Manual Control Behavior in Depth Control Tracking Tasks with Stereoscopic Vision

M. H. H. Kemna

Faculty of Aerospace Engineering

TU Delft Delft University of Technology

Challenge the future

MANUAL CONTROL BEHAVIOR IN DEPTH CONTROL TRACKING TASKS WITH STEREOSCOPