters of both test subjects could be manually adjusted by shifting the large neuromuscular frequency to the maximum input frequency of the target forcing function and investigating the influence of this shift on the VAF. If the influence on the VAF would be minimal, the smaller neuromuscular frequency could be accepted. This method has successfully been applied before during an investigation into quantifying loss of motor skills due to Parkinsons Disease [154]. However, this method was not applied during this preliminary experiment, because it would not solve the low VAF values found in Figure 5-9. Additionally, because of the extremely high lead time constants T_L, the extremely low pilot gains K_p of especially subject 1, and the low VAF values, it was decided that the pilot model parameter results of test subjects 1 and 2 were not considered reliable enough to be able to draw a conclusion on the suitability of the dual-axis aircraft roll and pitch tracking task for pilot model identification. Therefore, some additional tracking runs were performed by an experienced subject to be able to draw a definitive conclusion. The experienced subject did exhibit pilot model parameters as expected from previous experiments [134, 135]. Additionally, almost all pilot models had a VAF higher than 60%, a value similar to those observed in previous single-axis tracking task experiments in which individual experiment runs were analyzed [134, 135]. From these results, it was concluded that the dual-axis aircraft roll and pitch tracking task is suitable for pilot model identification.

5-5 Conclusions and Recommendations

The following conclusion was drawn from the preliminary experiment:

- The dual-axis aircraft roll and pitch tracking task is suitable for pilot model identification (**PE objective 2**).

The following recommendations were made for the final experiment design:

- The training phase should consist of 100 tracking runs in total, spread out over four consecutive days (**PE objective 1**).
- No more than 25 tracking runs should be performed each day to avoid effects of fatigue or boredom.
- A short break should be scheduled in the middle of each session to keep concentration levels as constant as possible.
- Several forcing function realizations should be used in order to avoid that later in training experiment participants start to recognize the target forcing functions.

Chapter 6

Experiment Design

After performing the literature review, it was found that it is extremely difficult to compare the results of previous research on the retention of manual control skills due to the different performance measures used. The current research will try to set a new standard for measuring skill retention by applying a cybernetic approach. This cybernetic approach has been used in training studies before, but not yet in retention studies. To investigate the retention of manual control skills, a human-in-the-loop experiment was performed. The preliminary experiment discussed in the previous chapter was conducted to optimize the final experiment design.

This chapter describes the final experiment design. Its supporting research question is posed in Section 6-1, after which the experiment setup is discussed in Section 6-2. Since the training phase and first retention tests of the experiment have already been performed, the experiment design has been frozen already.

6-1 Research Question

Now that a good understanding of the present state of research into the retention of manual control skills has been obtained and a preliminary experiment with a dual-axis aircraft roll and pitch tracking task has been performed, the objective of the current research can be defined further by narrowing down the preliminary research questions posed in the introduction. The objective of the current research can now be described by the research question and three sub-questions below.

Research question:	"To what extent do manual control skills of novices decay during periods of non-practice?"
Sub-question 1:	"What trend does the decay curve of manual control skills of novices follow?"
Sub-question 2:	"What is the optimal retention interval to ensure that manual control skills of novices do not decay significantly, while at the same time minimizing the amount of refresher training?"
Sub-question 3:	"How does the reacquisition rate of manual control skills of novices during retention testing compare to their initial acquisition rate?"

It was decided to focus on the skill retention of novices, because it is of utmost importance to the outcome of this research that during the retention interval participants do not involve themselves in activities that would influence their skill retention. Although it would have been desirable to have pilots participate in the experiment in order to maximize the relevance of this research, it is unrealistic to expect pilots to refrain from flying during the course of the experiment. Also, chances are that general aviation pilots would be involved in leisure flying during the retention interval. Moreover, finding enough (general aviation) pilots in time for such an extensive experiment is almost impossible.

6-2 Experiment Setup

To determine to what extent the manual control skills of novices decay during periods of non-practice, a human-in-the-loop experiment was performed in the HMI Laboratory at the Faculty of Aerospace Engineering at Delft University of Technology. This section describes the experiment setup.

6-2-1 Dual-Axis Control Task

The control task used was the same as the one in the preliminary experiment. The schematic representation of the dual-axis aircraft roll and pitch tracking task is depicted in Figure 5-1. However, if indications of noticeable crossfeed are found during data analysis, the human operator might be modeled as in the schematic representation shown in Figure 6-1.

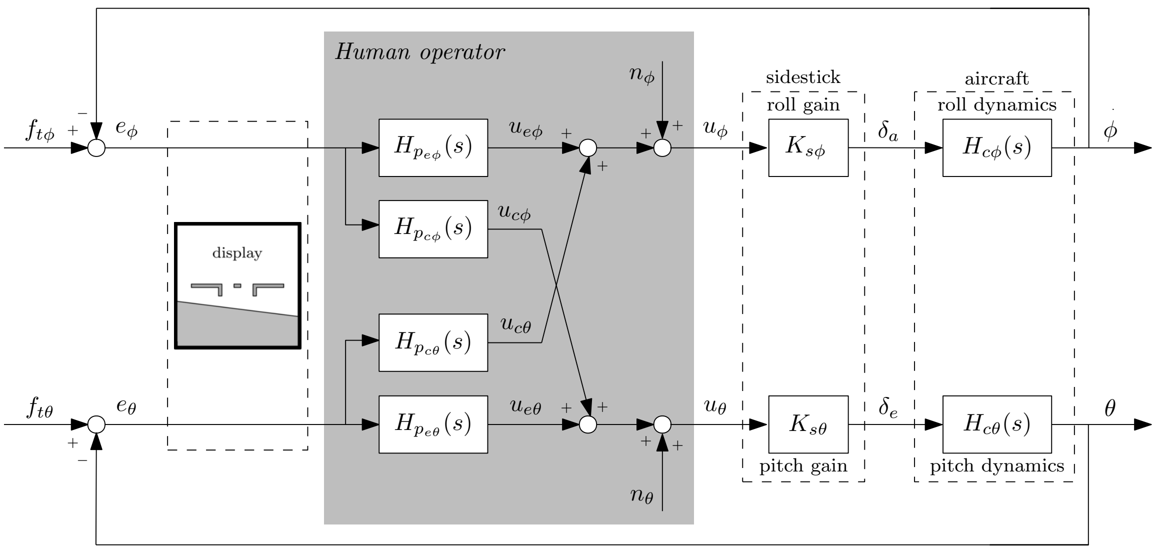


Figure 6-1: Schematic representation of the compensatory dual-axis aircraft roll and pitch tracking task with crossfeed.

6-2-2 Human Operator Modeling

The pilot model considered for data analysis is the same as the one used in the preliminary experiment. This pilot model is described by Equations (5-1) and (5-2).

6-2-3 Controlled Aircraft Dynamics

The controlled aircraft dynamics used were the same as the dynamics used in the preliminary experiment. These dynamics are described in Section 5-2-3.

6-2-4 Forcing Functions

The forcing functions used in the experiment were the same as the ones used in the second preliminary experiment. The parameters of the roll and pitch target forcing functions can be found in Tables 5-2 and 5-3, respectively. Five different forcing function realizations were used to ensure that participants do not remember parts of the target forcing functions.

6-2-5 Control Variables

The control variables in the experiment were the same as the ones in the preliminary experiment, complemented with a few additional ones. For completeness, the total list of control variables is stated below.

Control variables

- Control task
  - Display
  - Forcing functions
  - Controlled aircraft dynamics
  - Sidestick
- Training procedures
  - Experiment briefing (written and verbally)
  - Duration of single tracking run
  - Number of training runs
  - Division of training runs over training days

The control task is elaborately explained in Section 5-2-1. The display is depicted in Figure 5-2.

6-2-6 Independent Variables

In order to determine the trend of the skill decay curve and the 'optimal' retention interval to ensure that skills do not decay significantly, while at the same time minimizing the amount of refresher training, participants were split up into three groups. Each group performed refresher training after a different retention interval. The experiment procedures were designed such that all participants would perform their last retention test after the same amount of time, meaning that all groups performed a different number of retention tests based on their retention interval length.

The three different groups, which can be regarded as three different experiment conditions, are shown in Table 6-1. The retention intervals provided were calculated from the end of training. All subjects performed their last retention test six months after completing their training.

Table 6-1: Experimental conditions used in the human-in-the-loop experiment.

Group	Retention interval	Number of retention tests
1	6 months	1
2	3 months	2
3	2 months	3

Because of the between-subjects design of the experiment and the fact that each group had to consist of a considerable number of subjects due to the uncertainty whether subjects who had completed training would also be available for retention testing, it was decided that the experiment would be limited to only one independent variable.

Independent variable

- Retention interval length / number of retention tests

The retention interval length and the number of retention tests were coupled. This also determined the total number of retention runs participants had to perform, as is explained more elaborately in Section 6-2-9.

6-2-7 Apparatus

Exactly the same apparatus was used as in the preliminary experiment. The apparatus is described in Section 5-2-7.

6-2-8 Participants

A total of 39 task-naive participants completed the training phase of the experiment and all gave written consent for their participation. They also agreed on refraining from participation in any other tracking task or flying experiments until the last retention test had taken place. All participants were students at Delft University of Technology, except for one, who had graduated from the university five months before the training phase of the experiment. The majority of students were from the Faculty of Aerospace Engineering. The youngest participant was 18 years old at the time of training, the oldest 32. On average, the age was 21.0 years, with a standard deviation of $\pm$ 3 years. Twenty-nine participants were male, and ten female. Most of the participants were right-handed, but all of them were comfortable with operating the sidestick with their right hand.

Participant requirements

- Right-handed or comfortable operating the sidestick with the right hand
- No pilot experience and little tracking task experience
- No participation in other tracking task or flying experiments until the last retention test has taken place
- Available for training and a maximum number of three retention tests, where the last retention test will take place six months after training

6-2-9 Experiment Procedures

To evaluate the retention of manual control skills, the human-in-the-loop experiment consisted of two phases, referred to as the *training* phase and the *retention* phase. During the training phase, all task-naive participants received ab-initio training in the dual-axis tracking task under the same conditions. During the retention phase, the same tracking task was performed as during training.

Two to four days before the start of their training phase, participants received the experiment briefing, provided in Appendix C, which had to be read before the start of the experiment. In the experiment briefing the goal of the experiment, the dual-axis tracking task and the experiment procedures were explained. On the first day of training, before starting the experiment, participants signed an experiment consent form, provided in Appendix D and received a short in-person briefing covering the most important aspects of the written experiment briefing as well as safety procedures. This in-person briefing also provided participants with the opportunity to ask any questions they had after reading the experiment briefing. After the briefing, participants were demonstrated how to adjust their seating position. They could adjust the seating position to their liking to make sure that they could comfortably operate the sidestick. During the training phase, participants also filled out a one-time questionnaire, provided in Appendix E. This questionnaire was used to determine participants' previous experience with tracking skills.

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. For four consecutive working days, participants performed one session per day in order to enable skill improvement between training sessions, an effect known as offline learning (i.e, consolidation of learned motor skills while not physically performing the task). Sleep enables offline skill improvement following explicit (intentional) learning [174]. Although there is no solid consensus yet on the optimum amount of time between consecutive training sessions, in a meta-analysis by Kantak and Winstein [175] 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. 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 were notified of their performance in roll and pitch by displaying their scores on the PFD. These scores were expressed as the root mean square of the tracking error signals in roll and pitch, respectively. 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 an affirmative answer, the next run was started. Otherwise, participants were offered to take their time until they felt ready to perform the next run in order to ensure that participants' concentration levels were high and as constant as possible throughout the training session.

Although no actual evidence has been found favoring spaced practice over massed practice for the retention of motor skills, as can be read in Section 3-1, it was decided to apply a spaced practice schedule. 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. 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 in three experiment groups based on their training performance and their availability for retention testing.

In order to form retention groups, training performance of each participant was determined by averaging their tracking errors, control inputs and pilot model parameters of the last ten training runs (runs 91 - 100). Additionally, the learning curves of their tracking errors over the entire training phase were evaluated. Participants' training performance was evaluated for roll and pitch separately. Subsequently, groups were formed such that there were no significant differences in error scores, control activity, pilot model parameters or any of the learning curve parameters between the three groups.

The groups were differentiated from one another in their retention interval length and the number of retention tests they would perform, as was already illustrated in Table 6-1. The first group, *Group 1*, only performed a single retention test after a retention interval of six months. The second group, *Group 2*, performed two retention tests with retention intervals of three months in between. The last group, referred to as *Group 3*, performed a total of three retention tests with retention intervals of two months in between. This means that all participants performed their final retention test six months after the end of training. The entire experiment setup is illustrated in Figure 6-2. At every retention test participants were asked whether they had been involved in any activities during the retention interval that could either positively or negatively affect their retention performance. The last retention test of each group was built up in the same manner as the individual training sessions, meaning that the test consisted of 25 90-second runs with a five-minute break between the first 15 and last ten runs. The other retention tests, i.e. the first retention test of Group 2 and the first two retention tests of Group 3, consisted of only five 90-second runs without a break. The five-run retention tests were kept short on purpose, to be able to capture the performance of participants at that moment in time, while at the same time avoiding extensive additional learning. All last retention tests were longer again to be able to establish participants' relearning rate of lost skills, if any skill decay had occurred after six months, which would help in answering sub-question 3 in Section 6-1.

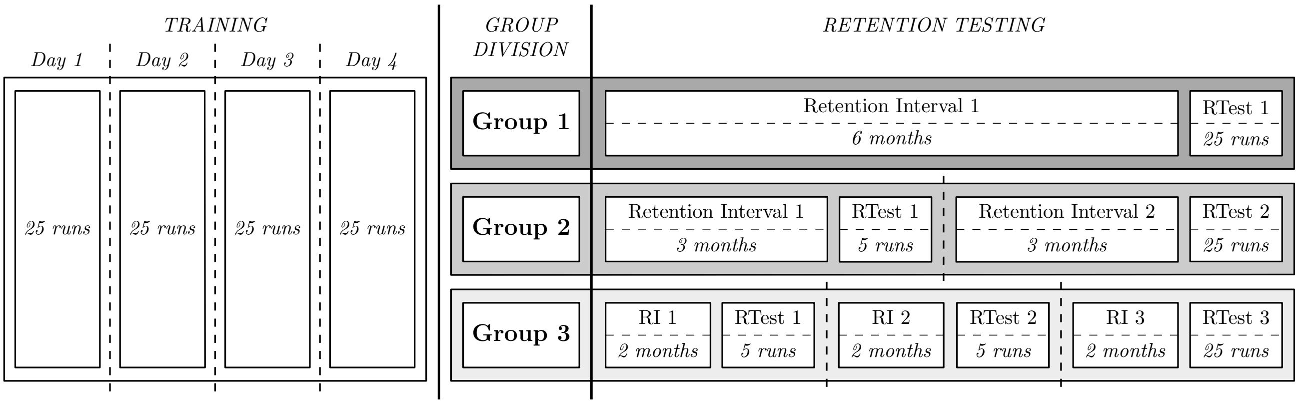


Figure 6-2: Experiment setup (RI = Retention Interval, RTest = Retention Test).

The first retention tests of each group were used to identify the trend of the skill decay curve, and thereby answering sub-question 1 in Section 6-1. The last retention tests of each group, meaning the retention tests six months after the end of training, were used to identify the 'optimal' retention interval to prevent skill decay, while at the same time minimizing the amount of refresher training, and thereby trying to answer sub-question 2.

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. The exact training schedules of all subjects can be found in Appendix F. This appendix also describes the exceptions to the experiment procedures.

6-2-10 Dependent Measures

To be able to quantify skill development, decay and retention, the error between the target and attitude signals, and participants' control stick inputs were logged every 0.01 seconds. The dependent measures were the same as the ones used in the preliminary experiment, complemented with a few additional ones. For completeness, the total list of dependent measures is stated below.

Dependent measures

- Root mean square of tracking error (RMS(e))
- Root mean square of control input (RMS(u))
- Human operator model parameters
  - Pilot gain K_p
  - Lead time constant T_L
  - Response delay τ_e
  - Neuromuscular frequency ω_{nm}
  - Neuromuscular damping ratio ζ_{nm}
- Variance Accounted For (VAF)
- Crossover frequency ω_c
- Phase margin ϕ_m

All ten dependent measures were analyzed in roll and pitch separately, meaning that a total of 20 objective dependent measures were determined. Also, the learning curve parameters p_0, p_a and F, as depicted in Equation (4-2), were determined for all dependent measures.

Moreover, next to participants' error signals and stick inputs, metadata was gathered. The participants' age was recorded, because control strategy might vary with the human controllers' age. Furthermore, some data regarding participants' previous experience with tracking was determined using a questionnaire.

6-2-11 Experiment Confounds

The complex experiment setup described in Section 6-2-9 introduces three important experiment confounds. These confounds are explained below.

Participants are not trained to the same "relative" level
As explained in Section 3-1, certain people have a more "natural" ability than others in performing a task without prior practice. Since, as described in Section 6-2-9, all participants receive the same number of training runs, this could lead to a situation where at the end of the training phase, certain participants have reached asymptotic performance, whereas others are still in the learning phase. This means that at the end of training, part of the participants will have overlearned the tracking task, whereas others might have actually "underlearned" the task. Since overlearning has been identified as having a positive influence on retention, as stated in Section 3-1, this has to be taken into account during data analysis. In an ideal situation all participants would be trained until they have just reached asymptotic performance, meaning that they neither overlearn nor underlearn the task. This is desired because the experiment is designed to only look at skill retention as a function of time, not at the effects of overlearning. However, training all participants till asymptotic performance would mean that the number of training runs would need to be tailored to the individual subjects and can only be determined while training is taking place. Unfortunately, this scheduling uncertainty cannot be accommodated for, because of the availability of the HMI Lab, having to avoid scheduling training on the weekends as well as the large number of participants required. Having to continue training after a weekend off would introduce a different confound. It would also be undesirable to have participants perform their entire training on a single day because of the possible introduction of fatigue or boredom, which could make it impossible to capture participants' true learning curves. Therefore, it was decided to accept the confound of "overlearning" and to take it into account during data analysis.

Participants not performing training on the same time every day
Because almost all participants are students, training has to be scheduled around lectures. This means that training cannot take place at the same time every day, which introduces a circadian confound. Because of the large number of participants required and the limited time available for the entire training phase, this confound cannot be avoided.

Retention intervals are not exact
The "ideal" retention intervals for the human-in-the-loop experiment are mentioned in Section 6-2-9. However, the real retention intervals will differ slightly from the ideal ones due to participant availability. Especially the two-month retention interval has some margin, because the first retention test of Group 3 will take place in the summer holidays of the Delft University of Technology. These differences in retention interval length between participants will have to be taken into account during data analysis.

6-2-12 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 hypotheses have been formulated for the current research.

> ***H1: Training causes an improvement in performance and task proficiency.***
>
> *This improvement is achieved through:*
> - *Decreased RMS(e)*
> - *Increased crossover frequency* ω_c
> - *Increased phase margin* ϕ_m
> - *Increased pilot gain* K_p
> - *Decreased lead time constant* T_L
> - *Decreased response delay* τ_e
> - *Increased VAF*

Clear effects of training are expected to occur, as has been seen in previous training experiments concerning tracking tasks. These effects are visible through the above mentioned changes in the dependent measures [25, 134, 135].

> ***H2: Participants perform better in pitch than in roll both during training and retention testing.***
>
> *This performance difference can be seen in lower RMS errors in pitch than in roll.*

It has been seen in previous dual-axis aircraft roll and pitch tracking task experiments that RMS errors were lower in pitch than in roll [150, 162–165]. Although none of these experiments included retention testing, there is no reason to believe that this would be different at later retesting moments.

> ***H3: Skill decay can be captured by a positively accelerating decay curve.***

In an earlier experiment concerning the skill retention of a visual approach and landing task it has been experienced that skill degradation was moderate for the first three months, but increased sharply after that, hence, following a positively accelerating decay curve [97]. The same skill decay trend was found in an experiment concerning the retention of helicopter flying skills, in which skill decay started to accelerate after a non-utilization period of six months [132]. Because these flying tasks are the most comparable to the control task used in the current research, it is hypothesized that the same skill decay trend will be seen in this experiment.

> ***H4: During the last retention test six months after training, best performance and task proficiency will be shown by Group 3, whereas worst performance and task proficiency will be shown by Group 1.***

As described in Section 3-1, individuals perform better at retention testing if they are provided with some form of practice during the retention interval [37, 96–98]. When comparing the last retention tests of each group, the experiment setup can also be seen as if all groups have a retention interval of six months, during which Group 2 receives one practice moment mid-interval and Group 3 receives two practice moments. Since Group 1 receives no practice at all during the retention interval and Group 3 receives the most practice, it is hypothesized that Group 1 will perform worst after six months and Group 3 will perform best.

> ***H5: During the last retention test six months after training, degraded control skills of all three groups will be reacquired at a faster rate than the initial acquiring rate during the training phase.***

As described in Section 3-4, previous retention experiments concerning motor skills have consistently shown that retraining after a retention interval up to performance levels achieved at the end of training requires less time than initial training [36, 79, 93]. There are no indications to believe that this will be different for the current research.

Chapter 7

Training Results

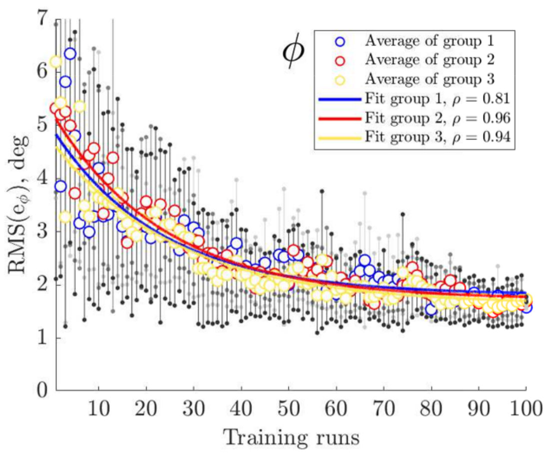
This chapter presents the results of the training phase of the final experiment, which, as stated in Chapter 6, has already been completed. Based on participants' training performance, which was evaluated in pitch and roll separately, three experiment groups were formed, as shown in Figure 6-2. The goal for the group forming was to have no significant differences in training performance between the three groups to be able to compare their retention performance in a reliable manner. Therefore, this chapter presents and compares the training results of the three different experiment groups. Section 7-1 presents the group learning curves for the dependent measures. After that, Section 7-2 shows the differences between the three groups in tracking performance, control activity and human operator model parameters at the end of training, an important criteria used in the forming of the experiment groups. This is followed in Section 7-3 by an evaluation of the differences between the three groups in the parameters of the tracking error learning curves, which was the second criteria used in the group forming. In Section 7-4 a discussion of the training results is provided. Lastly, a brief overview of the next steps of the data analysis is provided in Section 7-5. All figures in this chapter show the data from Group 1 in blue, data from Group 2 in red and data from Group 3 in yellow.

7-1 Group Learning Curves

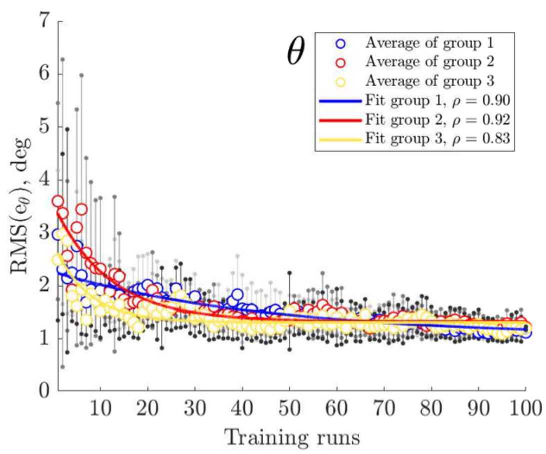
This section presents the group learning curves for tracking performance, control activity and human operator model parameters. The colored data points indicate the group means of each run, while the gray error bars indicate the 95% confidence intervals of the mean data. Additionally, for the tracking performance and the control activity a statistical analysis was performed to compare the measured data at the beginning and at the end of training in order to determine the significance of the training effects. This *training* comparison compared the averages of each participant on runs 1 - 5 and 96 - 100. When the group data was sufficiently normally distributed, the statistical test utilized was the dependent t test. However, in most cases the Wilcoxon signed-rank test was applied, since large between-subject variabilities often led to a violation of the normal distribution assumption.

7-1-1 Tracking Performance

Tracking performance was measured in terms of the root mean square (RMS) of the roll and pitch errors, i.e., the errors presented to the human operator on the PFD. The lower the value of RMS(e), the better the task performance was. Figure 7-1 shows the average group RMS(e) per experiment run, together with the 95% confidence intervals of the mean data and fitted learning curves as described in Section 4-1. The fit of the learning curves was evaluated using Pearson's correlation coefficient ρ. Figures 7-1a and 7-1b show the results in roll and pitch, respectively. The parameters of the fitted learning curves are presented in the left part of Table 7-1. As can be seen in Figure 7-1a, the RMS(e) learning curves in roll were very similar for the different groups, a desired result of the group division. At the start of training, the task-naive participants had an average tracking error in roll of approximately 5°, which decreased to approximately 1.8° at the end of training. However, the learning curves of the RMS(e) in pitch differed slightly more between the three groups, as can be seen in Figure 7-1b. Whereas at the start of training, the average tracking error in pitch differed between 2.3° and 3.6° between the groups, at the end of training this had decreased to about 1.3° for all three groups. These results can also be seen in the learning curve parameters in the left part of Table 7-1. The average tracking errors were slightly higher than those seen in an earlier dual-axis training experiment [150]. However, this is not surprising, since the earlier experiment was performed with motion feedback, and as stated in Section 4-2 task proficiency is often better when motion feedback is present [152, 153]. Additionally, the earlier experiment was performed with general aviation pilots instead of task-naive participants, which have of course more experience, and therefore, showed better performance. The statistical analysis results in the right part of Table 7-1 show that the performance improvement during training was significant in both roll and pitch for all three groups. The detailed statistical results can be found in Appendix H, in Figures H-1 through H-4. Pearson's correlation coefficients were relatively high for all learning curves, meaning that the learning curves accurately represent the mean group data of the individual experiment runs. Additionally, pitch tracking performance was consistently better than roll tracking performance throughout the entire training phase for all three experiment groups, as was expected from earlier dual-axis tracking task experiments [150, 162–165]. However, the difference in performance between pitch and roll decreased as training continued. While at the start of training, the performance difference in pitch and roll was approximately 2°, this difference had decreased to about 0.5° on average between the three groups at asymptotic performance, as can be seen in Table 7-1. Overall, reaching asymptotic performance in roll required more experiment runs than reaching asymptotic performance in pitch.



(a) Roll.



(b) Pitch.

Figure 7-1: Average root mean square of tracking error per group.

Table 7-1: Learning curve parameters and statistical analysis for tracking error per group.

RMS(e)	Learning Curve Parameters					
	Roll ϕ			Pitch θ		
	p_0, deg	p_a, deg	$F(\cdot 10^{-2})$	p_0, deg	p_a, deg	$F(\cdot 10^{-2})$
Group 1	4.97	1.81	4.32	2.26	1.04	2.17
Group 2	5.26	1.72	4.12	3.57	1.32	8.23
Group 3	4.72	1.64	3.75	2.81	1.31	12.7

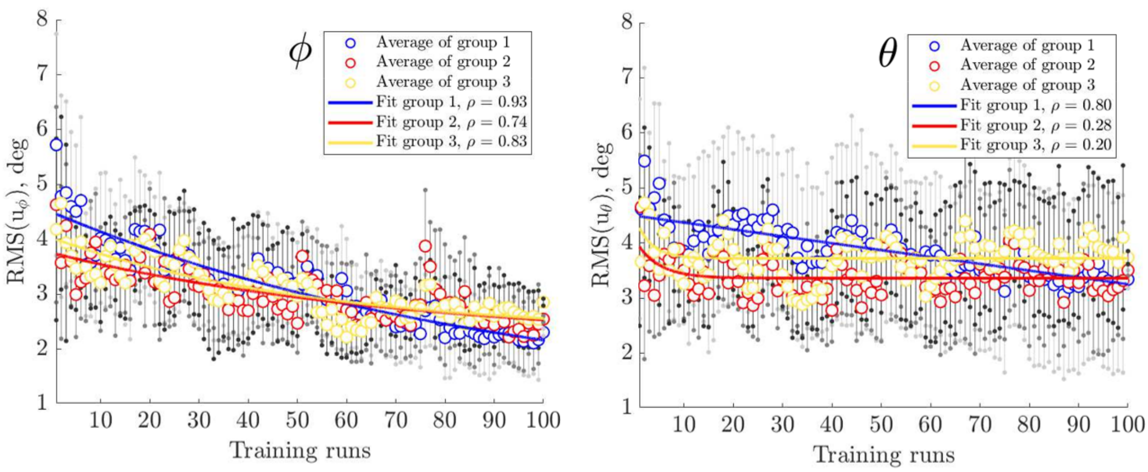
RMS(e)	Statistical Significance	
	Roll ϕ	Pitch θ
	Sig.	Sig.
Group 1	$**^a$	$**^a$
Group 2	$**^a$	$**^a$
Group 3	$**^a$	$**^a$

[a] At least one sample not normally distributed, Wilcoxon signed-rank test applied instead of dependent t test.

Legend:
$**$ = highly significant ($p < 0.01$)
$*$ = significant ($0.01 \leq p < 0.05$)
$-$ = not significant ($p > 0.05$)

7-1-2 Control Activity

Figure 7-2 shows the change in pilot control activity throughout training. As can be seen in Figure 7-2a, control input in roll decreased during training. However, control input of Group 1 decreased more than that of the other two groups. While at the start of training, Group 1 had a higher control input in roll than the other two groups, at the end of training, its control input was lower. However, the difference in control input between the start and end of training was significant for all groups, as shown in Table 7-2. Also, Groups 2 and 3 had very similar control input learning curves, especially in the second half of training. As can be seen in Table 7-2, their asymptotic control input was very similar. Furthermore, control input in pitch gradually decreased throughout training for Group 1. This can also be seen from the very low asymptotic value p_a and learning rate F in Table 7-2. According to the statistical analysis, the change in control input from the start to the end of training was significant. Control input in pitch of Groups 2 and 3 showed a relatively large amount of spread, as was also expressed through the low correlation coefficients. Therefore, no learning curve parameters for control input in pitch are provided for those two groups. The significant spread in control input data is consistent with earlier findings [134, 152, 153]. No significant difference in control input at the start and at the end of training was found for the two groups. The detailed statistical results can be found in Appendix H, in Figures H-5 through H-9.



(a) Roll. **(b)** Pitch.

Figure 7-2: Average root mean square of control input per group.

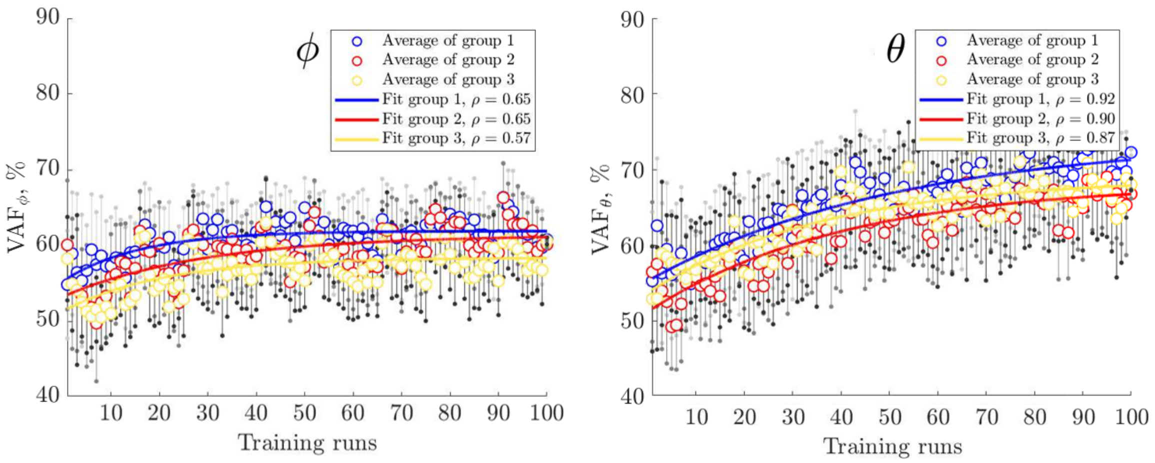
Table 7-2: Learning curve parameters and statistical analysis for control input per group.

RMS(u)	Learning Curve Parameters						Statistical Significance	
	Roll ϕ			Pitch θ			Roll ϕ	Pitch θ
	p_0, deg	p_a, deg	$F(\cdot 10^{-2})$	p_0, deg	p_a, deg	$F(\cdot 10^{-2})$	Sig.	Sig.
Group 1	4.50	0.89	1.04	4.50	-48.8	0.024	$**$	$*^a$
Group 2	3.75	2.11	1.35	n/a	n/a	n/a	$**^a$	$-^a$
Group 3	4.03	2.27	1.81	n/a	n/a	n/a	$**$	$-^a$

7-1-3 Human Operator Modeling Results

The development of skill-based control behavior throughout training was further analyzed by identifying pilot models for all individual training runs using Genetic Maximum Likelihood estimation [148]. An assessment of the quality of the fitted models was performed in terms of the VAF. The group results are shown below.

The average VAF of each group throughout training is shown in Figure 7-3. Often, human operator data is averaged between consecutive experiment runs to increase the accuracy of the human operator model, since data averaging results in a decrease in the amount of remnant noise in the signal used for identification as well as in an increase of the linearity of the measured human control behavior. In that case, the VAF is usually around 80% to 90% [148]. However, in the current research changes in pilot model parameters throughout training and retention testing are crucial to be able to evaluate the development, decay and retention of manual control skills. Averaging results between experiment runs would mask the training and retention effects. Therefore, it was decided to fit pilot models to each individual experiment run, resulting in a lower VAF. In previous single-axis tracking task experiments, in which individual experiment runs were evaluated, the majority of pilot models had VAFs between 60% and 80% [134, 135]. However, in dual-axis tracking task experiments, such as the current one, slightly lower VAFs can be expected, since pilots have to divide their attention between two axes, causing more non-linearities in pilot behavior.



(a) Roll. **(b)** Pitch.

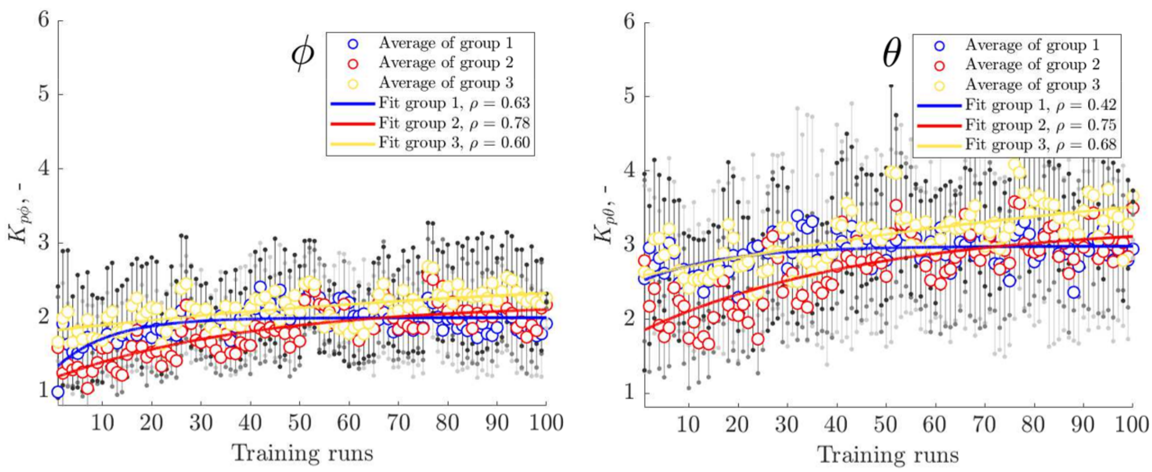
Figure 7-3: Average Variance Accounted For per group.

In Figure 7-3 it can be seen that the shape of the VAF learning curves was very similar for the three groups, both in roll and in pitch. Whereas Group 1 had the highest VAF in both roll and pitch, Group 3 had the lowest VAF in roll, and Group 2 had the lowest VAF in pitch. At the start of training, the VAF was very similar in roll and pitch, as can also be seen in Table 7-3, and took on average values somewhere between 51% and 55% for the three groups. However, at the end of training, the VAF was significantly higher in pitch, 66% to 72%, than in roll, 58% to 62%, for all three groups. Whereas the VAF in roll reached asymptotic performance in the second half of training, the VAF in pitch had not yet reached the asymptotic values presented in Table 7-3 at the end of training. It must also be noted that Pearson's correlation coefficient ρ was considerably higher for the VAF in pitch than in roll.

Table 7-3: Learning curve parameters for Variance Accounted For per group.

	VAF					
	Roll ϕ			Pitch θ		
	p_0, %	p_a, %	$F(\cdot 10^{-2})$	p_0, %	p_a, %	$F(\cdot 10^{-2})$
Group 1	55.3	61.9	6.09	55.2	74.1	1.89
Group 2	53.2	61.4	3.30	51.1	68.2	2.42
Group 3	51.3	58.4	4.68	53.7	68.9	2.67

Figure 7-4 shows the development of the pilot gain K_p throughout training. As was expected from earlier training studies [25,134], the pilot gain increased as training continued. Pilot gains of the three groups were consistently larger in pitch than in roll. This could have been expected, since a larger pilot gain is related to better performance [25,134] and pilots usually perform better in pitch than in roll in dual-axis tracking tasks [150,162–165], as is also seen in the current research. However, some differences between the groups were seen. Whereas the pilot gains in roll and pitch seemed to have stabilized after the first half of training for Group 1, the pilot gains in roll and pitch of Groups 2 and 3 had not yet reached their asymptotic values at the end of training, as can be seen in Figure 7-4 and Table 7-4. Pearson's correlation coefficients of the pilot gain learning curves were also slightly lower than the coefficients of for example the RMS(e) learning curves, meaning that the spread in the average pilot gains was bigger, due to which it was more difficult to describe them by a single learning curve. Table 7-4 does not show the parameters of the pilot gain learning curve in pitch for Group 1 because of the low correlation coefficient.



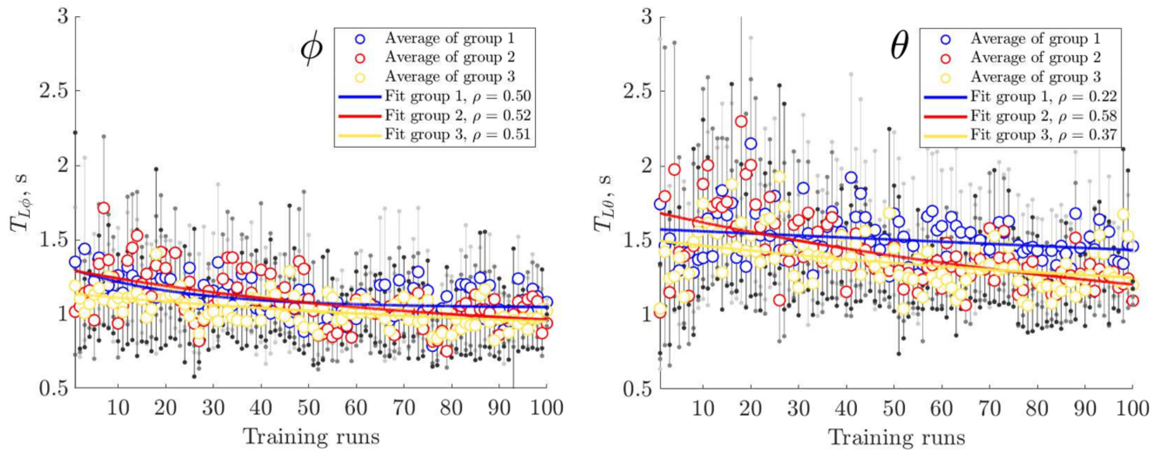
(a) Roll. **(b)** Pitch.

Figure 7-4: Average pilot gain K_p per group.

Table 7-4: Learning curve parameters for pilot gain K_p per group.

	Pilot gain K_p					
	Roll ϕ			Pitch θ		
	p_0, -	p_a, -	$F(\cdot 10^{-2})$	p_0, -	p_a, -	$F(\cdot 10^{-2})$
Group 1	1.28	1.99	8.80	n/a	n/a	n/a
Group 2	1.18	2.20	2.40	1.82	3.28	2.14
Group 3	1.82	2.68	0.88	2.57	4.21	0.85

As expected from earlier training studies [25, 134], a slight decrease in the lead time constant T_L occurred throughout training, as shown in Figure 7-5. The lead time constant was consistently higher in pitch than in roll, as was also observed in previous dual-axis tracking studies [150, 163–165]. A somewhat large spread was observed in the average lead time constants, especially in pitch, which led to slightly lower correlation coefficients. Therefore, no learning curve parameters are provided in Table 7-5 for learning curves with Pearson's correlation coefficient ρ smaller than 0.5.



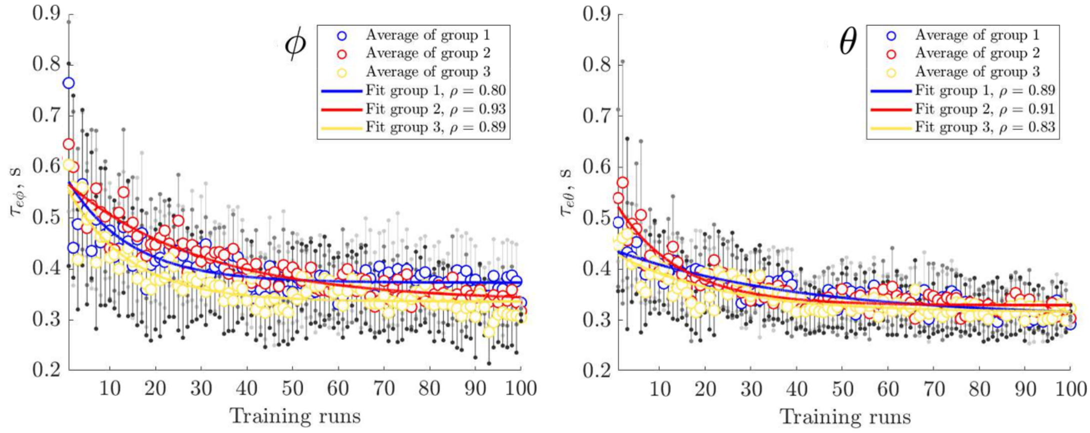
(a) Roll. **(b)** Pitch.

Figure 7-5: Average lead time constant T_L per group.

Table 7-5: Learning curve parameters for lead time constant T_L per group.

	Lead time constant T_L					
	Roll ϕ			Pitch θ		
	p_0, s	p_a, s	$F(\cdot 10^{-2})$	p_0, s	p_a, s	$F(\cdot 10^{-4})$
Group 1	1.31	1.05	4.04	n/a	n/a	n/a
Group 2	1.30	0.85	1.35	1.69	0.85	85.6
Group 3	1.14	0.85	1.02	n/a	n/a	n/a

The development of the response delay τ_e throughout training is shown in Figure 7-6. As was expected from earlier training studies [25, 134], the response delay decreased as training continued. Also, response delays of the three groups were consistently lower in pitch than in roll, as can be seen in Figure 7-6 and Table 7-6. This could have been expected, since a lower response delay is related to better performance [25, 134]. Little spread was shown in the average response delays, resulting in good fits of the learning curves, as was confirmed by the high correlation coefficients.



(a) Roll. **(b)** Pitch.

Figure 7-6: Average response delay τ_e per group.

Table 7-6: Learning curve parameters for response delay τ_e per group.

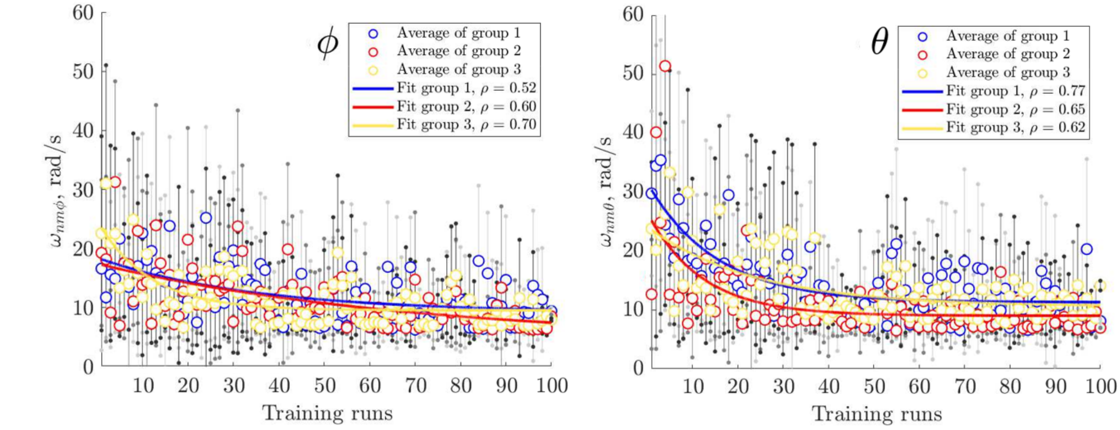
Response delay τ_e						
	Roll ϕ			Pitch θ		
	p_0, s	p_a, s	$F(\cdot 10^{-2})$	p_0, s	p_a, s	$F(\cdot 10^{-2})$
Group 1	0.58	0.37	7.04	0.44	0.30	2.41
Group 2	0.57	0.34	3.28	0.54	0.33	6.96
Group 3	0.57	0.34	7.16	0.43	0.32	4.26

A decrease in the neuromuscular frequency ω_{nm} throughout training is shown in Figure 7-7 and Table 7-7. However, the estimated neuromuscular frequencies early on in training are actually larger than the maximum input frequencies of the target forcing functions in roll and pitch. As already discussed in Section 5-4, these large neuromuscular parameters could be manually adjusted by shifting the neuromuscular frequency to the maximum input frequency of the target forcing function and investigating the influence of this shift on the VAF. If the influence on the VAF is minimal, the smaller neuromuscular frequency can be adopted. However, this method was not yet applied when evaluating the training results in order to form the retention groups due to a lack of time.

Table 7-7: Learning curve parameters for neuromuscular frequency ω_{nm} per group.

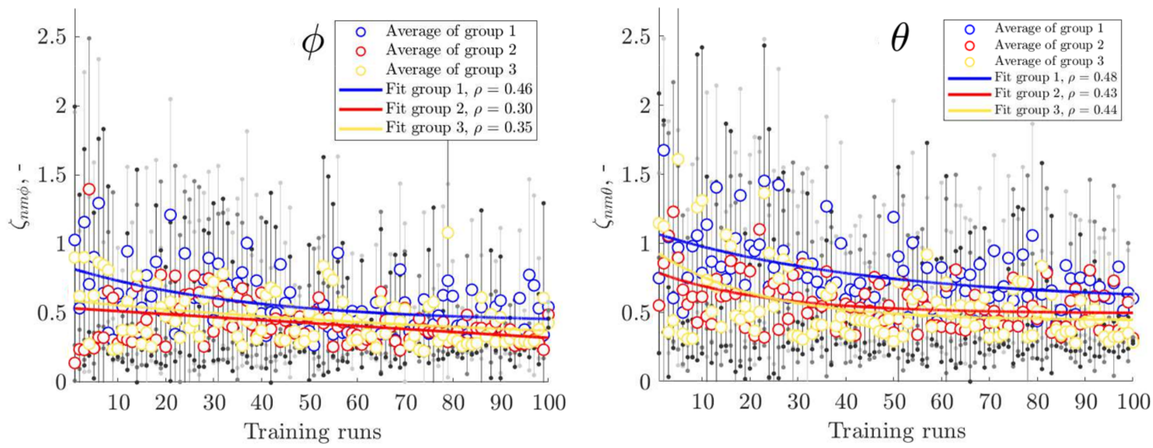
Neuromuscular frequency ω_{nm}						
	Roll ϕ			Pitch θ		
	p_0, rad/s	p_a, rad/s	$F(\cdot 10^{-2})$	p_0, rad/s	p_a, rad/s	$F(\cdot 10^{-2})$
Group 1	18.4	9.17	2.87	31.7	11.4	6.36
Group 2	17.4	4.28	1.41	26.6	9.08	7.96
Group 3	24.8	9.56	8.38	24.2	10.7	4.13

The development of the neuromuscular damping ratio ζ_{nm} throughout training is shown in Figure 7-8. During the first half of training, a considerable number of group means were larger than one. However, manually adjusting the neuromuscular frequency could have a large influence on these damping ratio values. Due to the large spread in the group means of the neuromuscular damping ratio, learning curves could not be accurately fitted to the data, as was expressed through the low correlation coefficients. Therefore, no learning curve parameters are provided.



(a) Roll. **(b)** Pitch.

Figure 7-7: Average neuromuscular frequency ω_{nm} per group.



(a) Roll. **(b)** Pitch.

Figure 7-8: Average neuromuscular damping ratio ζ_{nm} per group.

7-2 End of Training Results

In order to form the retention groups, training performance of each participant was determined by averaging their tracking errors, control inputs, estimated pilot model parameters and VAF of the last ten training runs (runs 91 - 100). The first criteria for the group forming was that there should be no significant differences between the three groups in tracking performance, control activity, estimated pilot model parameters or VAF. This section presents these end-of-training group results.

7-2-1 Tracking Performance

Figure 7-9 shows boxplots of the tracking errors of the three groups at the end of training. Tracking performance was clearly better in pitch than in roll. Although some boxplots were slightly larger than others, a statistical analysis proved that they did not differ significantly from one another, which was the goal during group forming. The statistical analysis results are provided in Appendix I, in Figures I-1 through I-4.

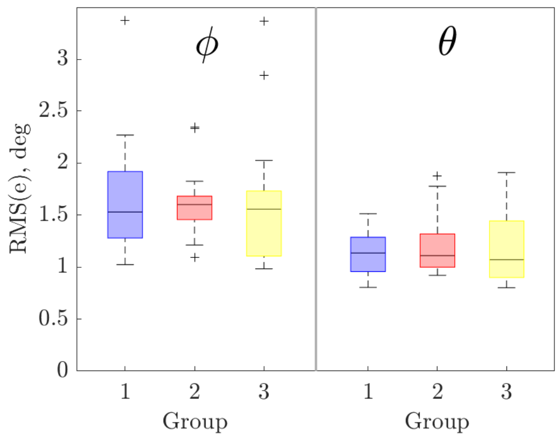


Figure 7-9: Root mean square of tracking error averaged over the last ten training runs (runs 91 - 100) per group.

7-2-2 Control Activity

Figure 7-10 shows boxplots of the control input of the three groups at the end of training. Although some boxplots were slightly larger than others, a statistical analysis proved that they did not differ significantly from one another, which was the goal during group forming. The statistical analysis results are provided in Appendix I, in Figures I-5 through I-8.

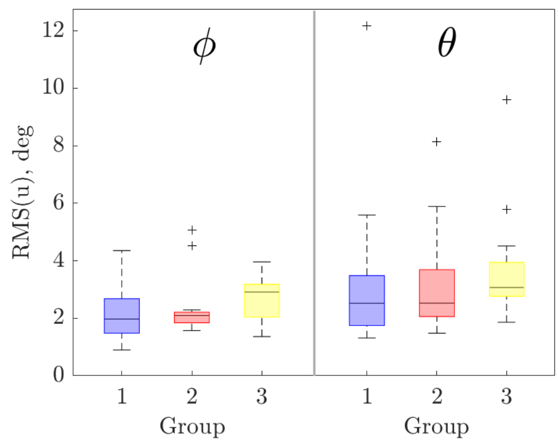
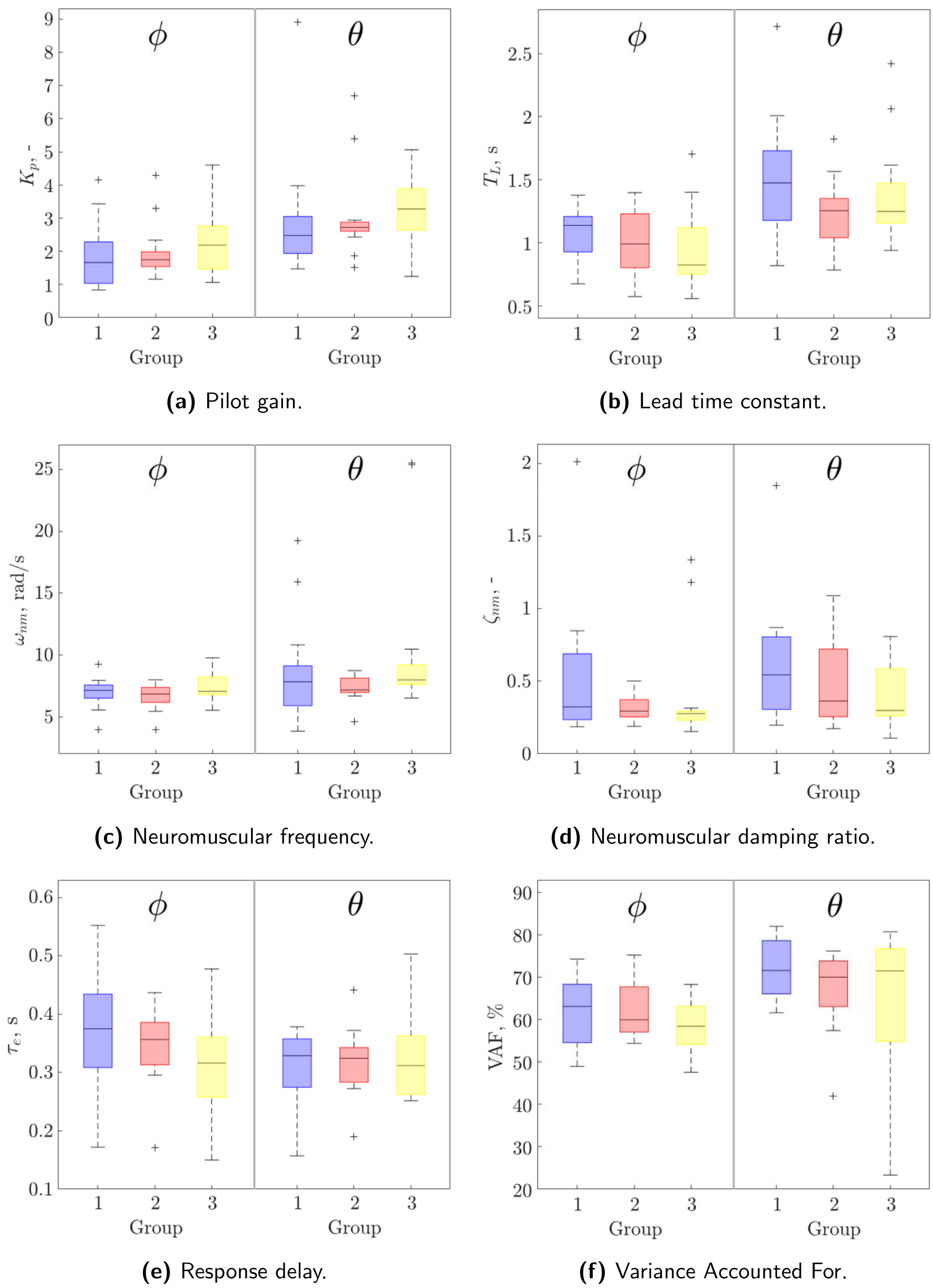


Figure 7-10: Root mean square of control input averaged over the last ten training runs (runs 91 - 100) per group.

7-2-3 Human Operator Modeling Results

Figure 7-11 presents boxplots of the pilot model parameters and VAF of the three groups. Again, Figure 7-11a shows that the pilot gain was slightly larger in pitch than in roll, as was to be expected for higher task proficiency. As was also shown in Figure 7-5, Figure 7-11b depicts a larger lead time constant in pitch than in roll. Additionally, the response delay in Figure 7-11e was slightly smaller in pitch than in roll. Lastly, Figure 7-11f reveals a higher VAF in pitch than in roll. Although it might sometimes be difficult to see with the naked eye, a statistical analysis proved that none of the human operator model parameters nor the VAF differed significantly between the groups, as was the desired result of group forming. The results of the statistical analysis are given in Appendix I, in Figures I-9 through I-36.



(a) Pilot gain. **(b)** Lead time constant.

(c) Neuromuscular frequency. **(d)** Neuromuscular damping ratio.

(e) Response delay. **(f)** Variance Accounted For.

Figure 7-11: Estimated human operator model parameters and Variance Accounted For averaged over the last ten training runs (runs 91 - 100) per group.

7-3 Learning Curve Parameters Tracking Error

In order to form the retention groups, not only participants' performance at the end of training was evaluated, but also their development throughout the entire training phase. This development was analyzed by fitting learning curves to the tracking error data of individual participants. The goal was to ensure that individual learning curve parameters did not differ significantly between the groups. The individual learning curves of participants are provided in Appendix G. In an ideal situation, learning curves would also have been fitted to the control input, pilot model parameters and VAF of each individual participant, to make sure there were also no significant differences between the groups for those dependent measures. However, fitting individual learning curves is a very time-consuming task, since it can take a lot of manual effort to make sure that learning curves are accurately fitted to the measured data. Since there was only a limited amount of time available to analyze the training results and form the retention groups, before the first retention tests had to be performed, there was unfortunately no time available to fit individual learning curves for any of the dependent measures other than the tracking error.

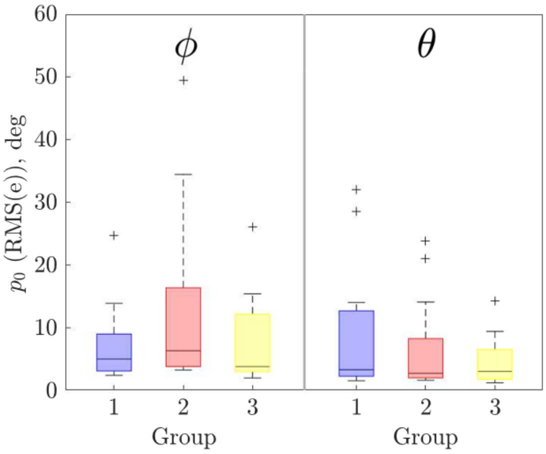
Figure 7-12 shows the differences in learning curve parameters between the groups. Only the parameters of learning curves with $\rho > 0.5$ were taken into account, since only for these learning curves the correlation between the measured human operator data and the learning curve was deemed sufficient. This meant that both in roll and in pitch one individual learning curve was excluded from the data analysis. A statistical analysis showed that the learning curve parameter differences between the three groups were not significant for any of the parameters. The statistical analysis results are provided in Appendix I, in Figures I-37 through I-49.

7-4 Discussion of Training Results

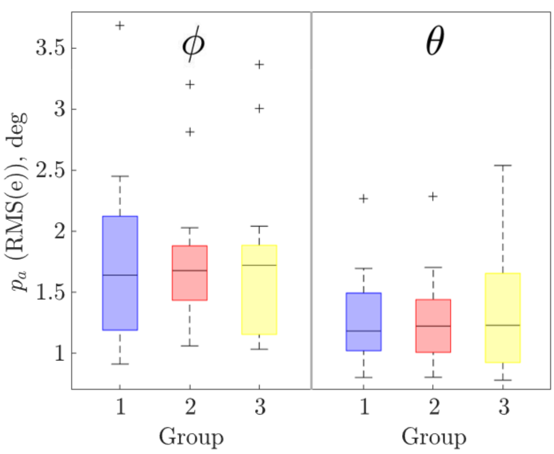
The goal of the training phase was to extensively train task-naive participants on a compensatory dual-axis aircraft roll an pitch tracking task in a fixed-base setting in order to be able to retest them after a period of non-practice on their retention of this control task. The experiment included 39 participants who completed the training phase, during which their tracking behavior was measured. Human operator modeling techniques were applied to quantify control skill improvement during training. The training results were used to divide all participants into three groups. The groups were formed such that at the end of training there were no significant differences in tracking performance, control activity or pilot model parameters between the three groups. This was done to be able to compare retention performance between the groups in a reliable manner. The three groups were retested after different retention intervals in order to evaluate the shape of the skill decay curve as well as to determine the "optimal" retention interval to ensure that manual control skills do not decay significantly, while at the same time minimizing the amount of refresher training.

Based on the results of earlier training experiments [25,134,135], clear effects of training were expected to occur (Hypothesis **H1**). With training, better tracking performance was shown, with a decrease in tracking errors. Statistical analysis confirmed that this decrease was significant for all three experiment groups in both pitch and roll. Contrary to the pilot model parameters identified in the preliminary experiment, a slight increase in pilot gain K_p, a slight decrease in the lead time constant T_L and a relatively large decrease in the response delay τ_e were observed during the training phase of the final experiment. Participants also became more consistent and linear in their control behavior, leading to an increase in the VAF.

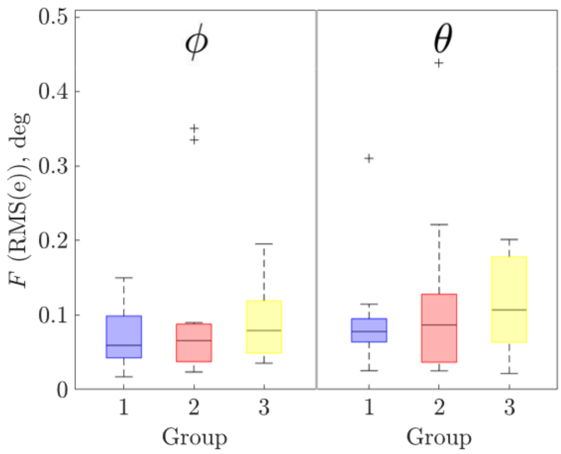
In line with previous dual-axis tracking task experiments [150, 162–165], it was hypothesized that tracking performance would be consistently better in pitch than in roll during training and retention testing (Hypothesis **H2**). During training, it was indeed observed that tracking errors were lower in pitch than in roll. Also, the asymptotic value of the response delay τ_e was slightly lower in pitch than



(a) Initial value p_0.



(b) Asymptotic value p_a.



(c) Learning rate F.

Figure 7-12: Learning curve parameters for root mean square of tracking error per group.

in roll and the pilot gain K_p and the VAF were larger in pitch than in roll. However, no conclusions can be drawn yet regarding the difference in tracking performance between pitch and roll during retention testing.

Since hypotheses 3 through 5 concern retention performance, nothing can be suggested yet about the validity of these hypotheses.

7-5 Next Steps Data Analysis

As discussed in Section 4-3, human control behavior is often modeled as being (quasi-)linear and time-invariant (LTI), whereas in real life manual control behavior is often nonlinear as well as time-varying. Due to the LTI modeling approach, only the overall, lumped response of a fully-trained human based on prolonged measurements can be identified [141,144,147,152,155,158]. However, these lumped responses provide limited insight into the development of manual control skills during training. Although it has not yet been proven, it is likely that some human operator parameters will change faster than others [141], i.e. different parameters have different "life expectancies". To gain a better insight into the temporal scales of learning, time-varying manual control identification and modeling methods should be applied. Examples of time-varying identification methods are those that rely on Kalman filtering [170,176,177], recursive least squares [176,178], wavelets [179] or windowed LTI human operator modeling [179]. Examples of time-varying parameter estimation and model fitting methods are time-domain modeling [164,165] and linear parameter varying model-based methods [157,159,160].

In the current research, the first step to determine whether the human operator's manual control behavior is time-varying is to split up single experiment runs into multiple parts and to analyze the VAF of the pilot model for all these separate parts. If it is found that the operator's manual control behavior is sufficiently time-varying, more time will be invested in analyzing the experiment data using time-varying manual control identification and modeling methods. The exact methods to be used in that case still have to be determined.

Next to investigating the time-varying nature of the operator's manual control behavior, an investigation into the presence of crossfeed will be performed. Crossfeed is a phenomenon which can be described as a form of task interference, in which the human operator is not able to completely decouple two tasks [180], such as a dual-axis tracking task. A first step would be to decompose the root mean square values of the tracking error and control input into contributions from the target signal of the evaluated axis, the target signal of the other axis (crossfeed) and the pilot remnant. If a considerable contribution is made by the crossfeed target signal, crossfeed will be identified for all experiment runs. The identification of crossfeed has been done before in pilot identification studies as well as in short training studies [162, 172, 180–183], but never in full training or skill retention studies. Whereas Van Lunteren [182] found evidence for LTI crossfeed with a visual origin in a multiple loop tracking task in which the loops were controlled by two separate manipulators, one using the left hand and the other using the right hand, Bekey et al. [183] only found crossfeed as a temporary or short-duration phenomenon in a two-axis compensatory tracking task with a single two-axis hand controller. This would suggest that crossfeed is time-varying instead of it being linear and time-invariant. If evidence of crossfeed is found in the current research, a more thorough investigation will need to be carried out to be able to draw any conclusions on the time-varying or time-invariant nature of crossfeed.

When decomposing tracking performance and control activity measures into contributions from the two target signals and the pilot remnant in order to investigate the presence of crossfeed, this data could also be used to investigate the significance of the pilot remnant. If it is shown that the remnant forms a considerable contribution to the total tracking error and input, a more thorough investigation into the remnant characteristics might be performed. Although the remnant is usually neglected, some previous attempts have been made to provide some rationale for the remnant component [184–187].

Lastly, although the crossover frequencies and phase margins were not yet calculated when analyzing the training phase data in order to form the retention groups, this will be done in the final data analysis.

Chapter 8

Summary and Future Research

The aim of this Preliminary Thesis was to determine the framework for the final research goal: performing an objective evaluation of the retention of manual control skills using a cybernetic approach. This newly-gained knowledge will be used to support the development of optimal recurrent training procedures for skill-based manual control in aviation in order to enhance the retention of manual flying skills as well as the development of optimal astronaut crew ground training and onboard refresher training in space flight.

This report included a literature review on the retention of manual control skills, a brief description of the cybernetic approach used to objectively analyze skill development, decay and retention, a preliminary experiment to be able to refine the design of the final experiment, the final experiment design as well as the results of the training phase of the final experiment.

Literature Study An extensive literature review into the retention of manual control skills was performed to gain insight into the variables affecting skill retention, the duration of skill retention, the skill decay curve and the time required for retraining after a retention interval to achieve old performance levels again. Many variables were identified to influence retention, of which the most important ones are the level of original learning, since overlearning is known to enhance retention, the length of the retention interval and the task difficulty [15, 23, 30–35] (**PT sub-question 1**). Previous literature was not conclusive on the trend of the skill decay curve [13, 37, 127, 132] (**PT sub-question 3**). This was due to the fact that in the little amount of previous research concerning skill retention different performance measures were used to measure retention, which could influence the shape of the skill decay curves. It was also found that retraining after a retention interval up to performance levels achieved at the end of initial training occurs rather quickly [36, 79, 93]. However, due to the limited amount of previous research, only a general observation could be made that the required amount of retraining is usually not more than 50% of the initial training time when low amounts of initial training are received [36, 46, 123, 125] and not more than 10% of the initial training time when moderate to high amounts of initial training are performed [124]. As could be expected, the required amount of retraining to achieve end-of-training performance levels usually increases with an increase in length of retention interval [79, 94] and/or in task difficulty [95] (**PT sub-question 4**). As is required to be able to finish the current research within the time allocated for the M.Sc. thesis, it was found that significant skill losses can be observed in a tracking task experiment with a retention interval of only six months while the subjects had reached asymptotic performance at the end of training, as long as the tracking task is sufficiently difficult [46] (**PT sub-question 2**).

Cybernetic Approach The novelty of the current skill retention research lies in the fact that a cybernetic approach will be used to objectively and explicitly quantify skill development, decay and retention. The advantage of this cybernetic approach over more conventional performance measures is that it uses multi-channel pilot models to allow for a quantitative analysis of pilots' use of multiple stimuli/cues for manual control, as well as the development of these control skills during training and after a period of non-practice.

The cybernetic approach models human operators using a linear, time-invariant response function and a remnant, of which the latter is usually neglected. In compensatory tracking tasks in which only visual feedback is present, the linear response function typically consists of equalization dynamics and limitation dynamics. The equalization dynamics could include a pilot gain K_p, a lead time constant T_L and a lag time constant T_I [25]. However, these dynamics are adjusted by the human operator such that the combined pilot-aircraft system approximates single-integrator dynamics for a wide frequency range, including the crossover frequency [145]. Considering the controlled aircraft dynamics used in the current research, the equalization dynamics used to model the human operator will contain a pilot gain K_p and a lead time constant T_L, but no lag time constant T_I. The limitation dynamics are split up into two parts. The first part is a time delay, modeled as an exponential function containing the response delay τ_e, to account for the time delays incurred in the perception and processing of the visual information. The second part are the neuromuscular actuation dynamics, which are modeled as a second-order mass-spring damper system with a neuromuscular frequency ω_{nm} and a neuromuscular damping ratio ζ_{nm} [145]. The human operator model accuracy in describing the pilot's control behavior can be assessed using the VAF. Furthermore, to quantitatively describe how the pilot's control behavior develops during training and after a period of non-practice, exponential learning curves can be fitted to the dependent measures of tracking error, control input, pilot model parameters and VAF [134, 135, 149, 150].

In the current research, Genetic Maximum Likelihood Estimation is used to estimate the parameters of the pilot models applied to the experiment data, at least for an initial data analysis. This one-step time-domain identification method yields more accurate and reliable results than two-step frequency-domain identification techniques [148].

In almost all training studies, clear effects of training can be observed [134, 135, 149, 151]. In tracking tasks, training often becomes evident from improved performance in terms of lower tracking errors [25, 134, 135, 149, 151], an increased pilot gain K_p and a decreased lead time constant T_L and response delay τ_e [25, 134]. In training studies with task-naive participants, also the neuromuscular damping ratio ζ_{nm} often decreases. Additionally, an increase in human operator consistency and linearity is often expected to occur during training [27]. This becomes apparent through an increase in the VAF. Lastly, better performance in skill-based manual control is often accompanied by a higher crossover frequency and phase margin [27].

Current-day cybernetics theory can only accurately model a human operator's control behavior in highly-constrained single-loop compensatory tracking tasks, without any preview on future control constraints, from the moment the operator is done learning. To be able to accurately model the learning human operator, the LTI models must be replaced by time-varying identification techniques, perhaps even by methods that can be employed in real-time. These time-varying identification techniques can be used to model the human operator during the full learning curve, from novice to expert. Identifying time-varying manual control models also provides the opportunity to determine to what extent the universal "time-invariance" assumption of cybernetics is valid.

Preliminary Experiment The objective of the preliminary experiment was twofold. First, the experiment was used to determine how many experiment runs were required for participants to reach asymptotic task performance. The second goal was to determine whether the dual-axis aircraft roll and pitch tracking task provides data suitable for pilot model identification. Two task-naive subjects participated in the preliminary experiment, which was performed in the HMI Laboratory at the Faculty of Aerospace Engineering at the Delft University of Technology. Since no definitive conclusion could be drawn from the measured data of those two participants on whether the tracking task was suitable for pilot model identification, some additional tracking runs were performed by an experienced subject. From this data, it was concluded that the dual-axis tracking task was indeed suitable for pilot model identification (**PE objective 2**). From the preliminary experiment results, the following recommendations were made for the final experiment design. Firstly, it was recommended that the training phase would consist of 100 tracking runs in total, spread out over four consecutive days consisting of 25 tracking runs each (**PE objective 1**). Performing more than 25 tracking runs per day would increase the changes that fatigue or boredom effects would influence the experiment data. A total of 100 tracking runs was recommended to be able to observe the (almost) full learning curve of participants. Furthermore, it was recommended that a short break was scheduled in the middle of each training session to keep concentration levels as constant as possible. Lastly, it was recommended that several forcing function realizations be used in order to avoid that later in training participants start to recognize the target forcing functions.

Experiment Design To objectively and explicitly quantify the retention of manual control skills, a human-in-the-loop experiment was performed in the HMI Laboratory. A total of 39 participants completed the training phase of the experiment. All participants were task-naive prior to starting training and committed to not participating in any other tracking or flying experiments until the last retention test had taken place. All subjects received the same training. They all completed 100 90-second tracking runs spread out over four consecutive working days consisting of 25 tracking runs each. The tracking task they performed was a dual-axis aircraft roll and pitch tracking task, the same as the one used in the preliminary experiment. Five different forcing function realizations were used to ensure that participants would not start to recognize the target signals at some point during training.

After all subjects had completed the training phase, they were divided into three different experiment groups. These groups were formed such that at the end of training there were no significant differences between the three groups in tracking performance, control activity, pilot model parameters or VAF. *Group 1* only performed a single retention test after a retention interval of six months. *Group 2* performed two retention tests with retention intervals of three months in between. *Group 3* performed a total of three retention tests with retention intervals of two months in between. This means that all participants performed their final retention test six months after the end of training. The last retention test of each group was built up in the same manner as the individual training sessions, meaning that the test consisted of 25 90-second tracking runs. The other retention tests, i.e. the first retention test of Group 2 and the first two retention tests of Group 3, consisted of only five 90-second tracking runs. The five-run retention tests were kept short on purpose, to be able to capture the performance of participants at that moment in time, while at the same time avoiding extensive additional learning. All last retention tests were longer again to be able to establish participants' relearning rate of lost skills, if any skill decay had occurred after six months (**Sub-question 3**).

The first retention tests of each group can be used to identify the trend of the skill decay curve (**Sub-question 1**). The last retention tests of each group, meaning the retention tests six months after the end of training, can be used to identify the 'optimal' retention interval to prevent skill decay, while at the same time minimizing the amount of refresher training (**Sub-question 2**).

Training Results As expected from earlier training studies, clear training effects were observed. For all three retention groups, end-of-training tracking performance was significantly better than performance at the start of training. Also, as expected, better performance in pitch than in roll was observed throughout the entire training phase. Regarding human operator modeling, an increase in the VAF throughout training was observed as well as a large decrease in the response delay τ_e. Lastly, a slight increase in the pilot gain K_p and a slight decrease in the lead time constant T_L were seen. No significant differences were observed between the three groups in their end-of-training tracking errors, control input, pilot model parameters or VAF. Also, no significant differences were found between the groups in the parameters of their tracking error learning curves.

Future Steps This Preliminary Thesis presented the results of an extensive literature review concerning the retention of manual control skills, the design and results of a preliminary experiment, as well as the final experiment design and even the results of the experiment training phase. The next phase of the M.Sc. thesis work will mainly focus on retention testing and data analysis. Below, some of the most important steps that still need to be taken in order to finish the current research are stated. These steps are not numbered since they can mostly be performed simultaneously.

- Perform retention testing of all three groups;
- Determine the presence of crossfeed in training data and first retention test data to decide whether crossfeed should be incorporated in the final data analysis;
- Determine whether the human operator's manual control behavior is time-varying by splitting up single experiment runs into multiple parts and analyzing the VAF of the human operator model for these separate parts;
- If the operator's manual control behavior is sufficiently time-varying, decide which time-varying manual control identification and modeling methods will be used to analyze the experiment data;
- Have a look into the contribution of the pilot remnant to the tracking error and control input during training and retention testing;
- Perform final data analysis of training and retention testing data possibly including the identification of time-varying behavior, crossfeed and pilot remnant characteristics;
- Draw final conclusions and recommendations based on the data analysis.

Future Research Suggestions The next research step that would need to be taken after the current research has been completed is to investigate the influence of training device fidelity on the retention of manual control skills. This is considered outside the scope of the current research. Successive research will have to be performed to determine whether the same results are obtained for more challenging flying tasks, such as landing an aircraft. A final research step would be to investigate the retention of manual flying skills in real flight, in order to observe if the outcome of the current research also holds for real flight.

References

[1] Anonymous. (n.d.). HMI Lab. Retrieved September 18, 2018, from `http://cs.lr.tudelft.nl/facilities/hmi-lab/`.

[2] Davenport, W. W. (1978). *Gyro! The Life and Times of Lawrence Sperry.* New York, NY: Scribner.

[3] Sweet, W. (1995). The Glass Cockpit [Flight Deck Automation]. *IEEE Spectrum*, *32*(9):30–38.

[4] Sheridan, T., & Johannsen, G. (1976). *Monitoring Behavior and Supervisory Control.* New York, NY: Plenum Publishing.

[5] Wiener, E. L. (1988). Cockpit Automation. In E. L. Wiener & D. C. Nagel (Eds.), *Academic Press Series in Cognition and Perception. Human Factors in Aviation* (pp. 433-461). San Diego, CA: Academic Press.

[6] Foushee, C. (1990). *National Plan to Enhance Aviation Safety Through Human Factors Improvements.* Washington, DC: Federal Aviation Administration.

[7] Bainbridge, L. (1983). Ironies of Automation. *Automatica*, *19*(6):775–779.

[8] Veillette, P. R. (1995). Differences in Aircrew Manual Skills in Automated and Conventional Flight Decks. *Transportation Research Record*, 1480:43–50.

[9] Nakamura, D. et al. (2013). *Operational Use of Flight Path Management Systems: Final Report of the Performance-Based Operations Aviation Rulemaking Committee/Commercial Aviation Safety Team Flight Deck Automation Working Group.* Washington, DC: Federal Aviation Administration. Retrieved from `https://www.faa.gov/aircraft/air_cert/design_approvals/human_factors/media/OUFPMS_Report.pdf`.

[10] Anonymous. (2017). SAFO 17007: Manual Flight Operations Proficiency. Washington, DC: Federal Aviation Administration.

[11] Anonymous. (2013). SAFO 13002: Manual Flight Operations. Washington, DC: Federal Aviation Administration.

[12] Mengelkoch, R. F., Adams, J. A., & Gainer, C. A. (1960). *The Forgetting of Instrument Flying Skills as a Function of the Level of Initial Proficiency* (NAVTRADEVCEN TR-71-16-18). Port Washington, NY: U.S. Naval Training Center.

[13] Wright, R. H. (1969). *Review of Behavioral Science Research Data Relevant to Army Proficiency Flying Programs,* HumRRO Consulting Report. Fort Rucker, AL: Human Resources Research Organization, Division No. 6 (Aviation).

[14] Wright, R. H. (1973). *Retention of Flying Skills and Refresher Training Requirements: Effects of Non-Flying and Proficiency Flying* (HumRRO Tech. Rep. 73-32). Alexandria, VA: Human Resources Research Organization.

[15] Prophet, W. W. (1976). *Long-Term Retention of Flying Skills: A Review of the Literature* (Report No. FR-ED(P)-76-35). Alexandria, VA: Human Resources Research Organization.

[16] Prophet, W. W. (1976). *Long-Term Retention of Flying Skills: An Annotated Bibliography.* Alexandria, VA: Human Resources Research Organization.

[17] Armstrong, M. B., Bleymaier, J. S., Hinkel, L. F., Levins, R., & Sheppard, R. R. (1975). *Flying Skill Retention and Proficiency Flying* (Research Report No. 0095-75). Montgomery, AL: Air Command and Staff College, Air University, Maxwell Air Force Base.

[18] Smith, J. F. & Matheny, W. G. (1976). *Continuation versus Recurrent Pilot Training* (Report No. AFHRL-TR-76-4). San Antonio, TX: Air Force Human Resources Laboratory, Brooks Air Force Base.

[19] Anonymous. (n.d.). *Road to Wings: Special Report of Accidents in Air Training Command.* San Antonio, TX: Air Training Command, Randolph Air Force Base.

[20] Barshi, I., & Dempsey, D. L. (2016). *Evidence Report: Risk of Performance Errors Due to Training Deficiencies.* Houston, TX: National Aeronautics and Space Administration, Lyndon B. Johnson Space Center.

[21] Chandler, F. (2007). *Learning from NASA mishaps: What separates success from failure.* Report presented to the Project Management Challenge, February 7, 2007. Retrieved from `https://spacese.spacegrant.org/Failure%20Reports/Chandler_Faith_LearningFromFailures.pdf`.

[22] Mars Architecture Steering Group (2009). *Human Exploration of Mars: Design Reference Architecture 5.0* (Report No. NASA/SP2009566). Washington, DC: National Aeronautics and Space Administration Headquarters.

[23] Gardlin, G. R., & Sitterley, T. E. (1972). *Degradation of Learned Skills: A Review and Annotated Bibliography* (Report No. D180-15080-1). Seattle, WA: The Boeing Company.

[24] Anonymous. (2016). Office of Inspector General Audit Report: Enhanced FAA Oversight Could Reduce Hazards Associated With Increased Use of Flight Deck Automation (Report No. AV-2016-013). Washington, DC: United States Department of Transportation.

[25] Pool, D. M., & Zaal, P. M. T. (2016). 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*, *49*(19):343–348.

[26] Rasmussen, J. (1983). Skills, Rules, and Knowledge; Signals, Signs, and Symbols, and Other Distinctions in Human Performance Models. *IEEE Transactions on Systems, Man, and Cybernetics*, *13*(3):257–266.

[27] McRuer, D. T., & Jex, H. R. (1967). A Review of Quasi-Linear Pilot Models. *IEEE Transactions on Human Factors in Electronics*, *8*(3):231–249.

[28] Tustin, A. (1947). The Nature of the Operator's Response in Manual Control, and Its Implications for Controller Design. *Journal of the Institution of Electrical Engineers - Part IIA: Automatic Regulators and Servo Mechanisms*, *94*(2):190–206.

[29] Elkind, J. I., & Forgie, C. D. (1959). Characteristics of the Human Operator in Simple Manual Control Systems. *IRE Transactions on Automatic Control*, *4*(1):44–55.

[30] Schendel, J. D., Shields, J. L., & Katz, M. S. (1978). *Retention of Motor Skills: Review* (Technical Paper 313). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

[31] Naylor, J. C., & Briggs, G. E. (1961). *Long-Term Retention of Learned Skills: A Review of the Literature* (ASD Tech. Rep. 61-390). Canton, OH: Aerospace Medical Laboratory, Wright-Patterson Air Force Base.

[32] Arthur, W., Jr., Bennett, W., Jr., Stanush, P. L., & McNelly, T. L. (1998). Factors That Influence Skill Decay and Retention: A Quantitative Review and Analysis. *Human Performance*, *11*(1):57–101.

[33] Farr, M. J. (1986). *The Long-Term Retention of Knowledge and Skills: A Cognitive and Instructional Perspective* (IDA Memorandum Report M-205). Alexandria, VA: Institute for Defense Analyses.

[34] Hurlock, R. E., & Montague, W. E. (1982). *Skill Retention and Its Implications for Navy Tasks: An Analytical Review* (NPRDC Special Report 82-21). San Diego, CA: Navy Personnel Research and Development Center.

[35] Annett, J. (1977). *Skill Loss: A Review of the Literature and Recommendations for Research* (Report No. AD-A159 187). London, England: Training Services Agency.

[36] Mengelkoch, R. F., Adams, J. A., & Gainer, C. A. (1971). The Forgetting of Instrument Flying Skills. *Human Factors*, *13*(5):397–405.

[37] Sitterley, T. E., & Berge, W. A. (1972). *Degradation of Learned Skills: Effectiveness of Practice Method on Simulated Space Flight Skill Retention* (Report No. D180-15081-1). Seattle, WA: Boeing Aerospace Company.

[38] Adams, J. A., & Hufford, L. E. (1962). Contributions of a Part-Task Trainer to the Learning and Relearning of a Time-Shared Flight Maneuver. *Human Factors*, *4*(3):159–170.

[39] Casner, S. M., Geven, R. W., Recker, M. P., & Schooler, J. W. (2014). The Retention of Manual Flying Skills in the Automated Cockpit. *Human Factors*, *56*(8):1506–1516.

[40] Adams, J. A. (1967). *Human Memory.* New York, NY: McGraw-Hill.

[41] Gentile, A. M., & Nacson, J. (1976). Organizational Processes in Motor Control. *Exercise and Sport Sciences Reviews*, *4*(1):1–33.

[42] Noble, M., Trumbo, D., Ulrich, L., & Cross, K. (1966). Task Predictability and the Development of Tracking Skill under Extended Practice. *Journal of Experimental Psychology*, *72*(1):85–94.

[43] Swink, J., Trumbo, D., & Noble, M. (1967). On the Length-Difficulty Relation in Skill Performance. *Journal of Experimental Psychology*, *74*(3):356–362.

[44] Trumbo, D., Noble, M., Cross, K., & Ulrich, L. (1965). Task Predictability in the Organization, Acquisition, and Retention of Tracking Skills. *Journal of Experimental Psychology*, *70*(3):252–263.

[45] Trumbo, D., Noble, M., & Swink, J. (1967). Secondary Task Interference in the Performance of Tracking Tasks. *Journal of Experimental Psychology*, *73*(2):232–240.

[46] Hammerton, M. (1963). Retention of Learning in a Difficult Tracking Task. *Journal of Experimental Psychology*, *66*(1):108–110.

[47] Fitts, P. M., & Seeger, C. M. (1953). S-R Compatibility: Spatial Characteristics of Stimulus and Response Codes. *Journal of Experimental Psychology*, *46*(3):199–210.

[48] Adams, J. A. (1954). Psychomotor Response Acquisition and Transfer as a Function of Control-Indicator Relationships. *Journal of Experimental Psychology*, *48*(1):10–14.

[49] Briggs, G. E. (1969). Transfer of Training. In E. A. Bilodeau (Ed.), *Principles of Skill Acquisition* (pp. 205-234). New York, NY: Academic Press.

[50] Gagné, R. M., Baker, K. E., & Foster, H. (1950). On the Relation between Similarity and Transfer of Training in the Learning of Discriminative Motor Tasks. *Psychological Review*, *57*(2):67–79.

[51] Lewis, D., & Shephard, A. H. (1951). Facilitation and Interference in Performance on the Modified Mashburn Apparatus: I. The Effects of Varying the Amount of Original Learning. *Journal of Experimental Psychology*, *41*(4):247–260.

[52] Melton, A. W. (1964). *Retention of Tracking Skills* (Report No. DA-49-007-MD-1020). Ann Arbor, MI: University of Michigan.

[53] Bahrick, H. P., Noble, M., and Fitts, P. M. (1954). Extra-Task Performance as a Measure of Learning a Primary Task. *Journal of Experimental Psychology*, *48*(4):298–302.

[54] Fleishman, E. A., & Rich, S. (1963). Role of Kinesthetic and Spatial-Visual Abilities in Perceptual-Motor Learning. *Journal of Experimental Psychology*, *66*(1):6–11.

[55] Grimsley, D. L. (1969). *Acquisition, Retention, and Retraining: Effects of High and Low Fidelity in Training Devices* (HumRRO Tech. Rep. 69-1). Alexandria, VA: Human Resources Research Office.

[56] Bernstein, B. R., & Gonzalez, B. K. (1969). *Learning, Retention, and Transfer, Volume I of II* (NAVTRADEVCEN Tech. Rep. 68-C-0215-1). Orlando, FL: U.S. Naval Training Devices Center.

[57] Johnson, S. L. (1981). Effect of Training Device on Retention and Transfer of a Procedural Task. *Human Factors*, *23*(3):257–272.

[58] Purdy, B. J., & Lockhart, A. (1971). Retention and Relearning of Gross Motor Skills after Long Periods of No Practice. *Research Quarterly*, *42*(1):265–272.

[59] Vineberg, R. (1975). *A Study of the Retention of Skills and Knowledge Acquired in Basic Training* (HumRRO Tech. Rep. 75-10). Alexandria, VA: Human Resources Research Organization.

[60] Grimsley, D. L. (1969). *Acquisition, Retention, and Retraining: Training Category IV Personnel With Low Fidelity Devices* (HumRRO Tech. Rep. 69-12). Alexandria, VA: Human Resources Research Organization.

[61] Healy, A. F., & Bourne, L. E., Jr. (2012). *Training Cognition: Optimizing Efficiency, Durability, and Generalizability.* New York, NY: Psychology Press.

[62] Lohse, K. R., Sherwood, D. E., & Healy, A. F. (2010). How Changing the Focus of Attention Affects Performance, Kinematics, and Electromyography in Dart Throwing. *Human Movement Science*, *29*(4):542–555.

[63] Lohse, K. R., Wulf, G., & Lewthwaite, R. (2012). Attentional Focus Affects Movement Efficiency. In N. J. Hodges & M. A. Williams (Eds.), *Skill Acquisition in Sport: Research, Theory and Practice* (2nd ed.). New York, NY: Routledge.

[64] Shea, C. H., & Wulf, G. (1999). Enhancing Motor Learning Through External-Focus Instructions and Feedback. *Human Movement Science*, *18*(4):553–571.

[65] Wulf, G. (2007). *Attention and Motor Skill Learning.* Champaign, IL: Human Kinetics.

[66] Beilock, S. L., Bertenthal, B. I., McCoy, A. M., & Carr, T. H. (2004). Haste Does Not Always Make Waste: Expertise, Direction of Attention, and Speed versus Accuracy in Performing Sensorimotor Skills. *Psychonomic Bulletin & Review*, *11*(2):373–379.

[67] Schmidt, R. A. (1975). *Motor Skills.* New York, NY: Harper & Row.

[68] Duncan, C. P., & Underwood, B. J. (1953). Retention of Transfer in Motor Learning after Twenty-Four Hours and after Fourteen Months. *Journal of Experimental Psychology*, *46*(6):445–452.

[69] McAllister, D. E. (1952). Retroactive Facilitation and Interference as a Function of Level of Learning. *The American Journal of Psychology*, 65(2):218–232.

[70] Holding, D. H. (1965). *Principles of Training.* Oxford, England: Pergamon Press.

[71] Holding, D. H. (1976). An Approximate Transfer Surface. *Journal of Motor Behavior*, *8*(1):1–9.

[72] Holding, D. H. (1977). Transfer of Training. In B. B. Wolman (Ed.), *International Encyclopedia of Psychiatry, Psychology, Psychoanalysis, and Neurology.* New York, NY: Aesculapius.

[73] Duncan, C. P. (1953). Transfer in Motor Learning as a Function of Degree of First-Task Learning and Inter-Task Similarity. *Journal of Experimental Psychology*, *45*:1–11.

[74] Mandler, G. (1954). Transfer of Training as a Function of Degree of Response Overlearning. *Journal of Experimental Psychology*, *47*(6):411–417.

[75] Wheaton, G. R., Rose, A. M., Fingerman, P. W., Korotkin, A. L., & Holding D. H. (1976). *Evaluation of the Effectiveness of Training Devices: Literature Review and Preliminary Model* (Research Memorandum 76-6). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

[76] Baker, K. E., Wylie, R. C., & Gagné, R. M. (1950). Transfer of Training to a Motor Skill as a Function of Variation in Rate of Response. *Journal of Experimental Psychology*, *40*(6):721–732.

[77] Lordahl, D. S., & Archer, E. J. (1958). Transfer Effects on a Rotary Pursuit Task as a Function of First-Task Difficulty. *Journal of Experimental Psychology*, *56*(5):421–426.

[78] Fleishman, E. A., & Parker, J. F., Jr. (1962). Factors in the Retention and Relearning of Perceptual-Motor Skill. *Journal of Experimental Psychology*, *64*(3):215–226.

[79] Ammons, R. B., Farr, R. G., Bloch, E., Neumann, E., Dey, M., Marion, R., & Ammons, C. H. (1958). Long-Term Retention of Perceptual-Motor Skills. *Journal of Experimental Psychology*, *55*(4):318–328.

[80] Melnick, M. J. (1971). Effects of Overlearning on the Retention of a Gross Motor Skill. *Research Quarterly*, *42*(1):60–69.

[81] Boucher, J.-L. (1974). Higher Processes in Motor Learning. *Journal of Motor Behavior*, *6*(3):131–137.

[82] Martens, R. (1974). Arousal and Motor Performance. *Exercise and Sport Sciences Reviews*, *2*(1):155–188.

[83] Bilodeau, E. A., Bilodeau, I. M., & Schumsky, D. A. (1959). Some Effects of Introducing and Withdrawing Knowledge of Results Early and Late in Practice. *Journal of Experimental Psychology*, *58*(2):142–144.

[84] Newell, K. M. (1974). Knowledge of Results and Motor Learning. *Journal of Motor Behavior*, *6*(4):235–244.

[85] Schmidt, R. A., & White, J. L. (1972). Evidence for an Error Detection Mechanism in Motor Skills. *Journal of Motor Behavior*, *4*(3):143–153.

[86] Schmidt, R. A., & Bjork, R. A. (1992). New Conceptualizations of Practice: Common Principles in Three Paradigms Suggest New Concepts for Training. *Psychological Science*, *3*(4):207–217.

[87] Rogers, C. A. (1974). Feedback Precision and Postfeedback Interval Duration. *Journal of Experimental Psychology*, *102*(4):604–608.

[88] Schmidt, R. A., Young, D. E., Swinnen, S., & Shapiro, D. C. (1989). Summary Knowledge of Results for Skill Acquisition: Support for the Guidance Hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(2):352–359.

[89] Adams, J. A., Goetz, E. T., & Marshall, P. H. (1972). Response Feedback and Motor Learning. *Journal of Experimental Psychology*, *92*(3):391–397.

[90] Adams, J. A., & Goetz, E. T. (1973). Feedback and Practice as Variables in Error Detection and Correction. *Journal of Motor Behavior*, *5*(4):217–224.

[91] Adams, J. A., Gopher, D., & Lintern, G. (1977). Effects of Visual and Proprioceptive Feedback on Motor Learning. *Journal of Motor Behavior*, *9*(1):11–22.

[92] Adams, J. A., Marshall, P. H., & Goetz, E. T. (1972). Response Feedback and Short-Term Motor Retention. *Journal of Experimental Psychology*, *92*(1):92–95.

[93] Hill, D. S. (1914). Minor Studies in Learning and Relearning. *Journal of Educational Psychology*, *5*(7):375–386.

[94] Neumann, E., & Ammons, R. B. (1957). Acquisition and Long-Term Retention of a Simple Serial Perceptual-Motor Skill. *Journal of Experimental Psychology*, *53*(3):159–161.

[95] Lersten, K. C. (1969). Retention of Skill on the Rho Apparatus after One Year. *Research Quarterly*, *40*(2):418–419.

[96] Leonard, R. L., Jr., Wheaton, G. R., & Cohen, F. D. (1976). *Transfer of Training and Skills Retention* (Technical Report TR-76-A3). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

[97] Sitterley, T. E. (1974). *Degradation of Learned Skills: Static Practice Effectiveness for Visual Approach and Landing Skill Retention* (Report No. D180-17876-1). Seattle, WA: Boeing Aerospace Company.

[98] Sitterley, T. E., Zaitzeff, L. P., & Berge, W. A. (1972). *Degradation of Learned Skills: Effectiveness of Practice Method on Visual Approach and Landing Skill Retention* (Report No. D180-15082-1). Seattle, WA: Boeing Aerospace Company.

[99] Cotterman, T. E., & Wood, M. E. (1967). *Retention of Simulated Lunar Landing Mission Skills: A Test of Pilot Reliability* (Report No. AMRL-TR-66-222). Dayton, OH: Aerospace Medical Research Laboratories, Wright-Patterson Air Force Base.

[100] Ammons, R. B. (1951). Effects of Distribution of Practice on Rotary Pursuit "Hits". *Journal of Experimental Psychology*, *41*(1):17–22.

[101] Singer, R. N. (1975). *Motor Learning and Human Performance: Application in Physical Education Skills*. London, England: Macmillan Publishers.

[102] Adams, J. A., & Reynolds, B. (1954). Effect of Shift in Distribution of Practice Conditions Following Interpolated Rest. *Journal of Experimental Psychology*, *47*(1):32–36.

[103] Lewis, D., & Lowe, W. F. (1956). Retention of Skill on the SAM Complex Coordinator. *Proceedings of the Iowa Academy of Science, Cedar Falls (IA)*, *63*:591–599.

[104] Reynolds, B., & Bilodeau, I. M. (1952). Acquisition and Retention of Three Psychomotor Tests as a Function of Distribution of Practice during Acquisition. *Journal of Experimental Psychology*, *44*(1):19–26.

[105] Catalano, J. F. (1978). The Effect of Rest Following Massed Practice of Continuous and Discrete Motor Tasks. *Journal of Motor Behavior*, *10*(1):63–67.

[106] Bilodeau, E. A., & Bilodeau, I. M. (1961). Motor Skills-Learning. *Annual Review of Psychology*, *12*:243–280.

[107] Goldstein, I. L. (1974). *Training: Program Development and Evaluation*. Monterey, CA: Brooks/Cole Publishing.

[108] Naylor, J. C. (1962). *Parameters Affecting the Relative Efficiency of Part and Whole Practice Methods: A Review of the Literature* (NAVTRADEVCEN Tech. Rep. 950-1). Orlando, FL: U.S. Naval Training Devices Center.

[109] Naylor, J. C., & Briggs, G. E. (1963). Effects of Task Complexity and Task Organization on the Relative Efficiency of Part and Whole Training Methods. *Journal of Experimental Psychology*, *65*(3):217–224.

[110] Roche, R. A. P., Commins, S., Agnew, F., Cassidy, S., Corapi, K., Leibbrand, S., et al. (2007). Concurrent Task Performance Enhances Low-Level Visuomotor Learning. *Perception & Psychophysics*, *69*(4):513–522.

[111] Hemond, C., Brown, R. M., & Robertson, E. M. (2010). A Distraction Can Impair or Enhance Motor Performance. *Journal of Neuroscience*, *30*(2):650–654.

[112] Goh, H. T., Sullivan, K. J., Gordon, J., Wulf, G., & Winstein, C. J. (2012). Dual-Task Practice Enhances Motor Learning: A Preliminary Investigation. *Experimental Brain Research*, *222*(3):201–210.

[113] Naylor, J. C., Briggs, G. E., & Reed, W. G. (1968). Task Coherence, Training Time, and Retention Interval Effects on Skill Retention. *Journal of Applied Psychology*, *52*(5):386–393.

[114] Bilodeau, E. A. (1969). Supplementary Feedback and Instructions. In E. A. Bilodeau (Ed.), *Principles of Skill Aqcuisition*. New York, NY: Academic Press.

[115] Reynolds, B., & Adams, J. A. (1953). Motor Performance as a Function of Click Reinforcement. *Journal of Experimental Psychology*, *45*(5):315–320.

[116] Kinkade, R. G. (1963). *A Differential Influence of Augmented Feedback on Learning and on Performance* (Report No. AMRL-TDR-63-12). Dayton, OH: Aerospace Medical Research Laboratories, Wright-Patterson Air Force Base.

[117] Bilodeau, I. M. (1969). Information Feedback. In E. A. Bilodeau (Ed.), *Principles of Skill Acquisition*. New York, NY: Academic Press.

[118] Buckhout, R., Naylor, J. C., & Briggs, G. E. (1963). *Effects of Modified Task Feedback during Training on Performance of a Simulated Attitude Control Task after Thirty Days* (Report No. AMRL-TDR-63-125). Dayton, OH: Aerospace Medical Research Laboratories, Wright-Patterson Air Force Base.

[119] Shanks, D. R., & Cameron, A. (2000). The Effect of Mental Practice on Performance in a Sequential Reaction Time Task. *Journal of Motor Behavior*, *32*(3):305–313.

[120] Driskell, J. E., Copper, C., & Moran A. (1994). Does Mental Practice Enhance Performance? *Journal of Applied Psychology*, *79*(4):481–492.

[121] Minas, S. C. (1978). Mental Practice of a Complex Perceptual Motor Skill. *Journal of Human Movement Studies*, *4*(2):102–107.

[122] Kohl, R. M., & Roenker, D. L. (1983). Mechanism Involvement during Skill Imagery. *Journal of Motor Behavior*, *15*(2):179–190.

[123] Bell, H. M. (1950). Retention of Pursuit Rotor Skill after One Year. *Journal of Experimental Psychology*, *40*(5):648–649.

[124] Eysenck, S. B. G. (1960). Retention of a Well-Developed Motor Skill after One Year. *The Journal of General Psychology*, *63*(2):267–273.

[125] Ammons, D. (1988). *Long-Term Retention of a Simple Motor Skill*, unpublished Masters Thesis. Missoula, MT: University of Montana.

[126] Battig, W. F., Nagel, E. H., Voss, J. F., & Brogden, W. J. (1957). Transfer and Retention of Bidimensional Compensatory Tracking after Extended Practice. *The American Journal of Psychology*, *70*(1):75–80.

[127] Youngling, E. W., Sharpe, E. N., Ricketson, B. S., & McGee, D. W. (1968). *Crew Skill Retention for Space Missions up to 200 Days* (Report No. F766). Berkeley, MO: McDonnell-Douglas Astronautics Company.

[128] Wilson, W. B. (1973). *The Effect of Prolonged Non-Flying Periods on Pilot Skill in Performance of a Simulated Carrier Landing Task*, unpublished Masters Thesis. Monterey, CA: Naval Postgraduate School.

[129] Smode, A. F., Hall, E. R., & Meyer, D. E. (1966). *An Assessment of Research Relevant to Pilot Training* (Report No. AMRL-TR-66-196). Dayton, OH: Aerospace Medical Research Laboratories, Wright-Patterson Air Force Base.

[130] Arthur, W., Jr., Day, E. A., Bennett, W., Jr., & Portrey, A. (2013). *Individual and Team Skill Decay: The Science and Implications for Practice*. New York, NY: Routledge.

[131] Ebbinghaus, H. (1885). *Memory: A Contribution to Experimental Psychology* (H. A. Ruger, & C. E. Bussenius, Trans.). New York, NY: Teachers College, Columbia University.

[132] Ruffner, J., Wick, W., & Bickley, W. (1984). 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)*, *28*(4):370–374.

[133] Bahrick, H. P. (1964). Retention Curves: Facts or Artifacts? *Psychological Bulletin*, *61*(3):188–194.

[134] Pool, D. M., Harder, G. A., & Van Paassen, M. M. (2016). Effects of Simulator Motion Feedback on Training of Skill-Based Control Behavior. *Journal of Guidance, Control, and Dynamics*, *39*(4):889–901.

[135] Mendes, M. F. S., Pool, D. M., & Van Paassen, M. M. (2017). Effects of Peripheral Visual Cues in Simulator-Based Training of Multimodal Control Skills. *Proceedings of AIAA Modeling and Simulation Technologies Conference, Denver (CO)*, no. AIAA-2017-3671.

[136] Wiener, N. (1961). *Cybernetics: Or Control and Communication in the Animal and the Machine* (2nd ed.). Cambridge, MA: MIT Press.

[137] McRuer, D. T., & Weir, D. H. (1969). Theory of Manual Vehicular Control. *IEEE Transactions on Man-Machine Systems*, *10*(4):257–291.

[138] McRuer, D. T., & Krendel, E. S. (1974). *Mathematical Models of Human Pilot Behavior* (AGARDograph AGARD-AG-188). Paris, France: Advisory Group for Aerospace Research and Development.

[139] Jagacinski, R. J., & Flach, J. M. (2003). *Control Theory for Humans: Quantitative Approaches to Modeling Performance*. Mahwah, NJ: Lawrence Erlbaum Associates.

[140] Mulder, M., Van Paassen, M. M., & Boer, E. R. (2004). Exploring the Roles of Information in the Manual Control of Vehicular Locomotion: From Kinematics and Dynamics to Cybernetics. *Presence: Teleoperators and Virtual Environments*, *13*(5):535–548.

[141] Mulder, M., Pool, D. M., Abbink, D. A., Boer, E. R., Zaal, P. M. T., Drop, F. M., Van der El, K., & Van Paassen, M. M. (2018). Manual Control Cybernetics: State-of-the-Art and Current Trends. *IEEE Transactions on Human-Machine Systems*, *48*(5):468–485.

[142] McRuer, D. T., & Krendel, E. S. (1959). The Human Operator as a Servo System Element. *Journal of the Franklin Institute*, *267*(5):381–403.

[143] Krendel, E. S., & McRuer, D. T. (1960). A Servomechanics Approach to Skill Development. *Journal of the Franklin Institute*, *269*(1):24–42.

[144] McRuer, D., Graham, D., Krendel, E., & Reisener, W., Jr. (1965). *Human Pilot Dynamics in Compensatory Systems: Theory, Models, and Experiments with Controlled Element and Forcing Function Variations* (Technical Report No. AFFDL-TR-65-15). Dayton, OH: Air Force Flight Dynamics Laboratory, Wright-Patterson Air Force Base.

[145] Pool, D. M., Zaal, P. M. T., Damveld, H. J., Van Paassen, M. M., Van der Vaart, J. C., & Mulder, M. (2011). Modeling Wide-Frequency-Range Pilot Equalization for Control of Aircraft Pitch Dynamics. *Journal of Guidance, Control, and Dynamics*, *34*(5):1529–1542.

[146] Van Paassen, M. M., & Mulder, M. (1998). 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*, 515-520.

[147] Nieuwenhuizen, F. M., Zaal, P. M. T., Mulder, M., Van Paassen, M. M., & Mulder, J. A. (2008). Modeling Human Multi-Channel Perception and Control Using Linear Time-Invariant Models. *Journal of Guidance, Control, and Dynamics*, *32*(2):366–377.

[148] Zaal, P. M. T., Pool, D. M., Chu, Q. P., Van Paassen, M. M., Mulder, M., & Mulder, J. A. (2009). Modeling Human Multimodal Perception and Control Using Genetic Maximum Likelihood Estimation. *Journal of Guidance, Control, and Dynamics*, *32*(4):1089–1099.

[149] Levison, W. H., Lancraft, R. E., & Junker, A. M. (1979). Effects of Simulator Delays on Performance and Learning in a Roll-Axis Tracking Task. *Proceedings of the 15th Annual Conference on Manual Control, Dayton (OH)*, 168-186.

[150] Zaal, P. M. T., & Mobertz, X. R. I. (2017). Effects of Motion Cues on the Training of Multi-Axis Manual Control Skills. *Proceedings of AIAA Modeling and Simulation Technologies Conference, Denver (CO)*, no. AIAA-2017-3473.

[151] Martin, E. A. (2008). *The Influence of Tactual Seat-Motion Cues on Training and Performance in a Roll-Axis Compensatory Tracking Task Setting* (Report No. AFRL-RH-WP-SR-2009-0002). Dayton, OH: Air Force Research Laboratory Cognitive Systems Branch, Wright-Patterson Air Force Base.

[152] Zaal, P. M. T., Pool, D. M., De Bruin, J., Mulder, M., & Van Paassen, M. M. (2009). Use of Pitch and Heave Motion Cues in a Pitch Control Task. *Journal of Guidance, Control, and Dynamics*, *32*(2):366–377.

[153] Pool, D. M., Zaal, P. M. T., Van Paassen, M. M., & Mulder, M. (2010). Effects of Heave Washout Settings in Aircraft Pitch Disturbance Rejection. *Journal of Guidance, Control, and Dynamics*, *33*(1):29–41.

[154] De Vries, R. J. (2016). *A Tracking Task for Quantifying Loss of Motor Skills due to Parkinson's Disease*, unpublished M.Sc. Thesis. Delft, The Netherlands: Faculty of Aerospace Engineering, Delft University of Technology.

[155] Mulder, M., Pool, D. M., Abbink, D. A., Boer, E. R., & Van Paassen, M. M. (2016). 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*, *49*(19):1–6.

[156] Franklin, D. W., Osu, R., Burdet, E., Kawato, M., & Milner, T. E. (2003). Adaptation to Stable and Unstable Dynamics Achieved By Combined Impedance Control and Inverse Dynamics Model. *Journal of Neurophysiology*, *90*(5):3270–3282.

[157] Pronker, A., Abbink, D. A., Van Paassen, M. M., & Mulder, M. (2017). Estimating Driver Time-Varying Neuromuscular Admittance Through LPV Model and Grip Force. *Proceedings of the 20th IFAC World Congress, Toulouse, France*, 15481-15486.

[158] Nieuwenhuizen, F. M., Mulder, M., Van Paassen, M. M., & Bülthoff, H. H. (2013). Influences of Simulator Motion System Characteristics on Pilot Control Behavior. *Journal of Guidance, Control, and Dynamics*, *36*(3):667–676.

[159] Tóth, R., Laurain, V., Gilson, M., & Garnier, H. (2012). Instrumental Variable Scheme for Closed-Loop LPV Model Identification. *Automatica*, *48*:2314–2320.

[160] Van Wingerden, J.-W., & Verhaegen, M. (2009). Subspace Identification of Bilinear and LPV Systems for Open- and Closed-Loop Data. *Automatica*, *45*(2):372–381.

[161] Stapleford, R. L., McRuer, D. T., & Magdaleno, R. E. (1967). Pilot Describing Function Measurements in a Multiloop Task. *IEEE Transactions on Human Factors in Electronics*, *4*(1):52–55.

[162] Barendswaard, S., Pool, D. M., & Mulder, M. (2016). 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*, *49*(19):189–194.

[163] Zaal, P. M. T., & Sweet, B. T. (2012). Identification of Time-Varying Pilot Control Behavior in Multi-Axis Control Tasks. *Proceedings of AIAA Modeling and Simulation Technologies Conference, Minneapolis (MN)*, no. AIAA-2012-4793.

[164] Zaal, P. M. T. (2016). Manual Control Adaptation to Changing Vehicle Dynamics in Roll-Pitch Control Tasks. *Journal of Guidance, Control, and Dynamics*, *39*(5):1046–1058.

[165] Zaal, P. M. T., & Pool, D. M. (2014). Multimodal Pilot Behavior in Multi-Axis Tracking Tasks with Time-Varying Motion Cueing Gains. *Proceedings of AIAA Modeling and Simulation Technologies Conference, National Harbor (MD)*, no. AIAA-2014-0810.

[166] Damveld, H. J. (2009). *A Cybernetic Approach to Assess the Longitudinal Handling Qualities of Aeroelastic Aircraft*, Ph.D. Thesis. Delft, The Netherlands: Faculty of Aerospace Engineering, Delft University of Technology.

[167] Van Paassen, M. M., & Mulder, M. (2006). Identification of Human Control Behavior. In W. Karwowski (Ed.), *International Encyclopedia of Ergonomics and Human Factors* (2nd ed.) (pp. 400-407). London, England: Taylor & Francis.

[168] Pool, D. M., Mulder, M., Van Paassen, M. M., & Van der Vaart, J. C. (2008). Effects of Peripheral Visual and Physical Motion Cues in Roll-Axis Tracking Tasks. *Journal of Guidance, Control, and Dynamics*, *31*(6):1608–1622.

[169] Beerens, G. C., Damveld, H. J., Mulder, M., Van Paassen, M. M., & Van der Vaart, J. C. (2009). Investigation into Crossover Regression in Compensatory Manual Tracking Tasks. *Journal of Guidance, Control, and Dynamics*, *32*(5):1429–1445.

[170] Popovici, A., Zaal, P. M. T., & Pool, D. M. (2017). Dual Extended Kalman Filter for the Identification of Time-Varying Human Manual Control Behavior. *Proceedings of the AIAA Modeling and Simulation Technologies Conference, Denver (CO)*, no. AIAA-2017-3666.

[171] Pool, D. M., Zaal, P. M. T., Damveld, H. J., Van Paassen, M. M., & Mulder, M. (2012). Evaluating Simulator Motion Fidelity using In-Flight and Simulator Measurements of Roll Tracking Behavior. *Proceedings of AIAA Modeling and Simulation Technologies Conference, Minneapolis (MN)*, no. AIAA-2012-4635.

[172] Barendswaard, S. (2016). *Investigating the Difference Between Single and Dual Axis Manual Control*, unpublished M.Sc. Thesis. Delft, The Netherlands: Faculty of Aerospace Engineering, Delft University of Technology.

[173] Damveld, H. J., Beerens, G. C., Van Paassen, M. M., & Mulder, M. (2010). Design of Forcing Functions for the Identification of Human Control Behavior. *Journal of Guidance, Control, and Dynamics*, *33*(4):1064–1081.

[174] Robertson, E. M., Pascual-Leone, A., & Press, D. Z. (2014). Awareness Modifies the Skill-Learning Benefits of Sleep. *Current Biology*, *14*(3):208–212.

[175] Kantak, S. S., & Winstein, C. J. (2012). Learning-Performance Distinction and Memory Processes for Motor Skills: A Focused Review and Perspective. *Behavioural Brain Research*, *228*(1):219–231.

[176] Boer, E. R., & Kenyon, R. V. (1998). Estimation of Time-Varying Delay Time in Nonstationary Linear Systems: An Approach to Monitor Human Operator Adaptation in Manual Tracking Tasks. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, *28*(1):89–99.

[177] Schiess, J. R., & Roland, V. R. (1975). *Kalman Filter Estimation of Human Pilot-Model Parameters* (Tech. Rep. NASA-TN-D-8024). Hampton, VA: NASA Langley Research Center.

[178] Olivari, M., Nieuwenhuizen, F. M., Bülthoff, H. H., & Pollini, L. (2014). Identifying Time-Varying Neuromuscular System with a Recursive Least-Squares Algorithm: A Monte-Carlo Simulation Study. *Proceedings of the 2014 IEEE International Conference on Systems, Man, and Cybernetics, San Diego (CA)*, 3573-3578.

[179] Zaal, P. M. T., & Sweet, B. T. (2011). Estimation of Time-Varying Pilot Model Parameters. *Proceedings of AIAA Modeling and Simulation Technologies Conference, Portland (OR)*, no. AIAA-2011-6474.

[180] Todosiev, E. P. (1967). Human Performance in a Cross-Coupled Tracking System. *IEEE Transactions on Human Factors in Electronics*, *8*(3):210–217.

[181] Todosiev, E. P., Rose, R. E., & Summers, L. G. (1967). Human Performance in Single and Two-Axis Tracking Systems. *IEEE Transactions on Human Factors in Electronics*, *8*(2):125–129.

[182] Van Lunteren, A. (1979). *Identification of Human Operator Describing Function Models with One or Two Inputs in Closed Loop Systems*, unpublished Ph.D. Thesis. Delft, The Netherlands: Faculty of Mechanical Engineering, Delft University of Technology.

[183] Bekey, G. A., Meissinger, H. F., & Rose, R. E. (1965). Mathematical Models of Human Operators in Simple Two-Axis Manual Control Systems. *IEEE Transactions on Human Factors in Electronics*, *6*(1):42–52.

[184] Levison, W. H., Baron, S., & Kleinman D. L. (1969). A Model for Human Controller Remnant. *IEEE Transactions on Man-Machine Systems*, *10*(4):101–108.

[185] Jex, H. R., & Magdaleno, R. E. (1969). Corroborative Data on Normalization of Human Operator Remnant. *IEEE Transactions on Man-Machine Systems*, *10*(4):137–140.

[186] Popovici, A., Zaal, P. M. T., Pool, D. M., & Mulder, M. (2016). Relating Eye Activity Measures to Human Controller Remnant Characteristics. *Proceedings of the 13th IFAC/IFIP/IFORS/IEA Symposium on Analysis, Design, and Evaluation of Human-Machine Systems, Kyoto, Japan*, *49*(19):13–18.

[187] Popovici, A., Zaal, P. M. T., Pool, D. M., Van Paassen, M. M., & Mulder M. (2017). Effects of Eye Measures on Human Controller Remnant and Control Behavior. *Proceedings of the AIAA Modeling and Simulation Technologies Conference, Grapevine (TX)*, no. AIAA-2017-1316.

[188] Gottsdanker, R. M. (1956). The Ability of Human Operators to Detect Acceleration of Target Motion. *Psychological Bulletin*, *53*(6):477–487.

Part III

Preliminary Report Appendices

NOTE:
This part has already been graded under AE4020

Appendix A

Previous Manual Control Skill Retention Experiments

This appendix summarizes 16 experiments concerning the retention of manual control skills in different types of tasks, ranging from a simple rotary pursuit task, to bidimensional compensatory tracking task, to complex aircraft flight maneuvers, to spacecraft landing tasks. These experiments were reviewed to determine for how long different kinds of manual control tasks are retained as well as to analyze the shape of the skill decay curves. The secondary objective was to determine whether an investigation into the retention of manual control skills would be feasible within the time set for a M.Sc. thesis and if so, what kind of tracking task could best be used.

Rotary pursuit task 1 [123]

Purpose

To investigate the long-term retention of a rotary pursuit task.

Method

In a rotary pursuit task, the participant's objective is to follow (pursue) a small disc on a rotating turntable. The performance measure is time-on-target. During initial training, 47 subjects were given 20 one-minute trials on the pursuit rotor separated by rest periods of one minute each. During retention testing, again 20 one-minute trials separated by a one-minute rest period were performed.

The retention interval was equal to one year.

Results

Over a one-year retention interval, time-on-target performance decreased by 29%. On the first retention trial participants yielded a performance level equal to that obtained on trial 9 during initial training. Participants had completely regained their end-of-training performance after eight trials. Slight performance gains were observed during the remaining retention trials.

Rotary pursuit task 2 [124]

Purpose

To investigate the long-term retention of a rotary pursuit task after extended practice. This experiment differed from the one above [123] in that it did not only use considerably longer trials, but also more training trials.

Method

During initial training, eight subjects received 50 15-minute training trials, with trials being performed on successive days excluding weekends. Retention testing consisted of three 15-minute trials performed on three successive days, meaning one retention trial per day.

The retention interval equaled one year.

Results

After the retention interval, seven out of eight subjects exhibited a performance loss in time-on-target in the first retention trial. Performance had declined significantly by 10% on average when compared to the last training trial, but individual losses were between 3% and 25%. One subject even showed a performance gain of 6%. However, when comparing only the first 50 seconds of the first retention trial to the performance of the last training trial, an average performance decline of 37% was observed. There was no apparent relationship between the absolute level of performance on the last training trial and the amount of performance loss during the first retention trial. Performance loss over the retention interval was overcome by all subjects by the beginning of their second retention trial.

Although subjects in this experiment received significantly more training than participants in the previous experiment [123], their initial performance after the retention interval was not significantly different from the initial retention performance of participants in the previous experiment.

Rotary pursuit task 3 [125]

Purpose

To investigate the long-term retention of a rotary pursuit task after an extraordinarily long retention interval.

Method

This experiment retested 13 subjects who had taken part in an earlier rotary pursuit acquisition study an average of 15.5 years after task acquisition. Individual retention intervals varied between 14 and 18 years. Initial training consisted of three eight-minute practice periods with five minutes rest between subsequent periods. Each eight-minute practice period was divided into 24 20-second trials. The performance measure was cumulative time-on-target recorded for each trial. Retention testing followed the same procedure as initial training.

Results

Subjects suffered a performance loss of 37% on average, when comparing initial retention performance with end-of-training performance. When performance during the first minute of the second eight-minute retention period is compared to performance at the end of training, a retention performance of 99.5% is observed, meaning that performance was almost the same.

Single-axis pursuit tracking task [44]

Purpose

To investigate the effects of task predictability on the acquisition and retention of pursuit tracking skills.

Method

A single-axis pursuit tracking task was performed by 250 participants. The subject's objective was to superimpose a vertical line cursor on a one-inch vertical target line which could appear at any one of 15 equi-distance positions along the horizontal axis of the screen, by controlling a lateral arm controller. Performance was measured in terms of the mean absolute distance between the target position and cursor position across each trial. During initial training, all subjects were given seven

60-second trials, followed on the next day by an extra 20 trials. Initial training included detailed instructions including various sources of tracking error. Additional training consisted of either an additional 30 trials after initial training or an additional 80 trials after initial training over the next three days. During additional training, subjects were only given feedback on their error scores after every fifth trial. Four different levels of task predictability were used for different subjects. In the predictable task setting, 12 targets appeared in the same order and were repeated five times per trial for all trials. In the first and second intermediate predictability settings, every second, respectively, third target was selected at random using the predictable target sequence, whereas in the random predictability setting each target was selected at random for each trial.

The retention interval lasted either one week, one month or five months.

Results

The focus of the results will lie on the retention performance as a function of retention interval. In both the groups with less and more additional training, little performance decline was observed after a one-week retention interval. With continued retraining, performance generally continued to improve and surpassed the performance achieved at the end of training. One-month and five-month retention intervals produced larger performance declines. In retraining, performance quickly recovered, but never reached final training performance for any of the predictability settings. However, it was not found how many retraining trials subjects performed.

Acceleration control tracking task [46]

Purpose

To investigate the retention of a tracking task with acceleration control.

Method

An experiment was performed by 18 participants, of which six had some previous experience with tracking, to investigate skill retention in an acceleration control tracking task. The subject's objective was to align a moving spot of light with a stationary short vertical line by means of a thumb joystick. The acceleration control meant that the acceleration of the spot across the screen was proportional to the deflection of the stick, a condition much harder to control than position or velocity control [188], where, respectively, the spot's position and speed, are proportional to the stick deflection. The initial distance between the spot and the target line was 22.5 mm. The subject had to move the spot to the target and keep it within a zone of 1.5 mm on either side from the target for two consecutive seconds. When the subject had achieved this goal, the trial was ended. Researchers measured the time it took for subjects to bring the spot from 22.5 mm to a zone of 1.5 mm from the stationary vertical line and keep it there for two seconds minus the 2-second holding time. Before the experiment started, the movements required in an acceleration control task were explained to the subjects. Furthermore, the task was demonstrated by the researcher. Immediately after each trial, subjects were given knowledge of results. All subjects received five training trials per day until they had reached two criteria. These criteria were (1) three successive daily mean scores below 12 seconds, and (2) three successive daily scores which did not differ significantly at the 5% level. Half of the participants then stopped training. The other half received an additional ten trials per day until they had reached three successive daily scores which did not differ significantly at the 1% level. Whenever possible, training took place on consecutive days, however, sometimes a few days were missed. No training took place on the weekends. Retention testing continued at five trials per day until subjects had achieved the earlier stated first two criteria again. During retention testing, the group with additional training did not have to reach the more strict performance level anymore. The retention measure used was a numerical score defined as (the time of the first training trial minus the time of the first retention trial) divided by (the time of the first training trial minus the time of the last training trial). The lower the score, the worse the performance during the first retention trial was when compared to the last training trial.

Retention testing started six months after the last training day.

Results

The first training phase took subjects between eight and 22 days, meaning between 40 and 110 trials. The additional training, performed by half of the subjects, took participants between nine and 17 extra days, meaning between 90 and 170 additional training trials. At retention testing, only one subject achieved a score of 1, meaning that their performance at the first retention trial was just as good as their performance during the last training trial. However, it must be noted that this person was a qualified pilot. The lowest retention score achieved was -2.15. On the first retention trial, a five-fold increase in target acquisition time was seen for the group with less initial training and a two-fold increase for those with more initial training. Although the group with additional training was significantly better at their first retention trial than the group with less training, the decline in performance at initial retention testing when compared to the last training trial was significant for both groups. For both groups it took only three subjects more than five days, meaning 25 retention trials, to meet the two earlier stated criteria. Hammerton concluded that if a task is sufficiently difficult, overlearning does not prevent, but only somewhat reduces, a decline in performance.

Bidimensional compensatory tracking task with extended practice [126]

Purpose

To investigate the retention of a bidimensional compensatory tracking task after extended practice.

Method

Three of the four authors of the article served as experiment subjects and all had previous experience in performing tracking tasks. The subject's objective was to move a circle with a diameter of 1/8-inch within a target-scoring area defined by a circle with a diameter of 1/2-inch by means of a control stick which could be manipulated with the right hand. The performance measures were azimuth and elevation distances between manipulated marker and target circle as well as time-on-target in azimuth and elevation. The standard movement condition associated right-left and backward-forward movement of the stick with right-left and up-down movement of the marker on the screen. Training consisted of ten one-minute trials per day for 100 days with 20 seconds between each trial, except for a one-minute rest between trial 5 and 6 each day. In each trial the same two target functions in azimuth and elevation direction were applied. After the 100 training days, 14 additional training days with ten trials each day were performed under various conditions. On days 101 and 102 the target functions were reversed in time. On days 103 through 106 subjects trained with the original target functions again. On days 107 through 110 a reversed direction of control movement relative to target movement was used for both azimuth and elevation, while the original target functions were applied. Lastly, an additional four training days (days 111 - 114) were taken to practice once more with the standard movement condition and the original target functions. Retention testing consisted of four retraining days of ten trials each with the standard movement condition and the original target functions.

The retention interval lasted eight months.

Results

Although each trial consisted of identical target functions, none of the subjects believed they had learned the target functions by the end of training. The task was considered to be of relatively great difficulty because an asymptotic level of performance was only reached by approximately the eightieth day, meaning after approximately 800 one-minute trials. No statistically significant improvement in performance occurred after that day. Reversal in direction of control movement on days 107 through 110 resulted in significant worse performance and high negative transfer. Even though performance with reversed control movement significantly improved over the four days, performance after 40 trials, meaning the complete four days with reversed control movement, was still worse than the initial level of performance under the standard control movement condition. Switching back to the standard control movement condition on day 111, recovery of performance to the performance level achieved at day 100 of training was completed in about 20 trials, meaning two days. After that, training performance did not significantly differ anymore from that on training day 100.

After the 8-month retention interval, a high degree of retention was observed. Although the initial retention trials showed numerically worse performance compared to performance at the end of training, none of the four retention testing daily performance means were significantly different from the daily performance means at the end of the 100 training days nor from the daily means at days 111 through 114 just prior to the retention interval.

Image motion compensation task [127]

Purpose

To investigate the effects of amount of training and duration of the retention interval on retention performance.

Method

An image motion compensation task was performed by 96 participants. The visual dynamics of an earth orbital flyby at 100 nautical miles were simulated. Subjects had the objective to null the motion of a photographic mosaic depicting an earth target area by means of a pencil stick controller. Disturbances in the simulated motion of the mosaic were introduced into the optical system by simulating thruster firings. Subjects received either 60 or 120 training trials of 40 seconds each. During retention testing, 25 retraining trials were performed which were identical in length to those given during initial training. The number of seconds per trial during which subjects successfully nulled image motion within a specific constraint were used as performance measure. Performance loss was determined by subtracting the duration of time that image motion was nulled during the first retention trial from the mean time that motion was nulled during the last 15 initial training trials.

The retention interval was either 30, 90 or 200 days.

Results

The performance loss of the group with a 30-day retention interval was on average 1.33 seconds, of the 90-day interval group 3.28 seconds and of the 200-day interval group 7.19 seconds. On average the absolute performance loss was more than twice as large for the group with 60 training trials compared to the group with 120 training trials, causing amount of initial training to have a significant effect on retention performance. A significant difference in the rate of skill reacquisition was found between the groups with 30-day and 200-day retention intervals. The 30-day group regained their skills more rapidly than the 200-day group. However, there was almost no difference in rate of skill reacquisition between the 30-day and 90-day groups. Additionally, it is unknown how many of the 25 retraining trials were required to regain end-of-training performance.

Simulated radar intercept mission tracking task [78]

Purpose

To investigate the retention performance of a slightly more complex tracking task.

Method

The tracking task studied, called the simulated radar intercept mission, was performed by 130 participants. The subject's objective was to maintain a target dot at the center of a display and at the same time keep the sideslip at zero. Subjects used a control stick and rudder pedals as the controls. The controls were constructed in such a way as to produce realistic aircraft compensatory movements on the display. Participants performed 17 training sessions distributed over six weeks, where one training session consisted of 21 one-minute trials. An absolute error score was calculated from absolute errors measured on the controls in terms of azimuth, elevation and sideslip. Retention testing consisted of four sessions of 21 one-minute trials. One group of participants performed all four sessions on the same day, whereas the other group performed one session per day for four consecutive days. One week after retention testing all subjects performed one additional session to investigate performance differences due to different retraining schedules.

Retention performance was measured after either one month, five months, nine months, 14 months or 24 months.

Results

Very high skill retention was found for no-practice intervals up to 24 months. Retention performance was not significantly different for retention intervals ranging from one to 14 months, not even in the first minute of retention testing. Virtually no performance losses were observed during any retention testing up to the 14-month retention interval. Very small losses were recovered in the first two to three minutes of retraining. After the 24-month retention interval, performance losses were slightly larger, but regaining the lost skills only took about 20 minutes. It must be noted that subjects receiving the distributed retention testing performed better than subjects receiving the massed retention testing. However, during the additional testing trial one week later, there were no significant differences in final performance between the two groups.

Note that the amount of initial training was much more than the training time in the current research would be.

Three-dimensional compensatory tracking task with concurrent discrete procedural task [113]

Purpose

To investigate the retention of a three-dimensional compensatory tracking task with a concurrent discrete procedural task.

Method

The experiment was performed by 128 participants. The tracking task required subjects to get the error scores in roll, pitch and yaw down to zero. In total, six meters were used to indicate the attitude error and rate error of the three attitudes. A three-dimensional control stick was used by the subjects to manipulate the attitudes. Concurrently, subjects were required to press buttons in an appropriate sequence in response to the presence or absence of a light on the procedural task panel. Initially, participants were trained on the tracking task and procedural task in alternating fashion, but after the fifth training session, practice was changed to whole-task training. Each training session consisted of 12 trials of 70 seconds. Subjects were trained for either two or three weeks in total. However, it was not found how many training sessions participants performed exactly. The performance measure for the tracking task was the absolute error for each of the six meters, attitude and rate error for the three attitudes.

The retention interval was either one or four weeks.

Results

Only the results of the tracking task will be discussed here. Superior training performance was attained by the group receiving the greater amount of training. Additionally, retention testing revealed that statistically significant better retention performance was attained by the group with three weeks of initial training compared to the group with two weeks of training. However, for both groups a performance decrement was observed after the four-week retention interval. For the one-week retention interval, little performance loss was seen. The researchers concluded that amount of training has a major influence on skill retention performance as retention varied directly with level of training.

Compensatory pursuit task for aircraft control [79]

Purpose

To determine the effects of amount of practice and the duration of the retention interval on task performance.

Method

A compensatory pursuit task was performed by 465 participants. The main objective was to maintain an aircraft in level straight ahead flight by using stick controls and food pedals to compensate for movements made by the aircraft model. Individual trials lasted about one minute with a rest of ten seconds between subsequent trials. After every 30 minutes, five minutes rest was given. Performance was measured in terms of time-on-target, where the illumination of a small red light was used to indicate to subjects that they were on target. Subjects received either one or eight hours of training. Each of the retention intervals was followed by two hours of retraining.

The retention interval was either one day, one month, six months, 12 months or 24 months.

Results

After the first 15 minutes of initial training, subjects were on target about 40% of the time. After the first hour, they were on target 70% of the time, and after eight hours 90% of the time. After a retention interval of six months, time-on-target scores of the group with eight hours of training were held up fairly well when compared to end-of-training performance, whereas for the group with less training performance degradation was noticeable. After retention intervals of one to two years, the initial retention performance for the one-hour training group was between 50% and 60% on target, meaning a proportional performance loss of 15% to 30% when compared to the end-of-training performance of 70% time-on-target. Performance improvement continued throughout one hour of retraining. For the eight-hour training group, 50% to 75% of the performance loss after each of the retention intervals was regained during the first five minutes of retraining. However, after a retention interval of two years very slight performance gains were still realized during the first 48 minutes of retraining. After the two-year retention interval, the absolute performance loss was about the same for both training groups.

Ammons et al. concluded that the absolute performance loss is independent of the amount of training, meaning that only the proportional loss is greater for individuals with less training, because they usually achieve lower performance at the end of initial training than individuals with more training, as could also be seen in this experiment. A similar observation was made in Figure 3-1 for differences in initial ability levels instead of in amount of training.

Complex simulated flight task [36]

Purpose

To investigate the effects of amount of training on the forgetting of instrument procedural and control skills.

Method

A variety of complex simulated flight tasks were performed in a moving-base flight simulator. Subjects were naive to flying and were given four hours of classroom instruction, one familiarization trial in the simulator and after that, either five or ten identical training trials. Each trial was a 50-minute mission of maneuvers and procedures from starting the engine and take-off to landing and shutting down the engine. The trial included take-off and climb, straight and level flight, level turns, climbing and descending turns, level-offs, changes in flight speed, gliding and final approach. After the retention interval, four retention trials identical to the initial training trials were given. Performance measures included 125 procedural items and five flight control parameters measured every ten seconds. The primary flight parameters recorded were altitude, airspeed, bank angle, roll-out on new heading angle and level-off at altitude. Proficiency in flight control was measured in terms of the extent to which basic flight instruments were kept aligned with a value defined as ideal for the specific maneuver. Each parameter was scored in terms of error deviation from the desired performance.

The retention interval was equal to four months.

Results

The retention of procedural skills will not be discussed here, because of the research focus on motor skills. In terms of control skills, only altitude and airspeed had a statistically significant loss over the retention interval for both training groups. The performance loss in bank angle was only statistically significant for the group with less training trials. However, all performance losses were operationally insignificant. The researcher also concluded that the absolute amount of retention loss was independent of the amount of initial training, as Ammons et al. also did in their compensatory pursuit task discussed above [79]. During the first retention trial, retention performance of the group with more initial training was superior to the performance of the group with less training, which was also seen at the end of initial training. This means that a high level of initial training leads to a higher performance level at retention testing. For all flight control parameters, the time required to relearn differed little as a function of amount of initial training. It took both groups about three trials on average to get the flight control parameters back at end-of-training performance levels.

Simulated carrier landing task [128]

Purpose

To investigate the effects of non-flying periods on the retention of aircraft landing skills.

Method

A simulated carrier approach and landing experiment was performed by fifteen carrier qualified Naval Aviator students with backgrounds in attack or fighter-type jet aircraft. Average carrier experience was two carrier deployments of six to ten months each and a total flight time of 1600 hours, but individual numbers varied from 300 to 3300 hours. Subjects were divided into three groups. The first group had all flown operationally within the last 60 days. The second group had not been operationally flying for the last ten to 17 months, whereas the last group had not flown operationally for 25 to 30 months. This meant that the experiment only consisted of retention testing. The previous experience of the pilots can be regarded as the 'training'.

The task involved a minimal aircraft simulation and a computer image visual display of the carrier. The pilot's objective was to safely approach and land as closely as possible to the center line of a carrier deck while maintaining the correct landing airspeed. Flight controls used were a control stick, providing roll, pitch and yaw inputs and mounted on the right arm of the pilot chair, and a throttle quadrant, mounted on the left arm of the chair. The visual display depicted an inside-out runway/carrier deck, a landing mirror and a set of indexer lights. Performance was measured in terms of the sink rate at touchdown, the line-up with respect to the centerline of the landing field, the airspeed and the landing result. Landing results could include a crash due to for example an extremely hard landing or a bolter, which is when an aircraft's tailhook misses the arresting gear on the carrier deck so that the aircraft is required to take off again without stopping. Prior to testing, one demonstration run was flown by the researcher to familiarize subjects with the equipment and the display. After that, three practice runs were flown by each pilot, followed by three runs to measure retention performance.

Results

A small retention difference was observed between the first and the second group, in favor of the first group, who had most recently been flying, whereas the second and third group performed almost identical. However, performance losses were restricted to primarily losses in procedural tasks, whereas continuous (tracking) skills were retained fairly well for non-flying periods up to 30 months.

Time-shared flight maneuver [38]

Purpose

To investigate the use of part-task trainers in the training and retention of complex time-shared tasks.

Method

A fairly complex flight maneuver, namely a bomb-toss, was trained and later retested. Two groups of ten subjects each, differing in part-task and whole-task learning, trained the procedural and flight control components of the bomb-toss. Training trials were three minutes in length. However, it could not be found how many initial training trials were given. During retention testing ten retention trials were performed.

The retention interval equaled ten months.

Results

After a 10-month retention interval, proficiency in procedures had dropped from 95% to 5%, whereas flight control skills were retained fairly well. Although control of the bank angle showed a statistically significant proficiency loss, the mean increase in bank angle was less than two degrees and therefore not considered of practical operational importance. Furthermore, although the mean vertical speed error increased from about 150 ft/min on the last initial training trial to about 500 ft/min on the first retention trial, end-of-training proficiency was regained by the second, respectively, fifth three-minute retention trial for groups with and without procedural retraining before the whole-task retention trials. All other flight control parameters remained at acceptable levels. It must be noted, however, that the sampling rate of the flight control parameters was rather limited.

Simulated space flight from lift-off to orbit insertion [37]

Purpose

To investigate the effectiveness of practice methods on simulated space flight skill retention.

Method

An experiment concerning the control of a reusable space vehicle during the boost phase was performed by 45 relatively task-naive subjects. Subject performance was measured on a continuous control task and a procedural task. However, in this experiment summary the focus will lie on the continuous control task. The subject's objective was to provide control inputs to fly the optimum trajectory from lift-off to orbit insertion using a three-axis side arm control stick on the right arm of the pilot's seat as well as controls mounted on a bulkhead. The control stick provided rate inputs as a linear function of angular displacement and stick force. Displayed information included altitude, altitude rate, altitude error, angle of attack, velocity, vehicle pitch and roll error, side slip and compass heading error. The vehicle pitch and roll errors were displayed on an early model of the Attitude Director Indicator. Performance was measured in terms of the absolute altitude and altitude rate errors at orbit insertion as well as the absolute altitude and pitch errors from the nominal altitude and pitch profiles integrated over the duration of the total manually controlled flight. Initial training started with a basic lecture about the training data package. Then, subjects were seated in the cockpit and familiarized with the operational instruments and indicators. After that, subjects observed the instruments while the simulator flew two automatic runs at ten times the real-time speed as well as one real-time automatic run. Next, subjects started their first manually controlled flight. The researcher talked the subjects through their first three training flights. Training continued in sets of five flights until the subject's average performance for five consecutive flights met specific training qualification test performance criteria. These training qualification test performance criteria were the same as for trained pilots performing real space missions. Each flight took approximately 6 minutes, 44 seconds to complete. Retention testing consisted of one retention trial with the optimum trajectory applied in initial training, followed by four more flight procedure sequences. Rest time between consecutive flights was approximately one minute. After these five retention flights, a five-minute break was provided for subjects to be able to stretch their legs, after which another series of five flights and procedure sequences were flown. During the retention interval pilots were not to perform any piloting functions in other flight simulators or actual aircraft.

Retention intervals varied from one to six months in steps of one month with the exception of a five-month retention interval, which did not exist.

Results

Training took on average 34 flights, but individual numbers varied between 13 and 68 flights. Acceptable flight control performance was retained for two months, but deteriorated rapidly thereafter by a factor of 1.7 to 3.1 depending on the performance measure. The performance data suggested that skill degradation had reached its peak at about four months, after which it remained stable. However, it was not found whether subjects reached the performance criteria of training again during retention testing, and if so, how many flights it took them.

Simulated space vehicle approach and landing under instrument and visual flight conditions [98]

Purpose

To investigate the effectiveness of practice methods on visual approach and landing skill retention.

Method

A simulated spacecraft approach and landing was trained by 15 experienced pilots currently not flying, with an average of 5.3 years since their last flight, but at the start of the experiment individual numbers varied between six months and 16 years since their last flight. The pilot's task was to control the vehicle from an altitude of 31,000 feet through a descending turn and make an approach and landing on a runway using a two-axis, sidearm controller providing proportional rate commands for pitch and roll and rudder pedals providing displacement commands for yaw. This flight profile took approximately 6 minutes, 45 seconds to complete. Training consisted of one hour of ground school on flight and procedural tasks, cockpit familiarization, procedural task training, visual flight and landing practice, full practice flights of instrument approaches to visual landings, and full mission flights including emergency procedure tasks. In this summary, the focus will lie on flight task performance instead of on the procedural tasks. Training was distributed over five days and continued until the means and standard deviations of selected performance parameters reached an asymptotic level of acceptable flight performance. Performance was measured by integrated errors determined from a nominal flight path between control check points or reference planes as well as by the instantaneous errors observed when crossing these control check points. In total, 11 parameters were evaluated. However, since not all parameters were evaluated the same number of times, a total of 32 flight control data measurements were obtained. The most important performance parameters included measurement of lateral and vertical offset from the glide path, airspeed error, rate of descent, down range and lateral errors upon landing and descent rate at touchdown. During the retention interval pilots were not to perform any piloting functions in other flight simulators or actual aircraft. Retention testing consisted of five retention trials. At the end of each flight pilots received feedback information on the distance down the runway and the descent rate at touchdown. After these five flights, pilots were allowed to stretch their legs during a ten-minute break. After the break, one more simulated approach and landing was performed.

The retention interval equaled four months.

Results

Training took an average of 48 flights per pilot to train to proficiency, but individual numbers varied between 30 and 76 flights. After a four-month retention interval, flight control performance had degraded significantly. Based upon normal operational limits, the majority of the performance parameters exhibited degradation of practical importance. The magnitude of visual flight control degradation was comparable to that found in the study on the retention of control skills required to control a spacecraft during its boost phase [37]. During the six retention test flights, end-of-training performance was not reached again. It must also be mentioned that the general conclusion was that the simulated flight was a realistic reproduction of an approach and landing representative for a large flight vehicle.

Simulated lunar landing [99]

Purpose

To obtain an estimate of the degree of skill loss which may be expected over time intervals up to three months in tasks typical of space vehicle operation.

Method

Four crews of three aerospace research pilots each were trained for a period of six weeks in the performance of a simulated seven-day lunar landing mission. A full-size, high-fidelity mock-up of the Apollo Command Module and Lunar Excursion Module were used. Although in the Command Module three operator stations were provided, namely one for a pilot, a navigator and an engineer, the focus here will lie on the pilot, because the pilot was in control of the main controls and instruments required for controlling the vehicle. The controls and displays were mainly used to monitor and control translatory accelerations and attitude changes in pitch, roll and yaw. The Excursion Module also contained controls and displays for translation and attitude. These controls consisted of two control sticks and an engine throttle.

The first three to five days were devoted to reconnaissance of the mission and familiarization with the vehicle systems, displays and controls. Study time was given for written material regarding the overall mission. After that, subjects were introduced to the simulator, in which they practiced various mission phases. Each phase consisted of a sequence of tasks. During the last five to seven days, all mission phases were practiced in order. The last part of training consisted of a simulation of the entire seven-day mission in real time. For retention testing purposes, the mission was compressed into a single 13-hour working day by the omission of non-critical flight phases, navigational tasks and certain secondary activities. Performance was measured by analyzing 22 flight control parameters considered critical to mission performance. The values of the flight control parameters were then converted to probabilities of meeting specific performance criteria.

The retention interval was either four weeks, eight weeks, nine weeks or 13 weeks. Each crew got assigned a different retention interval. The crews assigned to the eight-week and 13-week retention intervals received substantially more practice than the two crews with the four-week and nine-week retention intervals.

Results

Because the two crews with the four-week and nine-week retention intervals received less practice, their end-of-training performance was not good enough to use the data. For the two crews with the eight-week and 13-week retention intervals, a significant decline in performance over the retention interval was observed. During three days of additional retraining, they did not regain performance exhibited at the end of initial training.

Appendix B

Call for Experiment Participants

The following page presents the poster that was used during an AE2235-II Instrumentation and Signals lecture to recruit second-year bachelor students in Aerospace Engineering for participation in the experiment. This poster was also used to inform first-year and second-year bachelor students in Aerospace Engineering during a project session of AE1222-I Design and Construction, and AE2223-I Test, Analysis and Simulation, respectively, about the upcoming experiment. Additionally, this poster was hung up in the faculties of Aerospace Engineering, Applied Sciences, Civil Engineering and Geosciences, Electrical Engineering, Mathematics and Computer Science, Mechanical, Maritime and Materials Engineering, and Technology, Policy and Management, as well as the Aula Conference Centre of the Delft University of Technology. Furthermore, this poster was used to promote the experiment through various social media channels, including the official TU Delft Aerospace Engineering Facebook page.

LOOKING FOR EXPERIMENT PARTICIPANTS!

For my MSc thesis into the retention of manual control skills, I am looking for participants to perform a tracking task in the fixed-base simulator in the HMI Lab at the Faculty of Aerospace Engineering.

Experiment details:

- You will be ***trained on 4 consecutive days in the 4th period***. This will take about ***1 hour per day***.
- You will be ***retested at maximum 3 times***. A retest will take about ***1 hour***.
- The last retest will take place before Christmas.
- Try to achieve the best score of everyone at retesting!

Participant requirements:

- ***Right-handed***
- ***No pilot experience and little tracking task experience***
- ***No participation in other tracking or flying experiments until last retest***
- ***Available for training and a maximum number of 3 retests. The last retest will take place before Christmas.***

Experiment schedule

Academic calendar weeks																				
Training weeks				Retesting weeks																
4.4	4.5	4.6	4.8	5.1	5.2	5.3	5.8	1.1	1.2	1.3	1.4	1.5	1.6	2.1	2.2	2.3	2.4	2.5	2.6
1 hour/day @ 4 consecutive days in total				1 hour in total				1 hour in total						1 hour in total					

- ***No testing in exam periods!***
- Available in 1 of the training weeks (4 consecutive days) for one hour per day
- Retested a maximum number of 3 times
- A maximum of one retest per academic period (might include summer period)
- Every retest will take about one hour

Interested? Please send an email to:
r.wijlens@student.tudelft.nl

Appendix C

Experiment Briefing

On the following pages, the experiment briefing, mentioned in Section 6-2-9, is presented. The briefing was provided to participants 2 to 4 days before the start of the training phase of the experiment. The experiment briefing discussed the goal of the experiment globally, as well as the tracking task and the experiment procedures.

Experiment Briefing

The retention of manual control skills

Thank you for your contribution to this scientific endeavour! You will be participating in a tracking experiment in the Human-Machine Interaction Laboratory (HMI Lab) at TU Delft, in which the retention of manual control skills is investigated using a dual-axis aircraft roll and pitch tracking task. This briefing will introduce you to the experiment and what is expected of you as a participant.

Experiment Goal

The goal of this experiment is to investigate the retention of a dual-axis aircraft roll and pitch tracking task. Retention can be explained as the condition of retaining a specific skill or knowledge through a period of non-practice. The tracking data will be used to explicitly quantify skill development, decay and retention. The results of this experiment can be used to make recommendations on training procedures, training intervals as well as on refresher training.

Dual-Axis Aircraft Roll and Pitch Tracking Task

The task you will carry out is a dual-axis aircraft roll and pitch tracking task with compensatory display. Imagine you are flying an aircraft. It is your goal to actively and simultaneously minimize the aircraft roll and pitch errors. <u>It is extremely important that you constantly try to remove the smallest errors in pitch and roll. In order to do this, you have to actively steer the aircraft at all times. Do not wait until the aircraft levels out by itself.</u> The tracking task can be compared to landing an aircraft using an attitude indicator (which illustrates both roll and pitch attitudes).

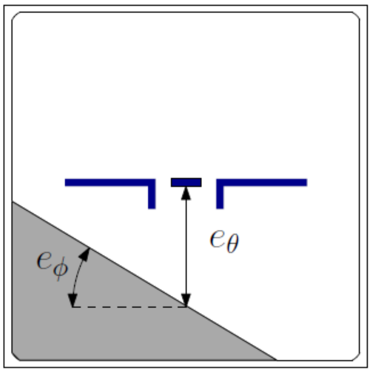


Figure 1: Experiment display.

The aircraft roll and pitch errors will be displayed on the primary cockpit display using a simplified artificial horizon instrument, as illustrated in Figure 1. The aircraft's attitude is displayed by fixed wings and the roll and pitch errors are displayed using a translating and rotating "ground" (illustrated in Figure 1 as the grey shape), on a contrasting background. Note that the error symbols e_φ and e_θ as well as their corresponding arrows, as illustrated in Figure 1, are not shown on the primary cockpit display.

During the experiment, you will control the aircraft's roll and pitch attitudes by actively and continuously providing smooth control inputs using a side-stick on the right-hand side of the seat. Roll is controlled by left-right stick movements and pitch is controlled by fore-aft stick movements. The display is positioned in front of you. Your main objective is to keep both errors on the display as close to zero as possible at all times. To correct the errors shown in Figure 1, a simultaneous pitch-down (forward push) and roll-right (pull to the right) input is required. After each run, two scores will be displayed on the primary cockpit display to indicate your performance in pitch and roll, respectively. <u>A lower score indicates a better performance.</u>

Experiment Procedures

The first part of the experiment is the training phase. The training phase consists of four consecutive days. On each training day, you will carry out 25 runs of 90 seconds. Short breaks can be taken between runs to alleviate any discomfort that might occur in the right hand, arm or shoulder due to controlling the side-stick or after sitting in a fixed position for a prolonged period of time. Also, a 5-minute break will be taken after you have performed the first 15 runs. During the training phase, <u>your goal should be to constantly attempt to improve your scores.</u> Each training session will take about one hour. This means that the training phase will take about four hours in total spread out over four consecutive days.

After the training period, you will be assigned to one of the three retention groups. Your group will determine how many retention tests you will perform (one, two or three). The duration of the retention intervals might vary from 8 - 9 weeks (if you have been selected for three retention tests), 12 - 14 weeks (if you have been selected for two retention tests) to 24 - 28 weeks (if you have been selected for only one retention test). The last retention test will always take place 24 - 28 weeks after training. During retention testing, your task is the same as during the training phase. Your goal is also the same as during the training phase: constantly attempt to improve your scores. Every retention test will take about one hour. This means that retention testing will take about one to three hours in total, depending on the number of retention tests you will perform.

The total time of the experiment will be between five and seven hours, depending on the number of retention tests, and spread out over a time period of several months.

Table 1 summarizes the schedule of the experiment.

<table>
<tr><td colspan="6">Training period</td></tr>
<tr><td colspan="3">Day 1</td><td colspan="3">25 runs</td></tr>
<tr><td colspan="3">Day 2</td><td colspan="3">25 runs</td></tr>
<tr><td colspan="3">Day 3</td><td colspan="3">25 runs</td></tr>
<tr><td colspan="3">Day 4</td><td colspan="3">25 runs</td></tr>
<tr><td colspan="6">You are assigned to Group 1, 2 or 3</td></tr>
<tr><td colspan="2">Group 1</td><td colspan="2">Group 2</td><td colspan="2">Group 3</td></tr>
<tr><td colspan="2" rowspan="5">RI 1 (±24 – 28 weeks)</td><td colspan="2" rowspan="2">RI 1 (±12 – 14 weeks)</td><td colspan="2">RI 1 (±8 – 9 weeks)</td></tr>
<tr><td colspan="2">Retention test 1</td></tr>
<tr><td colspan="2">Retention test 1</td><td colspan="2">RI 2 (±8 – 9 weeks)</td></tr>
<tr><td colspan="2" rowspan="2">RI 2 (±12 – 14 weeks)</td><td colspan="2">Retention test 2</td></tr>
<tr><td colspan="2">RI 3 (±8 – 9 weeks)</td></tr>
<tr><td colspan="2">Retention test 1</td><td colspan="2">Retention test 2</td><td colspan="2">Retention test 3</td></tr>
</table>

Table 1: Experiment schedule (RI = Retention Interval).

Experiment Execution

For each tracking run, the subsequent procedure will be followed:

1. The researcher applies the settings for the next run.
2. The researcher checks whether the participant is ready to proceed and initiates the run after a countdown from 3 (3-2-1-go).
3. The participant performs the tracking task.
4. The participant is notified of their performance in the completed run in terms of error scores in pitch and roll displayed on the primary cockpit display after the completed run.

<u>Contact information researcher:</u>
Rowenna Wijlens
r.wijlens@student.tudelft.nl
+31 6 38069280

<u>Contact information research supervisor:</u>
dr. ir. Daan Pool
d.m.pool@tudelft.nl
+31 15 2789611

Appendix D

Experiment Consent Form

The following page presents the experiment consent form, mentioned in Section 6-2-9. The experiment consent form was signed by all participants on their first day of training before the start of the experiment.

Experiment Consent Form
The retention of manual control skills

I hereby confirm that:

1. I volunteer to participate in the experiment conducted by the researcher (**Rowenna Wijlens**) under Supervision of **dr.ir. Daan Pool** from the Faculty of Aerospace Engineering of TU Delft. I understand that my participation in this experiment is voluntary and that I may withdraw and discontinue participation at any time, for any reason.
2. I have read the experiment briefing. Also, I affirm that I understand the experiment instructions and have had all remaining questions answered to my satisfaction.
3. I understand that my participation involves performing a simple manual control task in a fixed-base simulator setup. I confirm that the researcher has provided me with detailed safety and operational instructions for the hardware (simulator setup, hydraulic sidestick) used in the experiment.
4. I understand that my participation involves being available for a training period of four consecutive working days with training sessions of about one hour per day. I also understand that I am expected to be available for a maximum number of three one-hour retests, where the last retest will take place before January 2019.
5. I understand that I am expected to refrain from participation in other tracking task or flying experiments until the last retest has taken place.
6. I understand that the researcher will not identify me by name in any reports or publications that will result from this experiment, and that my confidentiality as a participant in this study will remain secure.
7. I understand that this research study has been reviewed and approved by the TU Delft Human Research Ethics Committee (HREC). To report any problems regarding my participation in the experiment, I know I can contact the researchers using the contact information below.
8. I have been given a copy of this consent form.

My signature	Date
My printed name	Signature of researcher

Contact information researcher:	Contact information research supervisor:
Rowenna Wijlens	dr. ir. Daan Pool
r.wijlens@student.tudelft.nl	d.m.pool@tudelft.nl
+31 6 38069280	+31 15 2789611

Appendix E

Experiment Questionnaire

The following page presents the experiment questionnaire mentioned in Section 6-2-9 which was filled out by all experiment participants during the training phase of the experiment. This questionnaire was meant to gain insight into participants' activities that could potentially influence their initial learning as well as their retention performance.

Questionnaire
The retention of manual control skills

Name:	
Age:	
Have you read the experiment briefing?	Yes / No
Do you (currently) play any video games in which any kind of tracking skill is used?	No / Yes, on a(n) (almost) daily basis / Yes, on a weekly basis / Yes, on a monthly basis / Only a few times per year
Have you been playing video games which require any kind of tracking skill for a long time already or have you played these kinds of video games in the past?	No / Yes, I am still playing these kinds of video games and have been for about ….. year(s) now / Yes, I have played these kinds of video games for about ….. year(s) on a regular basis, but I have quit playing ….. year(s) ago
Are there any other activities / things you do / have done that you believe influence / have influenced your ability to perform tracking tasks?	
Any other remarks you would like to make or you think might be worth mentioning for this research:	

Appendix F

Training Schedules

On the following pages, the training schedules of the 39 participants that completed the training phase of the experiment, are presented. The total training phase in which all participants were trained, was spread out over five weeks. Since participants 10, 20, 32 and 36 did not finish the training phase due to various reasons, their training sessions have been removed from the training schedules all together for clarity. Because 39 participants completed their training phase and four participants dropped out, the largest participant number found in the training schedules is 43. It was decided not to renumber the subjects to avoid mix ups with the experiment data during data analysis.

A few remarks about the training schedules have to be made:

- Although it cannot be seen from the training schedule of week 1, subject 9 was not able to perform 25 training runs per training session due to problems with the fixed-base simulator in the HMI Lab. In the end, he or she was able to perform a total of 100 training runs spread out over four consecutive days, but using the following distribution:
  - Day 1: 8 runs
  - Day 2: 28 runs
  - Day 3: 34 runs
  - Day 4: 30 runs
- As can be seen in the training schedule of week 2, subject 11 was not able to perform his or her second day of training due to illness. Missing the second training day was compensated for by performing two training sessions on day 4 of training, one at the start of the day and another one at the end of the day. This was considered the best viable option, since this still allowed for some rest time between the third and fourth training session.
- As can be seen in the training schedule of week 3, the training phase of subject 26 started one day later than that of the other subjects in that week and ended one day later as well.

Training week 1

Day 1		Day 2		Day 3		Day 4	
						7:30 – 8:30	**S3**
8:30 – 9:30	**S1**	8:30 – 9:30	**S7**	8:30 – 9:30	**S7**	8:30 – 9:30	**S7**
9:30 – 10:30	**S2**	9:30 – 10:30	**S8**	9:30 – 10:30	**S6**	9:30 – 10:30	**S8**
10:30 – 11:30	**S3**	10:30 – 11:30	**S3**	10:30 – 11:30	**S3**	10:30 – 11:30	**S4**
11:30 – 12:30	**S4**	11:30 – 12:30	**S4**	11:30 – 12:30	**S1**	11:30 – 12:30	**S2**
12:30 – 13:30	**S5**	12:30 – 13:30	**S5**	12:30 – 13:30	**S5**	12:30 – 13:30	**S5**
13:30 – 14:30	**S6**	13:30 – 14:30	**S6**	13:30 – 14:30	**S4**	13:30 – 14:30	**S6**
14:30 – 15:30	**S7**	14:30 – 15:30	**S1**	14:30 – 15:30	**S2**	15:00 – 16:00	**S1**
15:45 – 16:45	**S8**	16:00 – 17:00	**S2**	15:30 – 16:30	**S8**		
17:00 – 18:00	**S9**	17:00 – 18:00	**S9**	17:00 – 18:00	**S9**	17:00 – 18:00	**S9**

Training week 2

Day 1		Day 2		Day 3		Day 4	
				7:30 – 8:30	**S13**		
8:30 – 9:30	**S11**					8:30 – 9:30	**S11**
9:30 – 10:30	**S12**	9:30 – 10:30	**S15**			9:30 – 10:30	**S13**
		10:30 – 11:30	**S16**	10:30 – 11:30	**S12**	10:30 – 11:30	**S18**
11:30 – 12:30	**S13**	11:30 – 12:30	**S12**	11:30 – 12:30	**S17**	11:30 – 12:30	**S15**
12:30 – 13:30	**S14**	12:30 – 13:30	**S14**	12:30 – 13:30	**S14**	12:30 – 13:30	**S14**
13:30 – 14:30	**S15**			13:30 – 14:30	**S15**	13:30 – 14:30	**S12**
15:00 – 16:00	**S16**	15:00 – 16:00	**S17**	15:00 – 16:00	**S16**	15:00 – 16:00	**S16**
16:00 – 17:00	**S17**	16:00 – 17:00	**S13**	16:00 – 17:00	**S18**	16:00 – 17:00	**S17**
17:00 – 18:00	**S18**	17:00 – 18:00	**S18**	17:00 – 18:00	**S11**	17:00 – 18:00	**S11**

Training week 3

Day 1		Day 2		Day 3		Day 4	
9:30 – 10:30	**S19**	9:30 – 10:30	**S22**	9:30 – 10:30	**S19**	9:30 – 10:30	**S19**
		10:30 – 11:30	**S23**	10:30 – 11:30	**S23**	10:30 – 11:30	**S23**
11:30 – 12:30	**S21**	11:30 – 12:30	**S21**	11:30 – 12:30	**S21**	11:30 – 12:30	**S25**
12:30 – 13:30	**S22**	12:30 – 13:30	**S25**	12:30 – 13:30	**S26**	12:30 – 13:30	**S21**
13:30 – 14:30	**S23**	13:30 – 14:30	**S26**	13:30 – 14:30	**S25**	13:30 – 14:30	**S26**
		14:30 – 15:30	**S24**	14:30 – 15:30	**S22**	14:30 – 15:30	**S24**
16:00 – 17:00	**S25**	16:00 – 17:00	**S19**	15:45 – 16:45	**S24**	16:00 – 17:00	**S22**
17:00 – 18:00	**S24**						

Day 5	
17:15 – 18:15	**S26**

Training week 4

Day 1		Day 2		Day 3		Day 4	
17:30 – 18:30	**S27**	17:30 – 18:30	**S27**	17:30 – 18:30	**S27**	17:30 – 18:30	**S27**
18:30 – 19:30	**S28**	18:30 – 19:30	**S28**	18:30 – 19:30	**S28**	18:30 – 19:30	**S28**
19:45 – 20:45	**S29**	19:45 – 20:45	**S29**	19:45 – 20:45	**S29**	19:45 – 20:45	**S29**
20:45 – 21:45	**S30**	20:45 – 21:45	**S30**	20:45 – 21:45	**S30**	20:45 – 21:45	**S30**

Training week 5

Day 1		Day 2		Day 3		Day 4	
				7:30 – 8:30	**S31**		
8:30 – 9:30	**S31**	8:30 – 9:30	**S31**			8:30 – 9:30	**S31**
		9:30 – 10:30	**S38**			9:30 – 10:30	**S33**
10:30 – 11:30	**S33**	10:30 – 11:30	**S39**	10:30 – 11:30	**S42**	10:30 – 11:30	**S38**
11:30 – 12:30	**S34**	11:30 – 12:30	**S34**	11:30 – 12:30	**S34**	11:30 – 12:30	**S34**
12:30 – 13:30	**S35**	12:30 – 13:30	**S35**	12:30 – 13:30	**S35**	12:30 – 13:30	**S35**
14:30 – 15:30	**S37**	14:30 – 15:30	**S33**	14:30 – 15:30	**S37**	14:30 – 15:30	**S42**
15:30 – 16:30	**S38**			15:30 – 16:30	**S38**		
16:30 – 17:30	**S39**	16:30 – 17:30	**S37**	16:30 – 17:30	**S33**	16:30 – 17:30	**S37**
17:30 – 18:30	**S40**	17:30 – 18:30	**S40**	17:30 – 18:30	**S40**	17:30 – 18:30	**S40**
18:30 – 19:30	**S41**	18:30 – 19:30	**S41**	18:30 – 19:30	**S41**	18:30 – 19:30	**S41**
19:45 – 20:45	**S42**	19:45 – 20:45	**S42**	19:45 – 20:45	**S39**	19:45 – 20:45	**S39**
20:45 – 21:45	**S43**	20:45 – 21:45	**S43**	20:45 – 21:45	**S43**	20:45 – 21:45	**S43**

Appendix G

Training Results - Individual Tracking Performance

In this appendix, learning curves are fitted to the roll and pitch tracking errors of individual participants during training. The exponential learning curve model is described by Equation (4-2). The learning curve parameters were calculated using a nonlinear optimization algorithm (Matlab's `fitnlm`) to minimize the summed squared error between the measured human operator data and the learning curve model. Pearson's correlation coefficient ρ was determined to asses the fit of the learning curves. For each learning curve, Pearson's correlation coefficient is provided in the graph legend. However, only for learning curves with $\rho > 0.5$ the correlation between the measured human operator data and the learning curve was deemed sufficient enough to include the learning curve parameters in the group comparison of the tracking error development throughout training in Section 7-3.

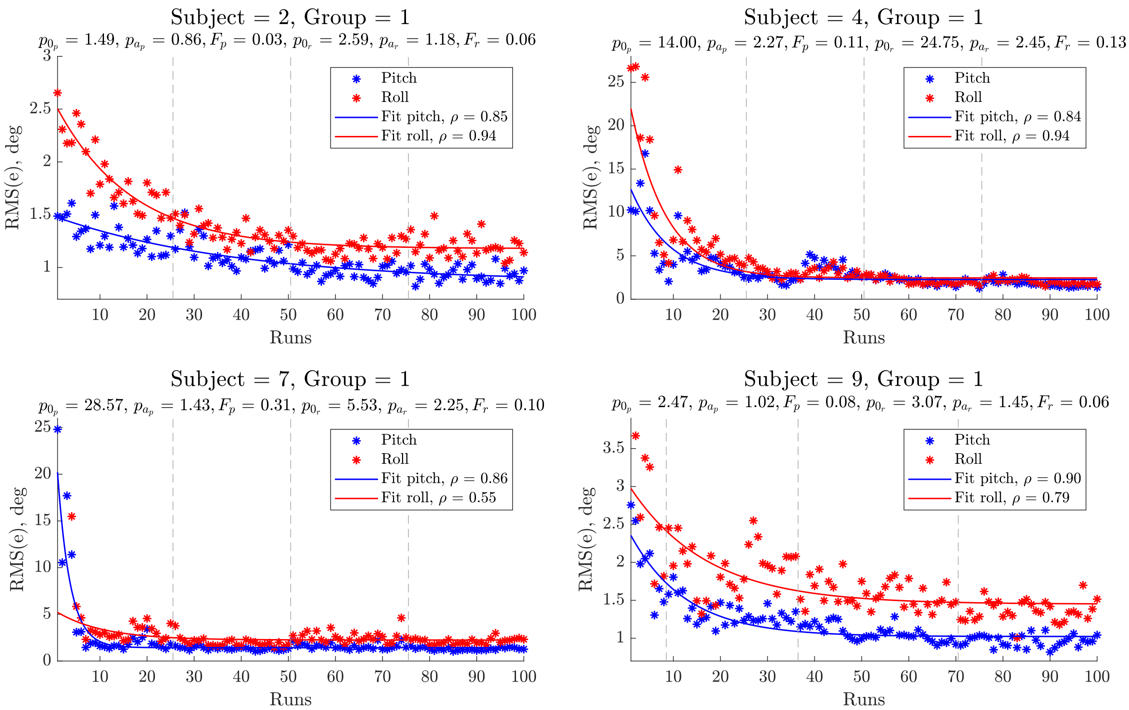


Figure G-1: Individual learning curves RMS(e) for participants of Group 1.

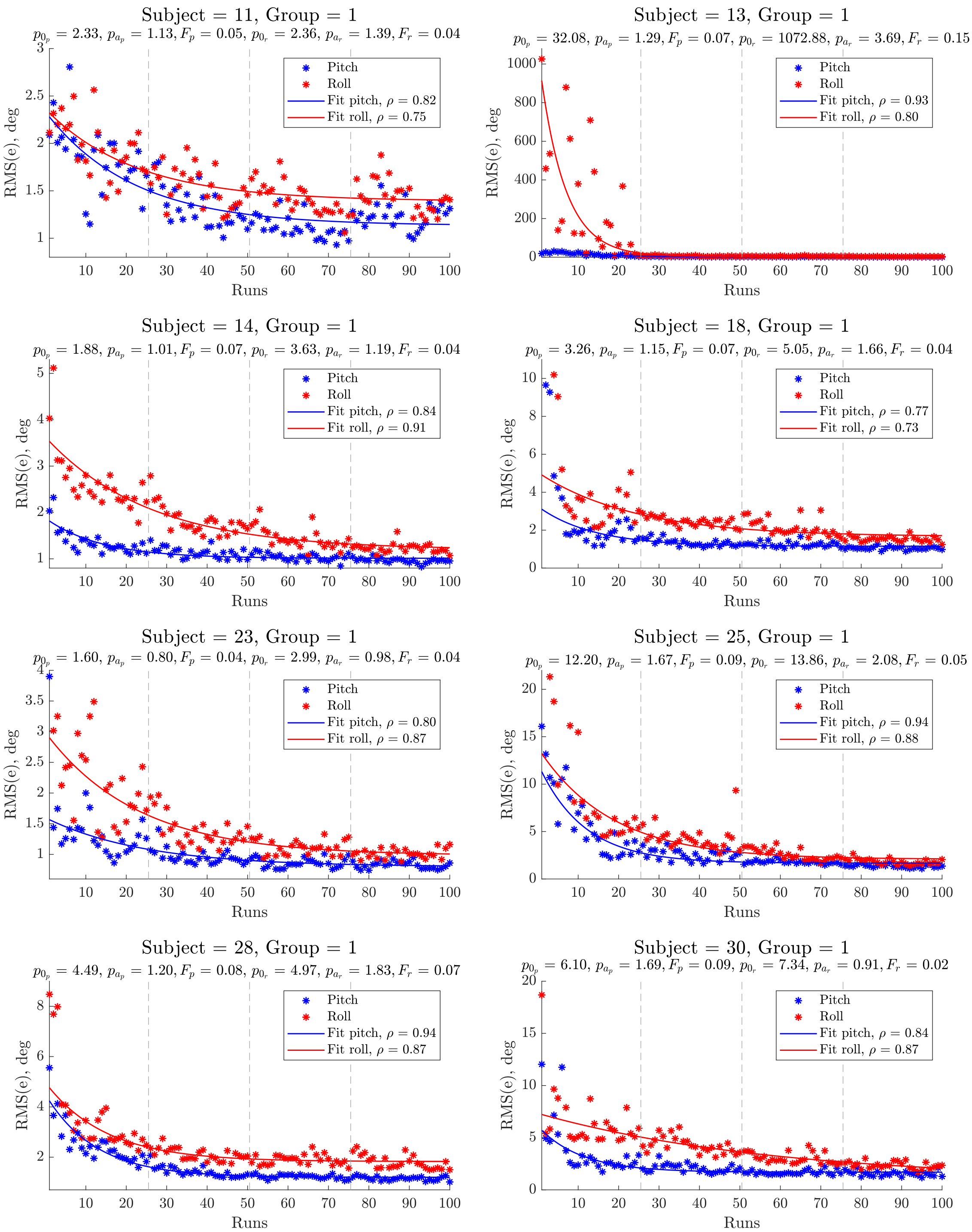


Figure G-1: Individual learning curves RMS(e) for participants of Group 1 (cont.).

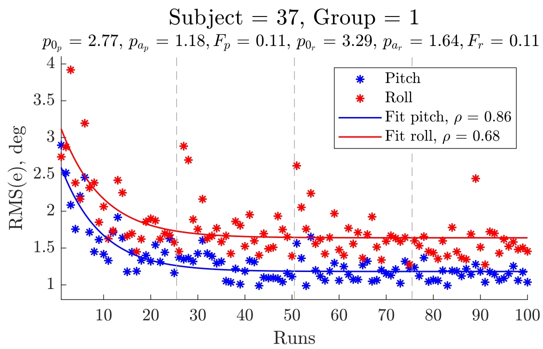


Figure G-1: Individual learning curves RMS(e) for participants of Group 1 (cont.).

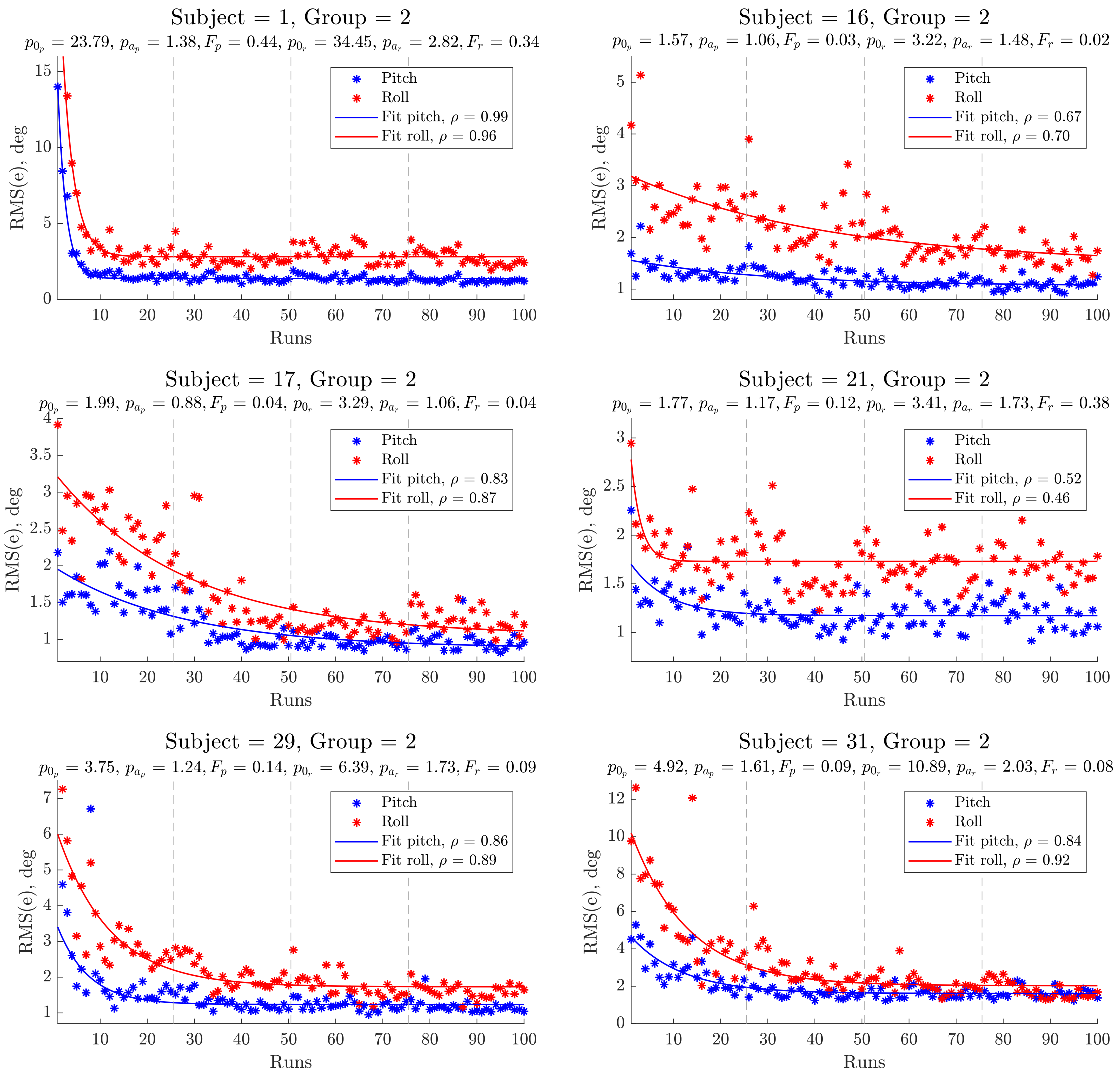


Figure G-2: Individual learning curves RMS(e) for participants of Group 2.

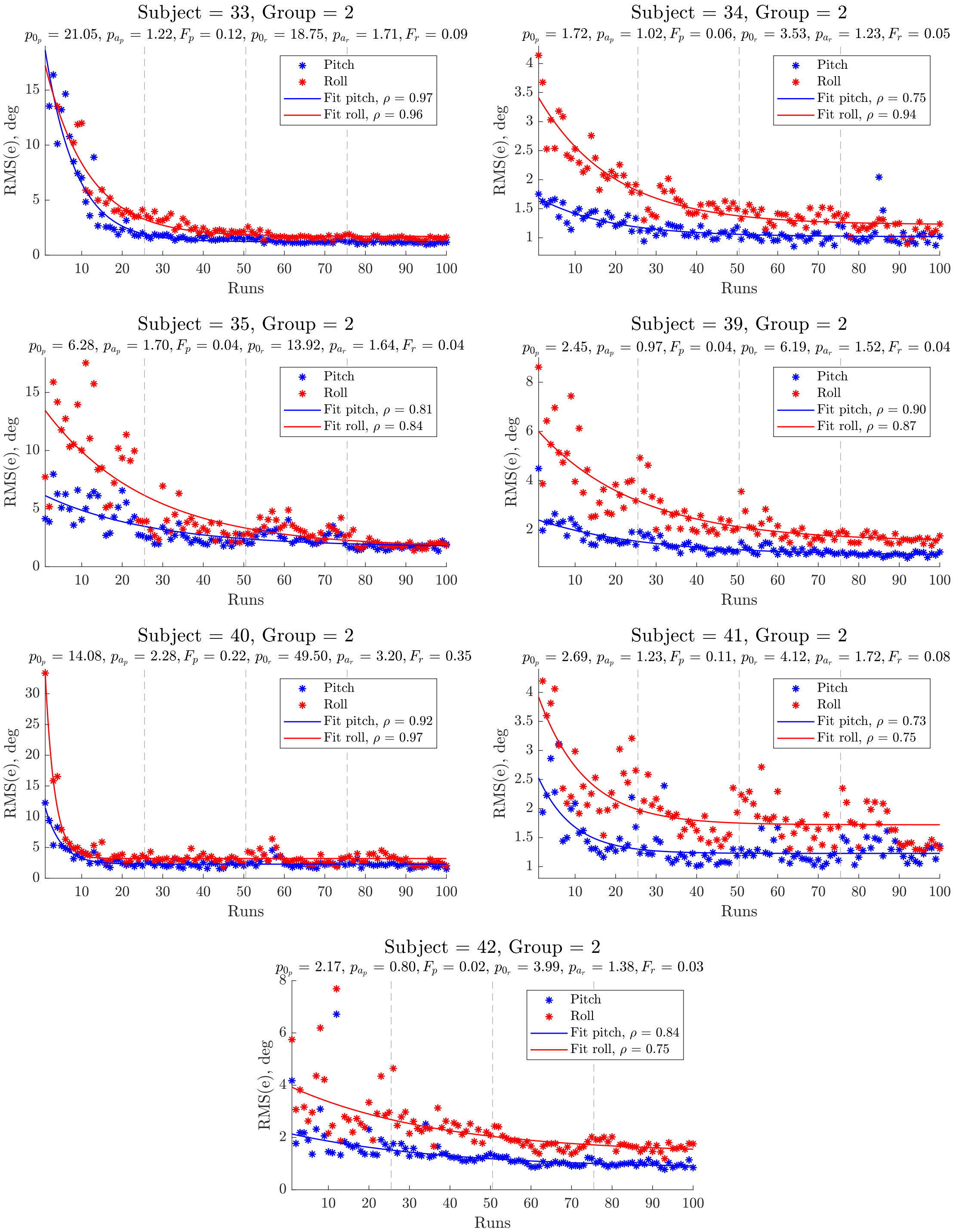


Figure G-2: Individual learning curves RMS(e) for participants of Group 2 (cont.).

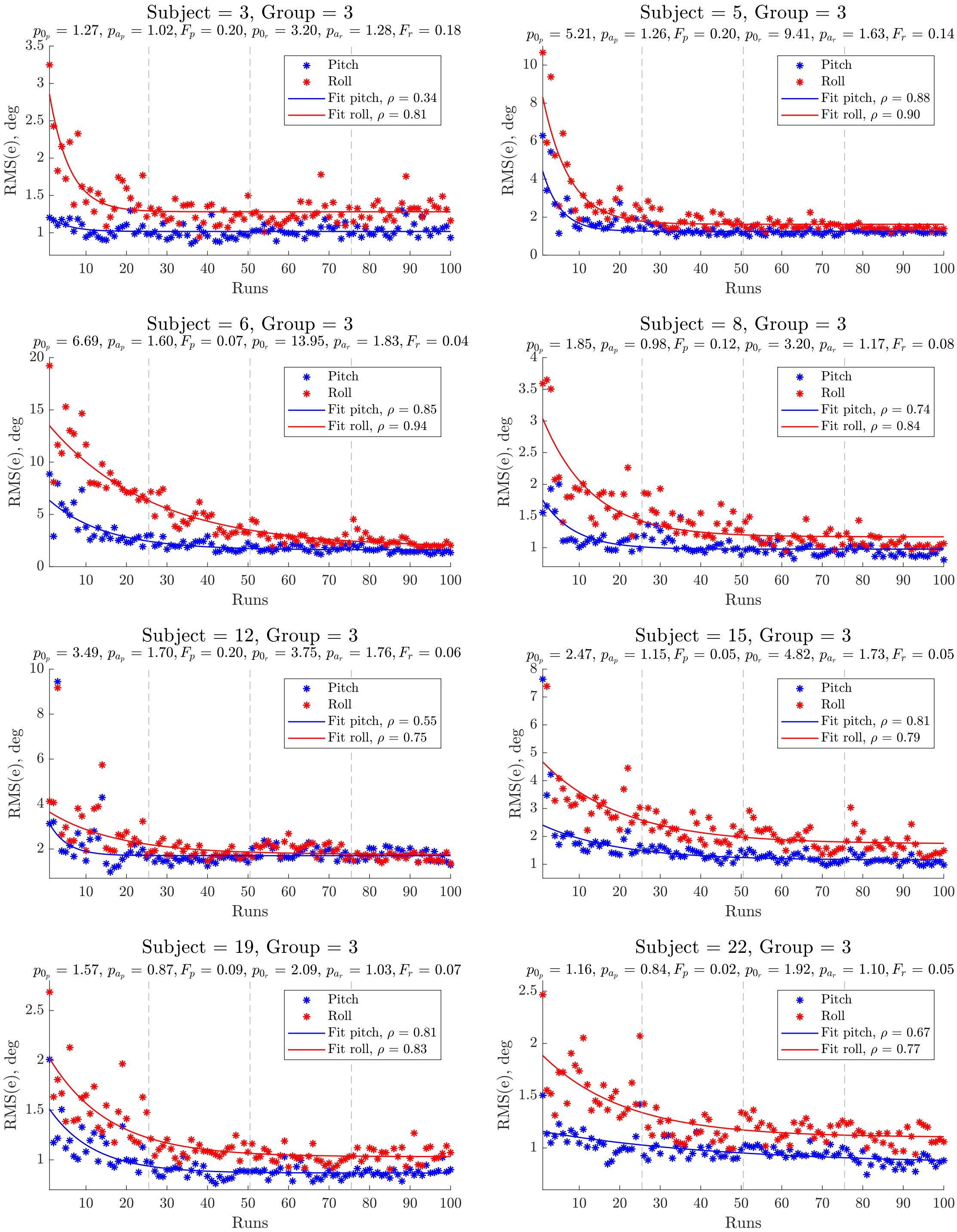


Figure G-3: Individual learning curves RMS(e) for participants of Group 3.

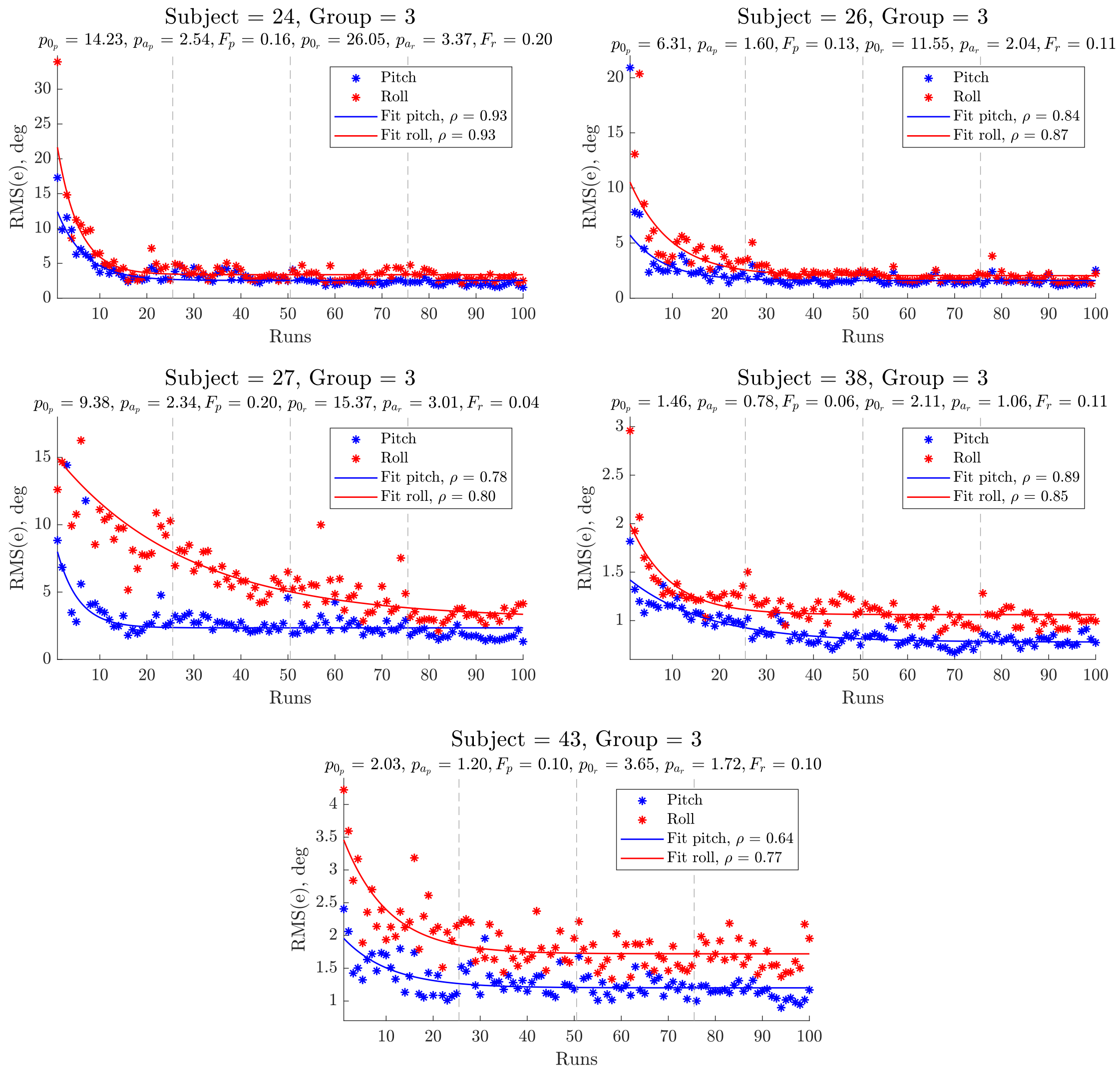


Figure G-3: Individual learning curves RMS(e) for participants of Group 3 (cont.).

Appendix H

Training Effects - Statistical Analysis

This appendix presents the detailed results of the statistical analysis of the training effects on the tracking errors and control inputs of the three retention groups, as was discussed in Chapter 7. The analysis was performed in SPSS 25. To analyze the parameter changes throughout the training phase, the five-run averages of these dependent measures at the start (runs 1 - 5) and at the end (runs 96 - 100) of training were subject to pairwise comparisons. If the data was approximately normally distributed, the dependent t test was performed. However, if one of the compared samples did not sufficiently fit a normal distribution, the nonparametric Wilcoxon signed-rank rest was performed instead of the parametric dependent t test.

H-1 Tracking Performance

Tests of Normality

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
RMSe_p_start_1	,224	13	,074	,807	13	,008
RMSe_p_end_1	,122	13	,200*	,955	13	,671
RMSe_p_start_2	,210	13	,120	,780	13	,004
RMSe_p_end_2	,260	13	,016	,801	13	,007
RMSe_p_start_3	,228	13	,064	,850	13	,029
RMSe_p_end_3	,236	13	,046	,892	13	,105

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Figure H-1: Tests of normality for RMS(e) in pitch averaged over five consecutive runs at the start (runs 1 - 5) and at the end (runs 96 - 100) of training.

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
RMSe_r_start_1	,422	13	,000	,594	13	,000
RMSe_r_end_1	,213	13	,110	,827	13	,014
RMSe_r_start_2	,288	13	,004	,620	13	,000
RMSe_r_end_2	,282	13	,006	,894	13	,109
RMSe_r_start_3	,327	13	,000	,631	13	,000
RMSe_r_end_3	,243	13	,035	,793	13	,006

a. Lilliefors Significance Correction

Figure H-2: Tests of normality for RMS(e) in roll averaged over five consecutive runs at the start (runs 1 - 5) and at the end (runs 96 - 100) of training.

Wilcoxon Signed-Rank Tests

	Null Hypothesis	Test	Sig.	Decision
1	The median of differences between RMSe_p_start_1 and RMSe_p_end_1 equals 0.	Related-Samples Wilcoxon Signed Rank Test	,001	Reject the null hypothesis.

	Null Hypothesis	Test	Sig.	Decision
1	The median of differences between RMSe_p_start_2 and RMSe_p_end_2 equals 0.	Related-Samples Wilcoxon Signed Rank Test	,001	Reject the null hypothesis.

	Null Hypothesis	Test	Sig.	Decision
1	The median of differences between RMSe_p_start_3 and RMSe_p_end_3 equals 0.	Related-Samples Wilcoxon Signed Rank Test	,001	Reject the null hypothesis.

Figure H-3: Wilcoxon signed-rank tests for RMS(e) in pitch averaged over five consecutive runs at the start (runs 1 - 5) and at the end (runs 96 - 100) of training.

	Null Hypothesis	Test	Sig.	Decision
1	The median of differences between RMSe_r_start_1 and RMSe_r_end_1 equals 0.	Related-Samples Wilcoxon Signed Rank Test	,001	Reject the null hypothesis.

	Null Hypothesis	Test	Sig.	Decision
1	The median of differences between RMSe_r_start_2 and RMSe_r_end_2 equals 0.	Related-Samples Wilcoxon Signed Rank Test	,001	Reject the null hypothesis.

	Null Hypothesis	Test	Sig.	Decision
1	The median of differences between RMSe_r_start_3 and RMSe_r_end_3 equals 0.	Related-Samples Wilcoxon Signed Rank Test	,001	Reject the null hypothesis.

Figure H-4: Wilcoxon signed-rank tests for RMS(e) in roll averaged over five consecutive runs at the start (runs 1 - 5) and at the end (runs 96 - 100) of training.

H-2 Control Activity

Tests of Normality

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
RMSu_p_start_1	,188	13	,200*	,926	13	,302
RMSu_p_end_1	,359	13	,000	,646	13	,000
RMSu_p_start_2	,247	13	,029	,768	13	,003
RMSu_p_end_2	,298	13	,002	,822	13	,013
RMSu_p_start_3	,170	13	,200*	,917	13	,229
RMSu_p_end_3	,269	13	,011	,726	13	,001

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Figure H-5: Tests of normality for RMS(u) in pitch averaged over five consecutive runs at the start (runs 1 - 5) and at the end (runs 96 - 100) of training.

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
RMSu_r_start_1	,168	13	,200*	,928	13	,325
RMSu_r_end_1	,245	13	,031	,869	13	,051
RMSu_r_start_2	,240	13	,038	,877	13	,065
RMSu_r_end_2	,357	13	,000	,704	13	,001
RMSu_r_start_3	,148	13	,200*	,938	13	,434
RMSu_r_end_3	,111	13	,200*	,962	13	,790

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Figure H-6: Tests of normality for RMS(u) in roll averaged over five consecutive runs at the start (runs 1 - 5) and at the end (runs 96 - 100) of training.

Wilcoxon Signed-Rank Tests

	Null Hypothesis	Test	Sig.	Decision
1	The median of differences between RMSu_p_start_1 and RMSu_p_end_1 equals 0.	Related-Samples Wilcoxon Signed Rank Test	,019	Reject the null hypothesis.

	Null Hypothesis	Test	Sig.	Decision
1	The median of differences between RMSu_p_start_2 and RMSu_p_end_2 equals 0.	Related-Samples Wilcoxon Signed Rank Test	,196	Retain the null hypothesis.

	Null Hypothesis	Test	Sig.	Decision
1	The median of differences between RMSu_p_start_3 and RMSu_p_end_3 equals 0.	Related-Samples Wilcoxon Signed Rank Test	,345	Retain the null hypothesis.

Figure H-7: Wilcoxon signed-rank tests for RMS(u) in pitch averaged over five consecutive runs at the start (runs 1 - 5) and at the end (runs 96 - 100) of training.

	Null Hypothesis	Test	Sig.	Decision
1	The median of differences between RMSu_r_start_2 and RMSu_r_end_2 equals 0.	Related-Samples Wilcoxon Signed Rank Test	,007	Reject the null hypothesis.

Figure H-8: Wilcoxon signed-rank test for RMS(u) in roll averaged over five consecutive runs at the start (runs 1 - 5) and at the end (runs 96 - 100) of training.

Dependent t test

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
					95% Confidence Interval of the Difference				
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper			
Pair 1	RMSu_r_start_1 - RMSu_r_end_1	3,28363	2,43407	,67509	1,81273	4,75452	4,864	12	,000

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
					95% Confidence Interval of the Difference				
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper			
Pair 1	RMSu_r_start_3 - RMSu_r_end_3	1,39068	1,51364	,41981	,47599	2,30536	3,313	12	,006

Figure H-9: Dependent t tests for RMS(u) in roll averaged over five consecutive runs at the start (runs 1 - 5) and at the end (runs 96 - 100) of training.

Appendix I

Group Division - Statistical Analysis

This appendix presents the detailed results of the statistical analysis of the group division as was discussed in Chapter 7. The analysis was performed in SPSS 25. If the data was approximately normally distributed and met the homogeneity of variances requirement, a parametric test was performed. The parametric test applied was a one-way independent Analysis of Variance (ANOVA). However, if the data was approximately normally distributed, but violated the homogeneity of variances requirement, or if the data did not fit a normal distribution at all, a nonparametric test was performed. The nonparametric test applied was the Kruskal-Wallis Test.

The statistical analysis was performed for the tracking error, the control input, the human operator model parameters, the VAF and the parameters of the tracking error learning curve, all both in pitch and roll.

I-1 Tracking Performance

Tests of Normality

Tests of Normality

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Group	Statistic	df	Sig.	Statistic	df	Sig.
RMSe_p	Group 1	,115	13	,200*	,956	13	,694
	Group 2	,292	13	,003	,802	13	,007
	Group 3	,197	13	,178	,920	13	,248

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Figure I-1: Tests of normality for RMS(e) in pitch averaged over the last ten training runs (runs 91 - 100).

Tests of Normality

	Group	Kolmogorov-Smirnov[a] Statistic	df	Sig.	Shapiro-Wilk Statistic	df	Sig.
RMSe_r	Group 1	,203	13	,145	,844	13	,024
	Group 2	,269	13	,011	,876	13	,062
	Group 3	,283	13	,006	,819	13	,012

a. Lilliefors Significance Correction

Figure I-2: Tests of normality for RMS(e) in roll averaged over the last ten training runs (runs 91 - 100).

Kruskal-Wallis Tests

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of RMSe_p is the same across categories of Group.	Independent-Samples Kruskal-Wallis Test	,961	Retain the null hypothesis.

Figure I-3: Kruskal-Wallis test for RMS(e) in pitch averaged over the last ten training runs (runs 91 - 100).

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of RMSe_r is the same across categories of Group.	Independent-Samples Kruskal-Wallis Test	,720	Retain the null hypothesis.

Figure I-4: Kruskal-Wallis test for RMS(e) in roll averaged over the last ten training runs (runs 91 - 100).

I-2 Control Activity

Tests of Normality

Tests of Normality

	Group	Kolmogorov-Smirnov[a] Statistic	df	Sig.	Shapiro-Wilk Statistic	df	Sig.
RMSu_p	Group 1	,276	13	,008	,659	13	,000
	Group 2	,282	13	,006	,790	13	,005
	Group 3	,270	13	,010	,751	13	,002

a. Lilliefors Significance Correction

Figure I-5: Tests of normality for RMS(u) in pitch averaged over the last ten training runs (runs 91 - 100).

Tests of Normality

	Group	Kolmogorov-Smirnov[a] Statistic	df	Sig.	Shapiro-Wilk Statistic	df	Sig.
RMSu_r	Group 1	,197	13	,178	,888	13	,092
	Group 2	,388	13	,000	,637	13	,000
	Group 3	,161	13	,200*	,958	13	,723

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Figure I-6: Tests of normality for RMS(u) in roll averaged over the last ten training runs (runs 91 - 100).

Kruskal-Wallis Tests

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of RMSu_p is the same across categories of Group.	Independent-Samples Kruskal-Wallis Test	,287	Retain the null hypothesis.

Figure I-7: Kruskal-Wallis test for RMS(u) in pitch averaged over the last ten training runs (runs 91 - 100).

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of RMSu_r is the same across categories of Group.	Independent-Samples Kruskal-Wallis Test	,289	Retain the null hypothesis.

Figure I-8: Kruskal-Wallis test for RMS(u) in roll averaged over the last ten training runs (runs 91 - 100).

I-3 Pilot Gain

Tests of Normality

Tests of Normality

	Group	Kolmogorov-Smirnov[a] Statistic	df	Sig.	Shapiro-Wilk Statistic	df	Sig.
Kp_p	Group 1	,295	13	,003	,655	13	,000
	Group 2	,384	13	,000	,720	13	,001
	Group 3	,104	13	,200*	,954	13	,655

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Figure I-9: Tests of normality for the pilot gain K_p in pitch averaged over the last ten training runs (runs 91 - 100).

Tests of Normality

	Group	Kolmogorov-Smirnov[a] Statistic	df	Sig.	Shapiro-Wilk Statistic	df	Sig.
Kp_r	Group 1	,198	13	,171	,881	13	,074
	Group 2	,327	13	,000	,762	13	,003
	Group 3	,201	13	,155	,902	13	,142

a. Lilliefors Significance Correction

Figure I-10: Tests of normality for the pilot gain K_p in roll averaged over the last ten training runs (runs 91 - 100).

Kruskal-Wallis Tests

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of Kp_p is the same across categories of Group.	Independent-Samples Kruskal-Wallis Test	,330	Retain the null hypothesis.

Figure I-11: Kruskal-Wallis test for the pilot gain K_p in pitch averaged over the last ten training runs (runs 91 - 100).

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of Kp_r is the same across categories of Group.	Independent-Samples Kruskal-Wallis Test	,364	Retain the null hypothesis.

Figure I-12: Kruskal-Wallis test for the pilot gain K_p in roll averaged over the last ten training runs (runs 91 - 100).

I-4 Lead Time Constant

Tests of Normality

Tests of Normality

	Group	Kolmogorov-Smirnov[a] Statistic	df	Sig.	Shapiro-Wilk Statistic	df	Sig.
TL_p	Group 1	,163	13	,200*	,929	13	,331
	Group 2	,154	13	,200*	,955	13	,674
	Group 3	,233	13	,052	,835	13	,018

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Figure I-13: Tests of normality for the lead time constant T_L in pitch averaged over the last ten training runs (runs 91 - 100).

Tests of Normality

	Group	Kolmogorov-Smirnov[a] Statistic	df	Sig.	Shapiro-Wilk Statistic	df	Sig.
TL_r	Group 1	,176	13	,200*	,939	13	,449
	Group 2	,127	13	,200*	,953	13	,637
	Group 3	,203	13	,146	,922	13	,268

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Figure I-14: Tests of normality for the lead time constant T_L in roll averaged over the last ten training runs (runs 91 - 100).

Kruskal-Wallis Test

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of TL_p is the same across categories of Group.	Independent-Samples Kruskal-Wallis Test	,266	Retain the null hypothesis.

Figure I-15: Kruskal-Wallis test for the lead time constant T_L in pitch averaged over the last ten training runs (runs 91 - 100).

Test of Homogeneity of Variances

Test of Homogeneity of Variances

		Levene Statistic	df1	df2	Sig.
TL_r	Based on Mean	1,018	2	36	,372
	Based on Median	,646	2	36	,530
	Based on Median and with adjusted df	,646	2	27,874	,532
	Based on trimmed mean	,984	2	36	,384

Figure I-16: Test of homogeneity of variances for the lead time constant T_L in roll averaged over the last ten training runs (runs 91 - 100).

One-Way ANOVA

ANOVA

TL_r

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	,064	2	,032	,436	,650
Within Groups	2,647	36	,074		
Total	2,711	38			

Figure I-17: One-way ANOVA for the lead time constant T_L in roll averaged over the last ten training runs (runs 91 - 100).

I-5 Response Delay

Tests of Normality

Tests of Normality

	Group	Kolmogorov-Smirnov[a] Statistic	df	Sig.	Shapiro-Wilk Statistic	df	Sig.
te_p	Group 1	,175	13	,200*	,890	13	,098
	Group 2	,148	13	,200*	,952	13	,624
	Group 3	,204	13	,142	,847	13	,026

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Figure I-18: Tests of normality for the response delay τ_e in pitch averaged over the last ten training runs (runs 91 - 100).

Tests of Normality

	Group	Kolmogorov-Smirnov[a] Statistic	df	Sig.	Shapiro-Wilk Statistic	df	Sig.
te_r	Group 1	,117	13	,200*	,985	13	,996
	Group 2	,166	13	,200*	,909	13	,179
	Group 3	,120	13	,200*	,953	13	,639

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Figure I-19: Tests of normality for the response delay τ_e in roll averaged over the last ten training runs (runs 91 - 100).

Kruskal-Wallis Test

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of te_p is the same across categories of Group.	Independent-Samples Kruskal-Wallis Test	,995	Retain the null hypothesis.

Figure I-20: Kruskal-Wallis test for the response delay τ_e in pitch averaged over the last ten training runs (runs 91 - 100).

Test of Homogeneity of Variances

Test of Homogeneity of Variances

		Levene Statistic	df1	df2	Sig.
te_r	Based on Mean	1,188	2	36	,317
	Based on Median	1,140	2	36	,331
	Based on Median and with adjusted df	1,140	2	33,820	,332
	Based on trimmed mean	1,207	2	36	,311

Figure I-21: Test of homogeneity of variances for the response delay τ_e in roll averaged over the last ten training runs (runs 91 - 100).

One-Way ANOVA

ANOVA

te_r

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	,019	2	,009	1,099	,344
Within Groups	,305	36	,008		
Total	,324	38			

Figure I-22: One-way ANOVA for the response delay τ_e in roll averaged over the last ten training runs (runs 91 - 100).

I-6 Neuromuscular Frequency

Tests of Normality

Tests of Normality

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Group	Statistic	df	Sig.	Statistic	df	Sig.
wnm_p	Group 1	,288	13	,004	,831	13	,016
	Group 2	,192	13	,200*	,891	13	,100
	Group 3	,381	13	,000	,579	13	,000

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Figure I-23: Tests of normality for the neuromuscular frequency ω_{nm} in pitch averaged over the last ten training runs (runs 91 - 100).

Tests of Normality

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Group	Statistic	df	Sig.	Statistic	df	Sig.
wnm_r	Group 1	,194	13	,194	,930	13	,338
	Group 2	,148	13	,200*	,919	13	,240
	Group 3	,187	13	,200*	,958	13	,722

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Figure I-24: Tests of normality for the neuromuscular frequency ω_{nm} in roll averaged over the last ten training runs (runs 91 - 100).

Kruskal-Wallis Test

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of wnm_p is the same across categories of Group.	Independent-Samples Kruskal-Wallis Test	,230	Retain the null hypothesis.

Figure I-25: Kruskal-Wallis test for the neuromuscular frequency ω_{nm} in pitch averaged over the last ten training runs (runs 91 - 100).

Test of Homogeneity of Variances

Test of Homogeneity of Variances

		Levene Statistic	df1	df2	Sig.
wnm_r	Based on Mean	,026	2	36	,975
	Based on Median	,018	2	36	,982
	Based on Median and with adjusted df	,018	2	34,941	,982
	Based on trimmed mean	,025	2	36	,975

Figure I-26: Test of homogeneity of variances for the neuromuscular frequency ω_{nm} in roll averaged over the last ten training runs (runs 91 - 100).

One-Way ANOVA

ANOVA

wnm_r

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	3,470	2	1,735	1,301	,285
Within Groups	48,002	36	1,333		
Total	51,471	38			

Figure I-27: One-way ANOVA for the neuromuscular frequency ω_{nm} in roll averaged over the last ten training runs (runs 91 - 100).

I-7 Neuromuscular Damping Ratio

Tests of Normality

Tests of Normality

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Group	Statistic	df	Sig.	Statistic	df	Sig.
dnm_p	Group 1	,515	13	,000	,323	13	,000
	Group 2	,246	13	,031	,778	13	,004
	Group 3	,362	13	,000	,462	13	,000

a. Lilliefors Significance Correction

Figure I-28: Tests of normality for the neuromuscular damping ratio ζ_{nm} in pitch averaged over the last ten training runs (runs 91 - 100).

Tests of Normality

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Group	Statistic	df	Sig.	Statistic	df	Sig.
dnm_r	Group 1	,240	13	,040	,683	13	,000
	Group 2	,173	13	,200*	,889	13	,094
	Group 3	,444	13	,000	,561	13	,000

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Figure I-29: Tests of normality for the neuromuscular damping ratio ζ_{nm} in roll averaged over the last ten training runs (runs 91 - 100).

Kruskal-Wallis Tests

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of dnm_p is the same across categories of Group.	Independent-Samples Kruskal-Wallis Test	,400	Retain the null hypothesis.

Figure I-30: Kruskal-Wallis test for the neuromuscular damping ratio ζ_{nm} in pitch averaged over the last ten training runs (runs 91 - 100).

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of dnm_r is the same across categories of Group.	Independent-Samples Kruskal-Wallis Test	,509	Retain the null hypothesis.

Figure I-31: Kruskal-Wallis test for the neuromuscular damping ratio ζ_{nm} in roll averaged over the last ten training runs (runs 91 - 100).

I-8 Variance Accounted For

Tests of Normality

Tests of Normality

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Group	Statistic	df	Sig.	Statistic	df	Sig.
VAF_p	Group 1	,174	13	,200*	,923	13	,271
	Group 2	,195	13	,187	,838	13	,020
	Group 3	,221	13	,082	,881	13	,074

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Figure I-32: Tests of normality for the Variance Accounted For in pitch averaged over the last ten training runs (runs 91 - 100).

Tests of Normality

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Group	Statistic	df	Sig.	Statistic	df	Sig.
VAF_r	Group 1	,131	13	,200*	,954	13	,666
	Group 2	,263	13	,014	,879	13	,069
	Group 3	,097	13	,200*	,980	13	,978

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Figure I-33: Tests of normality for the Variance Accounted For in roll averaged over the last ten training runs (runs 91 - 100).

Kruskal-Wallis Test

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of VAF_p is the same across categories of Group.	Independent-Samples Kruskal-Wallis Test	,387	Retain the null hypothesis.

Figure I-34: Kruskal-Wallis test for the Variance Accounted For in pitch averaged over the last ten training runs (runs 91 - 100).

Test of Homogeneity of Variances

Test of Homogeneity of Variances

		Levene Statistic	df1	df2	Sig.
VAF_r	Based on Mean	1,014	2	36	,373
	Based on Median	,790	2	36	,462
	Based on Median and with adjusted df	,790	2	33,992	,462
	Based on trimmed mean	,971	2	36	,388

Figure I-35: Test of homogeneity of variances for the Variance Accounted For in roll averaged over the last ten training runs (runs 91 - 100).

One-Way ANOVA

ANOVA

VAF_r

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	97,196	2	48,598	,955	,394
Within Groups	1831,888	36	50,886		
Total	1929,084	38			

Figure I-36: One-way ANOVA for the Variance Accounted For in roll averaged over the last ten training runs (runs 91 - 100).

I-9 Initial Value of Learning Curve Tracking Error

Tests of Normality

Tests of Normality

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Group	Statistic	df	Sig.	Statistic	df	Sig.
p0_p	Group 1	,291	13	,004	,714	13	,001
	Group 2	,295	13	,003	,701	13	,001
	Group 3	,208	12	,158	,831	12	,021

a. Lilliefors Significance Correction

Figure I-37: Tests of normality for the initial value p_0 of the individual RMS(e) learning curves in pitch.

Tests of Normality

	Group	Kolmogorov-Smirnov[a] Statistic	df	Sig.	Shapiro-Wilk Statistic	df	Sig.
p0_r	Group 1	,509	13	,000	,326	13	,000
	Group 2	,262	12	,022	,729	12	,002
	Group 3	,274	13	,009	,795	13	,006

a. Lilliefors Significance Correction

Figure I-38: Tests of normality for the initial value p_0 of the individual RMS(e) learning curves in roll.

Kruskal-Wallis Tests

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of p0_p is the same across categories of Group.	Independent-Samples Kruskal-Wallis Test	,735	Retain the null hypothesis.

Figure I-39: Kruskal-Wallis test for the initial value p_0 of the individual RMS(e) learning curves in pitch.

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of p0_r is the same across categories of Group.	Independent-Samples Kruskal-Wallis Test	,360	Retain the null hypothesis.

Figure I-40: Kruskal-Wallis test for the initial value p_0 of the individual RMS(e) learning curves in roll.

I-10 Asymptotic Value of Learning Curve Tracking Error

Tests of Normality

Tests of Normality

	Group	Kolmogorov-Smirnov[a] Statistic	df	Sig.	Shapiro-Wilk Statistic	df	Sig.
pa_p	Group 1	,201	13	,154	,896	13	,116
	Group 2	,231	13	,055	,886	13	,085
	Group 3	,186	12	,200*	,891	12	,122

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Figure I-41: Tests of normality for the asymptotic value p_a of the individual RMS(e) learning curves in pitch.

Tests of Normality

	Group	Kolmogorov-Smirnov[a] Statistic	df	Sig.	Shapiro-Wilk Statistic	df	Sig.
pa_r	Group 1	,162	13	,200*	,882	13	,076
	Group 2	,290	12	,006	,847	12	,033
	Group 3	,223	13	,075	,841	13	,022

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Figure I-42: Tests of normality for the asymptotic value p_a of the individual RMS(e) learning curves in roll.

Test of Homogeneity of Variances

Test of Homogeneity of Variances

		Levene Statistic	df1	df2	Sig.
pa_p	Based on Mean	1,491	2	35	,239
	Based on Median	,989	2	35	,382
	Based on Median and with adjusted df	,989	2	32,037	,383
	Based on trimmed mean	1,455	2	35	,247

Figure I-43: Test of homogeneity of variances for the asymptotic value p_a of the individual RMS(e) learning curves in pitch.

One-Way ANOVA

ANOVA

pa_p

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	,129	2	,064	,301	,742
Within Groups	7,484	35	,214		
Total	7,613	37			

Figure I-44: One-way ANOVA for the asymptotic value p_a of the individual RMS(e) learning curves in pitch.

Kruskal-Wallis Test

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of pa_r is the same across categories of Group.	Independent-Samples Kruskal-Wallis Test	,955	Retain the null hypothesis.

Figure I-45: Kruskal-Wallis test for the asymptotic value p_a of the individual RMS(e) learning curves in roll.

I-11 Learning Rate of Learning Curve Tracking Error

Tests of Normality

Tests of Normality

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Group	Statistic	df	Sig.	Statistic	df	Sig.
F_p	Group 1	,296	13	,003	,663	13	,000
	Group 2	,248	13	,028	,733	13	,001
	Group 3	,158	12	,200*	,924	12	,322

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Figure I-46: Tests of normality for the learning rate F of the individual RMS(e) learning curves in pitch.

Tests of Normality

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Group	Statistic	df	Sig.	Statistic	df	Sig.
F_r	Group 1	,235	13	,049	,905	13	,159
	Group 2	,381	12	,000	,653	12	,000
	Group 3	,160	13	,200*	,898	13	,127

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Figure I-47: Tests of normality for the learning rate F of the individual RMS(e) learning curves in roll.

Kruskal-Wallis Tests

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of F_p is the same across categories of Group.	Independent-Samples Kruskal-Wallis Test	,510	Retain the null hypothesis.

Figure I-48: Kruskal-Wallis test for the learning rate F of the individual RMS(e) learning curves in pitch.

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of F_r is the same across categories of Group.	Independent-Samples Kruskal-Wallis Test	,495	Retain the null hypothesis.

Figure I-49: Kruskal-Wallis test for the learning rate F of the individual RMS(e) learning curves in roll.

Part IV

MSc Thesis Report Appendices

Appendix J

Individual Experiment Schedules

This appendix presents the individidual experiment schedules of the 38 participants who completed the experiment, including the times at which their training sessions and retention tests took place as well as the length of their retention interval(s).

As also already stated in Appendix F, a few remarks regarding the individual experiment schedules have to be made:

- Although it cannot be seen from the experiment schedule in Table J-1, subject 9 was not able to perform 25 training runs per training session due to problems with the fixed-base simulator in the HMI Lab. In the end, he was able to perform a total of 100 training runs spread out over four consecutive days, but using the following distribution:
  - Day 1: 8 runs
  - Day 2: 28 runs
  - Day 3: 34 runs
  - Day 4: 30 runs
- As can be seen in the experiment schedule in Table J-1, subject 11 was not able to perform his second day of training due to illness. Missing the second training day was compensated for by performing two training sessions on day 4 of training, one at the start of the day and another one at the end of the day. This was considered the best viable option, as this still allowed for some rest time between the third and fourth training session.
- As stated in the paper in Part I and in Chapter 6, it was not possible to completely honor the 24-hour break - which had been found by Kantak and Winstein [175] to be favorable for promoting offline learning - 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. As an indication of the time of day during which a participant performed his or her individual training sessions and retention test(s), color coding was applied in Tables J-1 – J-3, in which different colors are used to define the morning, afternoon and evening sessions as well as the sessions taking place in the transition period around noon and the transition period from afternoon to evening. The exact times for which each color is used are defined below:

  = morning sessions starting between 7:30 and 11:15
  = late morning/early afternoon sessions starting between 11:30 and 13:15
  = afternoon sessions starting between 13:30 and 16:15
  = late afternoon/early evening sessions starting between 16:30 and 18:15
  = evening sessions starting from 18:30 onwards

Table J-1: Individual experiment schedules of group 1.

	Participant	TD 1	TD 2	TD 3	TD 4	RI 1 (days)	RT 1	Total RP (days)
	2	9:30 - 10:30	16:00 - 17:00	14:30 - 15:30	11:30 - 12:30	181	11:30 - 12:30	181
	4	11:30 - 12:30	11:30 - 12:30	13:30 - 14:30	10:30 - 11:30	183	13:00 - 14:00	183
	7	14:30 - 15:30	8:30 - 9:30	8:30 - 9:30	8:30 - 9:30	181	12:30 - 13:30	181
	9	17:00 - 18:00	17:00 - 18:00	17:00 - 18:00	17:00 - 18:00	182	16:00 - 17:00	182
	11	8:30 - 9:30		17:00 - 18:00	8:30 - 9:30/ 17:00 - 18:00	181	18:00 - 19:00	181
	13	11:30 - 12:30	16:00 - 17:00	7:30 - 8:30	9:30 - 10:30	182	19:00 - 20:00	182
	14	12:30 - 13:30	12:30 - 13:30	12:30 - 13:30	12:30 - 13:30	181	11:30 - 12:30	181
	18	17:00 - 18:00	17:00 - 18:00	16:00 - 17:00	10:30 - 11:30	187	11:00 - 12:00	187
	23	13:30 - 14:30	10:30 - 11:30	10:30 - 11:30	10:30 - 11:30	181	14:30 - 15:30	181
	25	16:00 - 17:00	12:30 - 13:30	13:30 - 14:30	11:30 - 12:30	181	18:00 - 19:00	181
	28	18:30 - 19:30	18:30 - 19:30	18:30 - 19:30	18:30 - 19:30	181	17:45 - 18:45	181
	30	20:45 - 21:45	20:45 - 21:45	20:45 - 21:45	20:45 - 21:45	182	19:30 - 20:30	182
	37	14:30 - 15:30	16:30 - 17:30	14:30 - 15:30	16:30 - 17:30	182	12:30 - 13:30	182
Mean						**181.9**		**181.9**
STD						**1.6**		**1.6**
Ideal						*182*		*182*

Legend:

RI	=	Retention Interval	STD	=	Standard Deviation
RP	=	Retention Period	TD	=	Training Day
RT	=	Retention Test			

Table J-2: Individual experiment schedules of group 2.

	Participant	TD 1	TD 2	TD 3	TD 4	RI 1 (days)	RT 1	RI 2 (days)	RT 2	Total RP (days)
	1	8:30 - 9:30	14:30 - 15:30	11:30 - 12:30	15:00 - 16:00	91	10:00 - 10:30	92	8:30 - 9:30	183
	16	15:00 - 16:00	10:30 - 11:30	15:00 - 16:00	15:00 - 16:00	90	15:00 - 15:30	91	15:00 - 16:00	181
	17	16:00 - 17:00	15:00 - 16:00	11:30 - 12:30	16:00 - 17:00	91	10:00 - 10:30	90	9:00 - 10:00	181
	21	11:30 - 12:30	11:30 - 12:30	11:30 - 12:30	12:30 - 13:30	90	14:00 - 14:30	92	15:00 - 16:00	182
	29	19:45 - 20:45	19:45 - 20:45	19:45 - 20:45	19:45 - 20:45	90	18:15 - 18:45	93	19:00 - 20:00	183
	31	8:30 - 9:30	8:30 - 9:30	7:30 - 8:30	8:30 - 9:30	89	9:30 - 10:00	93	7:30 - 8:30	182
	33	10:30 - 11:30	14:30 - 15:30	16:30 - 17:30	9:30 - 10:30	90	12:45 - 13:15	96	17:45 - 18:45	186
	34	11:30 - 12:30	11:30 - 12:30	11:30 - 12:30	11:30 - 12:30	89	15:20 - 15:50	92	9:30 - 10:30	181
	35	12:30 - 13:30	12:30 - 13:30	12:30 - 13:30	12:30 - 13:30	89	14:30 - 15:00	96	12:30 - 13:30	185
	39	16:30 - 17:30	10:30 - 11:30	19:45 - 20:45	19:45 - 20:45	89	10:00 - 10:30	93	20:00 - 21:00	182
	40	17:30 - 18:30	17:30 - 18:30	17:30 - 18:30	17:30 - 18:30	90	17:30 - 18:00	91	10:30 - 11:30	181
	41	18:30 - 19:30	18:30 - 19:30	18:30 - 19:30	18:30 - 19:30	90	18:00 - 18:30	92	19:00 - 20:00	182
	42	19:45 - 20:45	19:45 - 20:45	10:30 - 11:30	14:30 - 15:30	90	13:45 - 14:15	92	21:00 - 22:00	182
Mean						**89.8**		**92.5**		**182.4**
STD						**0.7**		**1.7**		**1.5**
Ideal						*91*		*91*		*182*

Legend:

RI = Retention Interval
RP = Retention Period
RT = Retention Test
STD = Standard Deviation
TD = Training Day

Table J-3: Individual experiment schedules of group 3.

	Participant	TD 1	TD 2	TD 3	TD 4	RI 1 (days)	RT 1	RI 2 (days)	RT 2	RI 3 (days)	RT 3	Total RP (days)
	3	10:30 - 11:30	10:30 - 11:30	10:30 - 11:30	7:30 - 8:30	55	11:10 - 11:40	63	9:00 - 9:30	63	13:45 - 14:45	181
	5	12:30 - 13:30	12:30 - 13:30	12:30 - 13:30	12:30 - 13:30	55	10:00 - 10:30	64	10:00 - 10:30	64	9:30 - 10:30	183
	6	13:30 - 14:30	13:30 - 14:30	9:30 - 10:30	13:30 - 14:30	60	9:00 - 9:30	63	13:30 - 14:00	67	13:30 - 14:30	190
	8	15:45 - 16:45	9:30 - 10:30	15:30 - 16:30	9:30 - 10:30	57	18:15 - 18:45	62	8:00 - 8:30	68	8:00 - 9:00	187
	12	9:30 - 10:30	11:30 - 12:30	10:30 - 11:30	13:30 - 14:30	55	17:15 - 17:45	64	16:00 - 16:30	63	11:00 - 12:00	182
	19	9:30 - 10:30	16:00 - 17:00	9:30 - 10:30	9:30 - 10:30	60	12:00 - 12:30	59	10:00 - 10:30	63	9:00 - 10:00	182
	22	12:30 - 13:30	9:30 - 10:30	14:30 - 15:30	16:00 - 17:00	61	10:00 - 10:30	63	10:45 - 11:15	58	10:45 - 11:45	182
	24	17:00 - 18:00	14:30 - 15:30	15:45 - 16:45	14:30 - 15:30	61	12:30 - 13:00	62	14:00 - 14:30	58	12:30 - 13:30	181
	26	13:30 - 14:30	12:30 - 13:30	13:30 - 14:30	17:15 - 18:15	61	13:00 - 13:30	59	12:45 - 13:15	61	13:30 - 14:30	181
	27	17:30 - 18:30	17:30 - 18:30	17:30 - 18:30	17:30 - 18:30	53	9:30 - 10:00	64	13:00 - 13:30	65	18:30 - 19:30	182
	38	15:30 - 16:30	9:30 - 10:30	15:30 - 16:30	10:30 - 11:30	60	10:00 - 10:30	62	12:00 - 12:30	59	15:30 - 16:30	181
	43	20:45 - 21:45	20:45 - 21:45	20:45 - 21:45	20:45 - 21:45	60	17:30 - 18:00	59	17:30 - 18:00	63	17:30 - 18:30	182
Mean						**58.2**		**62.0**		**62.7**		**182.8**
STD						**2.8**		**1.9**		**3.1**		**2.7**
Ideal						*60.7*		*60.7*		*60.7*		*182*

Legend:

RI = Retention Interval
RP = Retention Period
RT = Retention Test
STD = Standard Deviation
TD = Training Day

Appendix K

Experiment Questionnaire Answers

This appendix presents the answers to the experiment questionnaire provided in Appendix E. As mentioned in Appendix E, the questionnaire was meant to gain insight into participants' activities that could potentially influence their initial learning as well as their retention performance. Two questions regarding participants' car driving experience were later added to the questions listed in Appendix E. These two questions are as follows: (1) Are you in possession of a driver's license, and if so, when did you obtain it (month/year)?, and (2) If you are in possession of a driver's license, how many kilometers do you drive on average each year? Additionally, participants were asked about the study programme they were following, resulting in another three additional questions: (1) What study programme are you following?, (2) If you are studying Aerospace Engineering, from which year of the curriculum do you follow most of your courses at the moment?, and (3) If you are pursuing a MSc in Aerospace Engineering, are you specializing in Control & Simulation (Y/N)? Note that the participants' age and study year provided in this appendix were the participants' age and study year at the time of the training phase of the experiment. The answers to the additional questions mentioned above are also discussed in this appendix. Finally, at every retention test participants were asked whether they had been involved in any activities during the retention interval that could either positively or negatively affect their retention performance. Participants' answers given during the retention tests are discussed below as well. In the end, the answers to all these questions were used to investigate whether demographic factors could (partially) explain the retention results of the experiment.

Participant Age and Gender

Table K-1: Participants' age and gender (F = Female, M = Male).

	Group 1			Group 2			Group 3		
	Participant	Age (years)	Gender	Participant	Age (years)	Gender	Participant	Age (years)	Gender
	2	21	M	1	20	M	3	19	M
	4	20	M	16	32	M	5	21	F
	7	19	M	17	19	M	6	26	F
	9	23	M	21	20	M	8	20	M
	11	20	M	29	19	M	12	20	M
	13	23	F	31	18	M	19	22	M
	14	18	M	33	18	F	22	22	M
	18	19	F	34	27	M	24	20	F
	23	22	M	35	20	F	26	26	M
	25	21	F	39	19	M	27	20	F
	28	25	M	40	19	M	38	23	M
	30	18	F	41	18	M	43	23	M
	37	19	M	42	21	M			
Mean		20.6			20.8			21.8	
STD		2.1			3.9			2.2	

Participants' Studies

Table K-2: Participants' study programs and progress (Gr = Graduated).

Group 1				Group 2				Group 3			
Participant	Studies	AE year	C&S	Participant	Studies	AE year	C&S	Participant	Studies	AE year	C&S
2	AE	1	No	1	AE	2	No	3	AE	2	No
4	AE	1	No	16	AE	5	No	5	AE	3	No
7	AE	1	No	17	AE	1	No	6	AE	5	Yes
9	AE	5	Yes	21	AE	2	No	8	ME	—	—
11	AE	1	No	29	AE	1	No	12	AE	2	No
13	IDE	—	—	31	AE	1	No	19	AE	5	No
14	AE	1	No	33	Arch	—	—	22	AE	4	No
18	LST	—	—	34	AE	4	Yes	24	ME	—	—
23	BME and S&C	—	—	35	AP	—	—	26	ME	—	—
25	TPM	—	—	39	AE	1	No	27	AE	1	No
28	AE	Gr	Yes	40	AE	1	No	38	AE	5	Yes
30	AE	1	No	41	AE	1	No	43	ME	—	—
37	LST	—	—	42	AE	1	No				

Legend:	AE	=	Aerospace Engineering	IDE	=	Industrial Design Engineering
	AP	=	Applied Physics	LST	=	Life, Science & Technology
	Arch	=	Architecture	ME	=	Mechanical Engineering
	BME	=	Biomechanical Engineering	S&C	=	Systems & Control
	C&S	=	Control & Simulation	TPM	=	Technology, Policy & Management

Experiment Briefing

All participants answered the question about whether they had read the experiment briefing before the first training session of the experiment with a 'yes'.

Participants' Car Driving Experience

Table K-3: Participants' car driving experience (DL = Driver's License).

Group 1			Group 2			Group 3		
Participant	DL obtained	Estimated km/year	Participant	DL obtained	Estimated km/year	Participant	DL obtained	Estimated km/year
2	Feb. 2015	14,000	1	—	—	3	Jan. 2016	3,500
4	—	—	16	May 2003	200	5	—	—
7	—	—	17	May 2016	600	6	May 2010	3,000
9	Nov. 2012	2,500	21	Jan. 2016	500	8	Mar. 2015	10,000
11	Jan. 2016	8,000	29	—	—	12	Aug. 2016	1,500
13	Nov. 2015	850	31	—	—	19	July 2017	3,500
14	—	—	33	—	—	22	Sept. 2013	5,000
18	June 2017	2,000	34	July 2008	2,000	24	—	—
23	Sept. 2013	250	35	Aug. 2016	1,750	26	June 2010	5,000
25	Nov. 2013	400	39	May 2017	2,000	27	—	—
28	Sept. 2010	1,500	40	Nov. 2016	2,000	38	May 2013	750
30	Sept. 2017	500	41	—	—	43	June 2013	100
37	June 2017	1,000	42	June 2015	1,500			

Participants' Gaming Experience

Table K-4: Participants' gaming frequency.

Do you (currently) play any video games in which any kind of tracking skill is used?					
Group 1		Group 2		Group 3	
Participant		Participant		Participant	
2	No	1	No	3	Only a few times per year
4	Yes, on a weekly basis	16	No	5	No
7	No	17	Yes, on a weekly basis	6	No
9	Only a few times per year	21	No	8	Only a few times per year
11	Yes, on a monthly basis	29	No	12	No
13	Yes, on a monthly basis	31	Only a few times per year	19	Yes, on a weekly basis
14	Yes, on a weekly basis	33	Only a few times per year	22	Yes, on a weekly basis
18	No	34	Yes, on a weekly basis	24	Yes, on a monthly basis
23	Yes, on a(n) (almost) daily basis	35	No	26	No
25	No	39	No	27	No
28	No	40	Yes, on a monthly basis	38	Yes, on a monthly basis
30	No	41	Yes, on a(n) (almost) daily basis	43	Yes, on a weekly basis
37	Yes, on a(n) (almost) daily basis	42	No		

Table K-5: Participants' gaming experience in years.

Have you been playing video games which require any kind of tracking skill for a long time already or have you played these kinds of video games in the past?

Group 1		**Group 2**		**Group 3**	
Participant		Participant		Participant	
2	Yes, I have played these kinds of video games for about 2 years on a regular basis, but I have quit playing 1 year ago.	1	No	3	Yes, I am still playing these kinds of video games a few times per year and have been for about 6 years now.
4	Yes, I am still playing these kinds of video games and have been for about 6 years now.	16	Yes, I have played these kinds of video games for about 5 years on a regular basis, but I have quit playing 10 years ago.	5	Yes, I have played these kinds of video games for about 6 years on a regular basis, but I have quit playing 4 years ago.
7	No	17	Yes, I am still playing these kinds of video games and have been for about 1 year now.	6	No
9	Yes, I am still playing these kinds of video games a few times per year and have been for about 10 years now.	21	No	8	Yes, I am still playing these kinds of video games a few times per year and have been for about 8 years now.
11	Yes, I am still playing these kinds of video games and have been for about 3 years now.	29	No	12	No
13	Yes, I am still playing these kinds of video games and have been for about half a year now.	31	Yes, I have played these kinds of video games for about 2 years on a regular basis, but I have quit playing regularly 2 years ago. Now I only play a few times per year.	19	Yes, I have played these kinds of video games for about 5 years on a daily basis, but I have quit playing daily 7 years ago. Now I play them on a weekly basis.
14	Yes, I am still playing these kinds of video games and have been for about 6 years now.	33	Yes, I have played these kinds of video games for about 8 years on a regular basis, but I have quit playing regularly 6 year ago. Now I only play a few times per year.	22	Yes, I am still playing these kinds of video games and have been for about 10 years now.
18	No	34	Yes, I am still playing these kinds of video games and have been for about 12 years now.	24	Yes, I am still playing these kinds of video games and have been for about 2 years now.
23	Yes, I am still playing these kinds of video games and have been for about 10 years now.	35	No	26	Yes, I have played these kinds of video games for about 8 years on a regular basis, but I have quit playing 4 year ago.
25	No	39	No	27	No
28	Yes, I have played these kinds of video games for about 3 years on a regular basis, but I have quit playing 4 years ago.	40	Yes, I am still playing these kinds of video games and have been for about 6 years now.	38	Yes, I am still playing these kinds of video games and have been for about 6 years now.
30	No	41	Yes, I am still playing these kinds of video games and have been for about 5 years now.	43	Yes, I am still playing these kinds of video games and have been for about 10 years now.
37	Yes, I am still playing these kinds of video games and have been for about 7 years now.	42	Yes, I have played these kinds of video games for about 10 years on a regular basis, but I have quit playing 1 year ago.		

Participants' Non-Gaming and Non-Driving Related Activities

Table K-6: Participants' non-gaming and non-driving related activities influencing tracking skills.

Are there any other activities / things you do / have done that you believe influence / have influenced your ability to perform tracking tasks?					
Group 1		**Group 2**		**Group 3**	
Participant		Participant		Participant	
2	I have played many car racing simulator games in the past.	1	Perhaps sailing as a hobby of mine might have influenced it.	3	I have been playing classical piano for a long time already.
4	Cycling	16	I have played a lot of baseball in the past.	5	Riding my bike and I have played Mario Kart during high school.
7	—	17	—	6	—
9	I sometimes play Pokemon on the Nintendo DS.	21	—	8	I have played volleyball for 13 years. I also used to play the piano quite often by teaching myself with a phone app, but I have never had any piano lessons.
11	—	29	I play guitar.	12	—
13	—	31	I play piano.	19	I have played some flight simulator games in the past.
14	I have played flight simulator games in the past with a computer keyboard.	33	—	22	I play shooter games and car simulator games on the computer.
18	—	34	I am used to drive a car in strong winds and in snow. I bike a lot, I longboard and I skydive.	24	Driving lessons
23	I do combat sports.	35	—	26	I play racket sports.
25	I play piano.	39	I have been sailing for about 9 years.	27	—
28	—	40	I have been driving a car daily for almost two years.	38	I have been a goalie in soccer in high school.
30	—	41	—	43	—
37	I row and play field hockey.	42	I do martial arts.		

Table K-7: Participants' additional remarks worth mentioning.

Any other remarks you would like to make or you think might be worth mentioning for this research?					
Group 1		**Group 2**		**Group 3**	
Participant		Participant		Participant	
2	—	1	—	3	—
4	—	16	—	5	—
7	—	17	—	6	—
9	—	21	—	8	I work a lot with drawing programs on my computer, which could perhaps influence my motor skills.
11	—	29	—	12	—
13	I was born prematurely, which has influenced my fine motor skills.	31	—	19	—
14	—	33	—	22	—
18	—	34	—	24	—
23	—	35	—	26	—
25	—	39	I write with my left hand, but I do everything else with my right hand.	27	—
28	—	40	—	38	My strategy is to make smaller amplitude inputs, but with a higher frequency.
30	—	41	I have never used a joystick before.	43	—
37	—	42	—		

Retention Interval Activities/Incidents

Table K-8: Activities and incidents during the retention intervals of groups 1 and 2.

Have you done anything or has anything happened during the retention interval which might have influenced your retention performance?				
Group 1		**Group 2**		
Participant	RI 1	Participant	RI 1	RI 2
2	—	1	—	—
4	—	16	—	—
7	—	17	—	—
9	—	21	—	—
11	—	29	—	—
13	—	31	—	—
14	—	33	—	—
18	—	34	—	—
23	—	35	—	—
25	—	39	—	No activities during the retention interval, but I felt very tired during the retention test.
28	—	40	—	—
30	—	41	—	No activities during the retention interval, but I had difficulty focusing on the screen during the retention test.
37	—	42	—	—

Table K-9: Activities and incidents during the retention intervals of group 3.

Have you done anything or has anything happened during the retention interval which might have influenced your retention performance?			
Group 3			
Participant	RI 1	RI 2	RI 3
3	—	—	—
5	—	—	—
6	—	—	—
8	I have been looking at a screen all day before the retention test.	—	I feel like I suffer from sleep deprivation.
12	—	—	—
19	—	—	—
22	No activities during the retention interval, but I felt that I had to get accustomed again during the first few runs of the retention test.	I have suffered a concussion about 7 weeks before the retention test, but now I feel well again.	—
24	—	—	—
26	—	I have driven about 1800 km in 8 consecutive days during this retention interval.	—
27	—	—	—
38	—	—	—
43	—	—	—

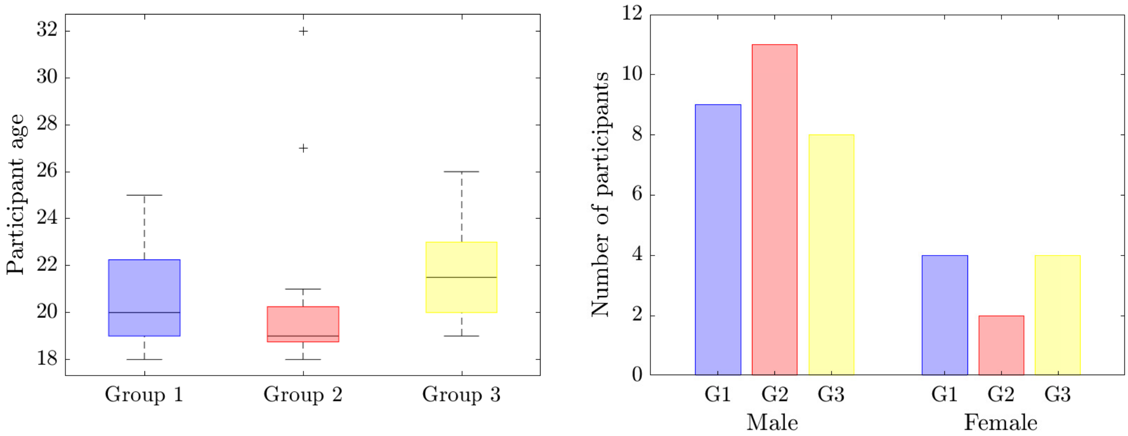
Appendix L

Group Demographics

This appendix presents the demographics of the three different groups used in this study. These demographics were analyzed to investigate whether there could be a demographic cause for the curious retention results. In the 6-month retention test group 1 performed very similar to group 3 and group 2 performed the worst, despite the fact that group 1 did not have any ‘practice’ opportunities in the past six months, group 2 had one ‘practice’ opportunity and group 3 had two. However, the demographic analysis results did not point out a demographic factor that could potentially have caused these curious retention results.

In figures where a demographic factor is plotted against retention performance, $\bar{\Delta}\mathrm{RMS}(e)$ is the instantaneous change in tracking performance between end of training and the start of the 6-month retention test. This instantaneous change was calculated by subtracting $\mathrm{RMS}(e)$ at the end of training from $\mathrm{RMS}(e)$ at the start of the 6-month retention test and subsequently dividing by $\mathrm{RMS}(e)$ at the end of training, as also explained in the paper in Part I. The instantaneous change in tracking performance is expressed as a nondimensional number to facilitate an easier comparison between individuals.

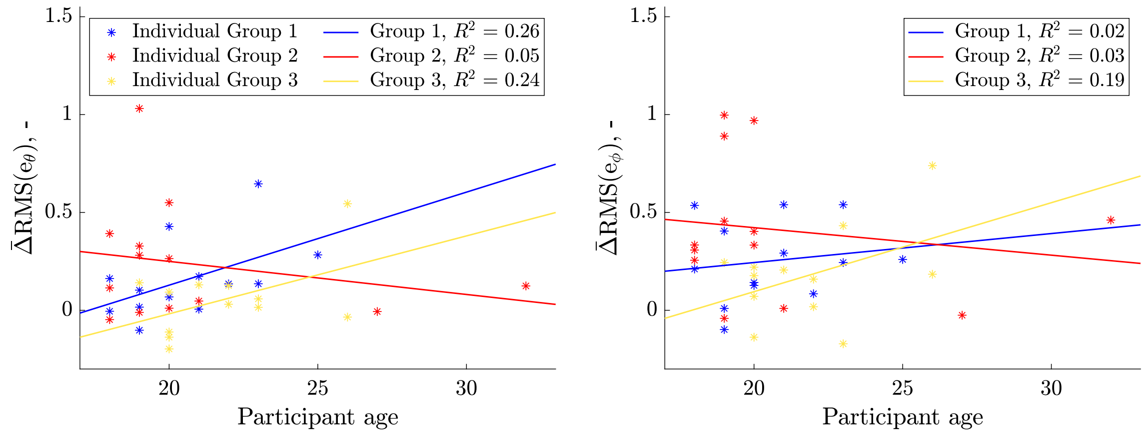
Participant Age and Gender



(a) Participant age

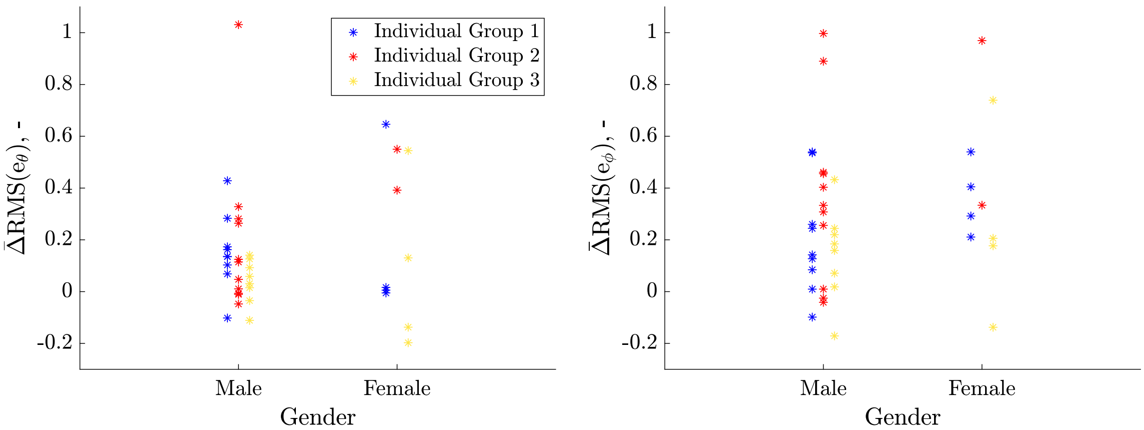
(b) Participant gender

Figure L-1: Participants’ age and gender per experiment group (G = Group).



(a) Pitch θ **(b)** Roll ϕ

Figure L-2: Correlation $\bar{\Delta}$RMS(e) and participant age.

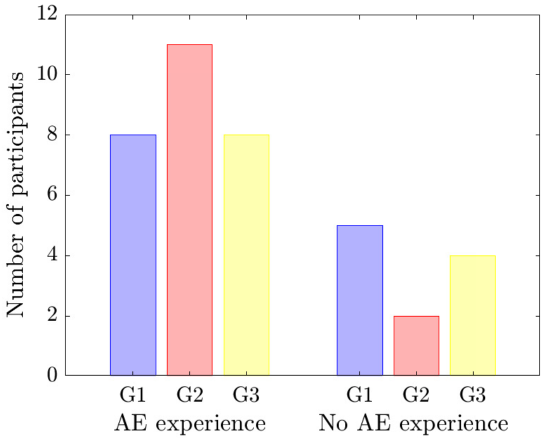


(a) Pitch θ **(b)** Roll ϕ

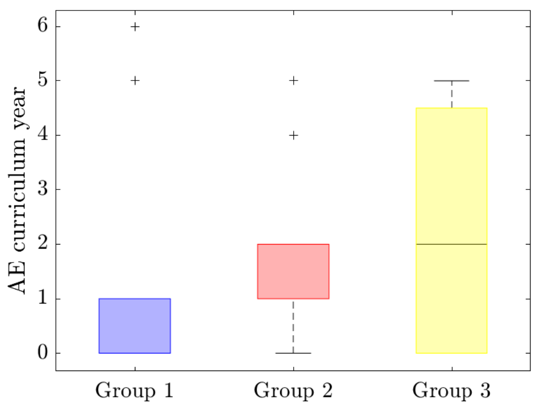
Figure L-3: $\bar{\Delta}$RMS(e) vs. participant gender.

Participants' Aerospace Engineering Experience

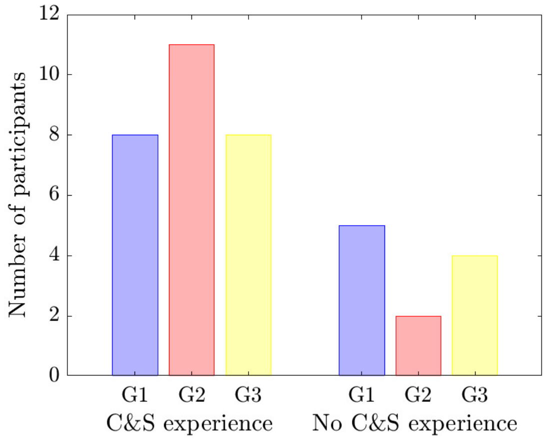
Participants' aerospace engineering experience is expressed in two different manners: (1) as whether or not the participant is pursuing or has obtained a BSc and/or MSc degree in Aerospace Engineering, and (2) as the Aerospace Engineering curriculum year from which the participant is following the most courses at the time of training. If a participant is not pursuing or has not obtained a BSc or MSc degree in Aerospace Engineering, the curriculum year is quantified as 0. If a participant has already obtained his/her MSc degree in Aerospace Engineering, the curriculum year is quantified as 6 in order to distinguish this participant from participants who are in the fifth year of the curriculum. Additionally, participants' control & simulation experience/knowledge is analyzed as whether or not the participant is pursuing or has obtained a MSc degree in Aerospace Engineering with a specialization in Control & Simulation.



(a) Aerospace Engineering experience

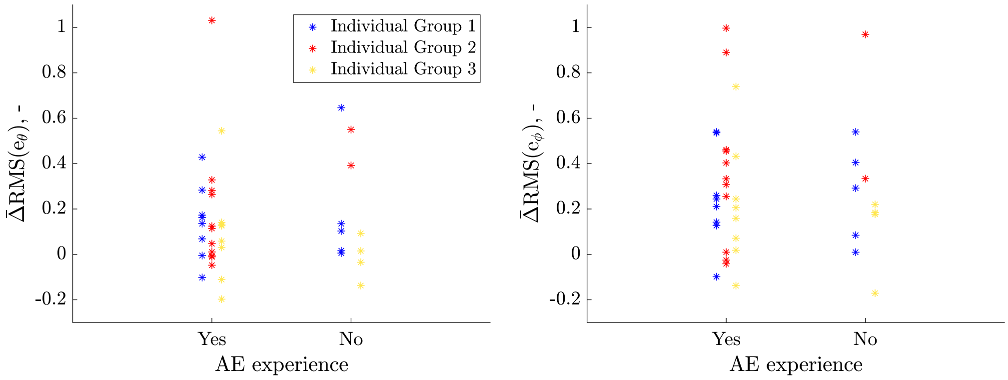


(b) Aerospace Engineering curriculum year



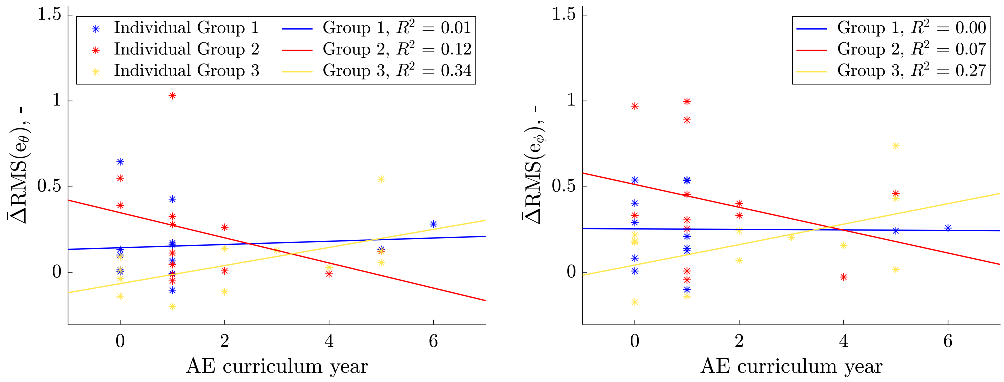
(c) Control & Simulation experience

Figure L-4: Participants' Aerospace Engineering and Control & Simulation experience (AE = Aerospace Engineering, C&S = Control & Simulation, G = Group).



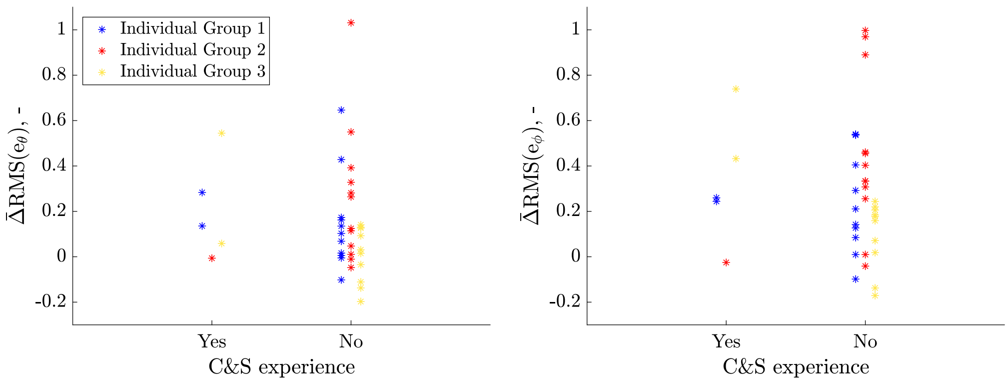
(a) Pitch θ **(b)** Roll ϕ

Figure L-5: $\bar{\Delta}$RMS(e) vs. Aerospace Engineering experience (AE = Aerospace Engineering).



(a) Pitch θ **(b)** Roll ϕ

Figure L-6: Correlation $\bar{\Delta}$RMS(e) and Aerospace Engineering curriculum year (AE = Aerospace Engineering).

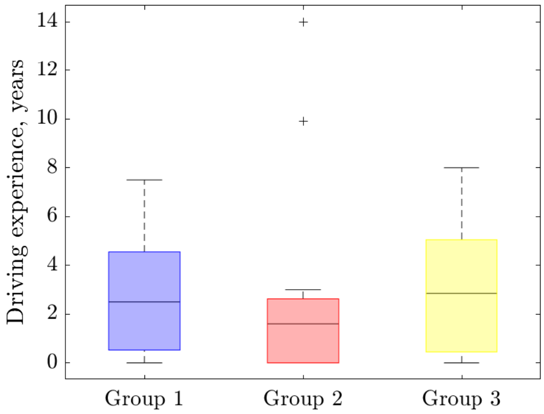


(a) Pitch θ **(b)** Roll ϕ

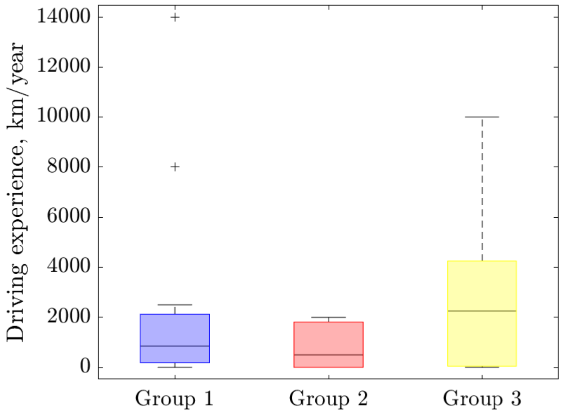
Figure L-7: $\bar{\Delta}$RMS(e) vs. Control & Simulation experience (C&S = Control & Simulation).

Participants' Car Driving Experience

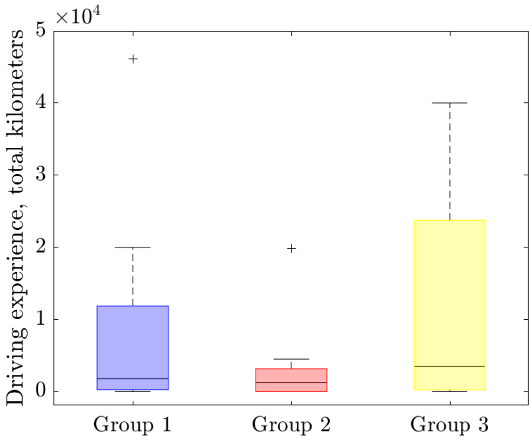
Participants' car driving experience is expressed in three different manners: (1) as years of driving experience, (2) as estimated km/year being driven, and (3) as estimated total kilometers driven.



(a) Years of car driving experience

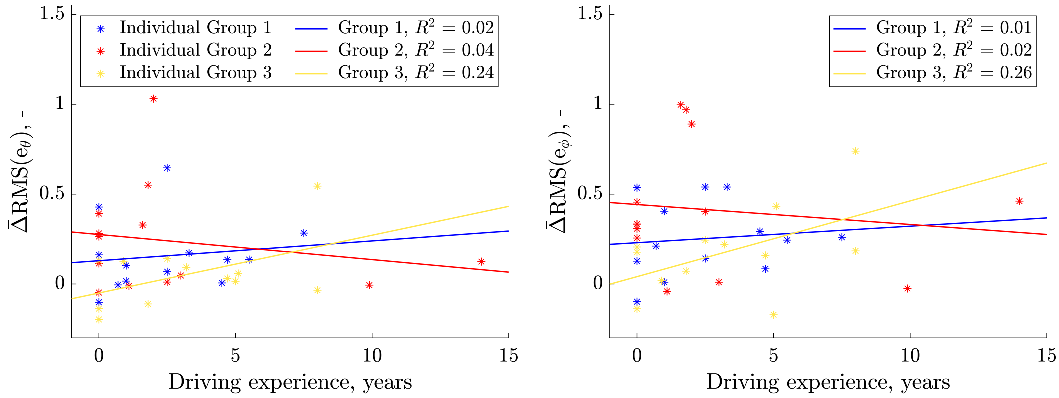


(b) Estimated km/year being driven



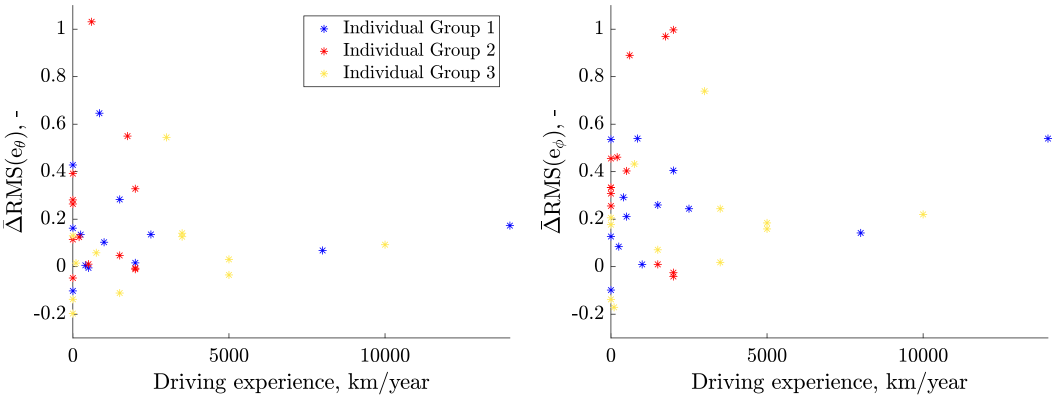
(c) Estimated total kilometers driven

Figure L-8: Participants' car driving experience.



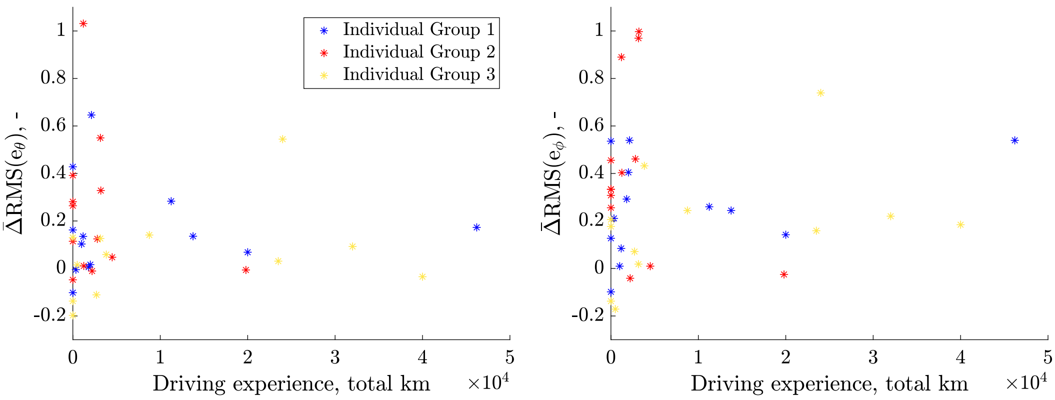
(a) Pitch θ **(b)** Roll ϕ

Figure L-9: Correlation $\bar{\Delta}$RMS(e) and years of car driving experience.



(a) Pitch θ **(b)** Roll ϕ

Figure L-10: $\bar{\Delta}$RMS(e) vs. estimated km/year being driven by car.

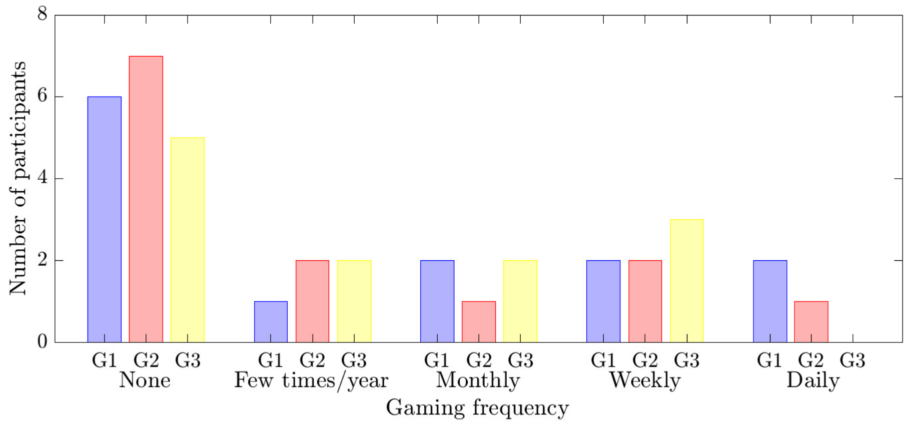


(a) Pitch θ **(b)** Roll ϕ

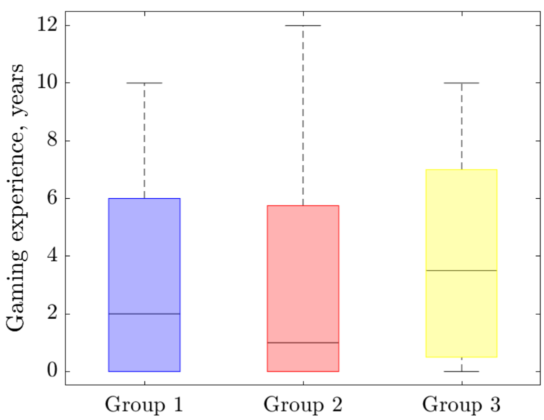
Figure L-11: $\bar{\Delta}$RMS(e) vs. estimated total kilometers driven by car.

Participants' Gaming Experience

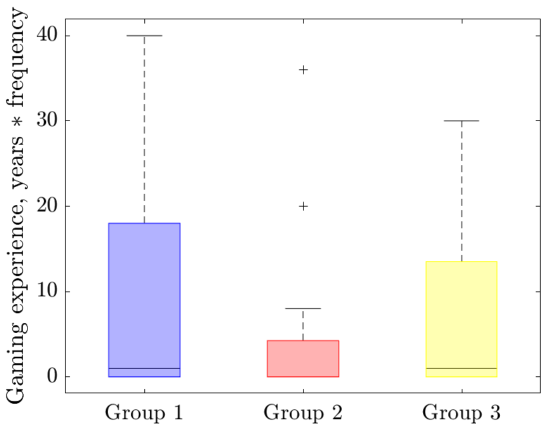
Participants' gaming experience is expressed in three different manners: (1) as participants' current gaming frequency, (2) as the number of years since participants started gaming, and (3) as an interaction score between participants' years of gaming experience and their gaming frequency. To calculate this interaction score, participants' years of gaming experience are multiplied by their gaming frequency. For gaming frequency the following values are used: none = 0, a few times per year = 1, monthly = 2, weekly = 3, and daily = 4.



(a) Gaming frequency

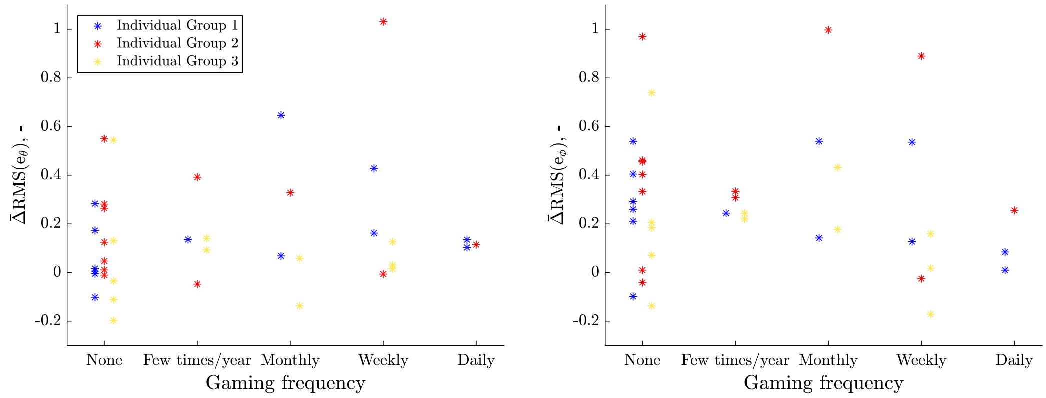


(b) Years of gaming experience



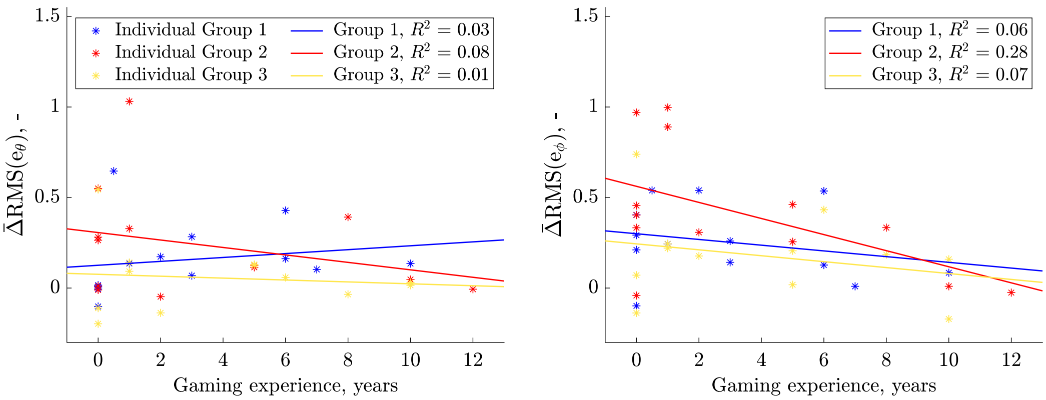
(c) Gaming experience (years ∗ frequency)

Figure L-12: Participants' gaming experience (G = Group).



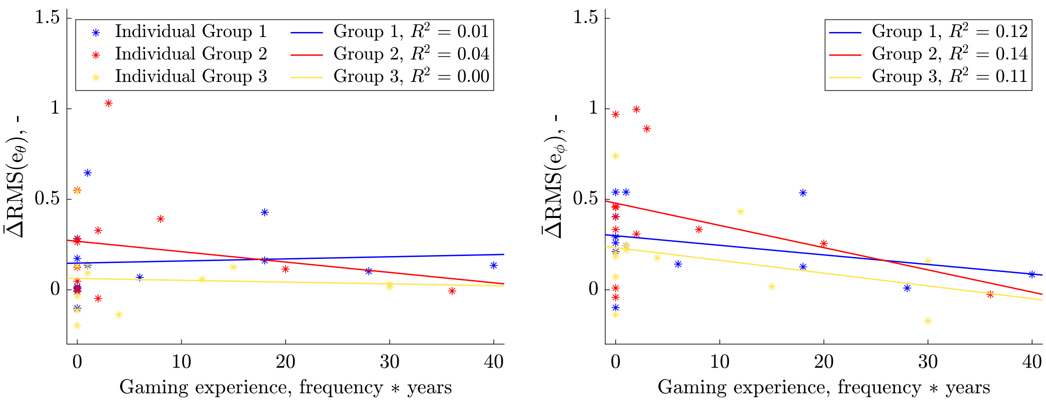
(a) Pitch θ **(b)** Roll ϕ

Figure L-13: $\bar{\Delta}$RMS(e) vs. gaming frequency.



(a) Pitch θ **(b)** Roll ϕ

Figure L-14: Correlation $\bar{\Delta}$RMS(e) and years of gaming experience.



(a) Pitch θ **(b)** Roll ϕ

Figure L-15: Correlation $\bar{\Delta}$RMS(e) and gaming experience (years $*$ frequency).

Appendix M

Experiment Curiosities

To recruit experiment participants:

- 1 lecture talk was given in the second-year course AE2235-II Instrumentation and Signals to call for experiment participants.
- 74 people were contacted personally to ask whether they would either be interested in participating in the experiment or whether they knew of any other people who might be interested in participating.
  - of which 65 had been contacted in person and/or through Whatsapp and 9 by email.
- 95 posters were hung up around the campus of Delft University of Technology to call for experiment participants...
  - ... of which 15 at the Faculty of Aerospace Engineering and 45 at the Faculty of Mechanical, Maritime and Materials Engineering (3mE)...
    - ... which led to 1 student expressing their interest in participation after seeing the poster at the Faculty of 3mE.
- 141 flyers were handed out to first-year and second-year Aerospace Engineering students during project hours to call for experiment participants.
- 1 post was published on the official TU Delft Aerospace Engineering Facebook page to call for experiment participants.

To schedule the experiment:

- 38 emails were sent to daily supervisor Daan with a request to make a reservation for the HMI Laboratory...
  - ... excluding the drop by's at his office to request a reservation in person.
- 15 experiment sessions were rescheduled due to last-minute cancellations or no-shows.

To actually perform the experiment:

- 227 hours were spent in the Human Machine Interaction (HMI) Laboratory performing the experiment...
  - ... excluding 45 hours of starting up the simulator and waiting for participants to (not) show up.
- the simulator was started up 98 times.

- a total of 5,370 count downs were initiated for the start of each run "*3... 2... 1... GO*"...
  - ... excluding the approximately 200 tracking runs performed to 'warm up' the simulator.
- 2 complete blackouts of the Primary Flight Display were experienced during the experiment.
- 1 refusal of access to the simulator software in the HMI Laboratory was experienced after a university-wide policy change which they had forgotten to notify university staff about...
  - ... due to which 1 participant had to drop out, because the training session could not take place.
- 15.37 kg of chocolate in the form of 79 chocolate bars was handed out to 39 participants as gratitude gifts.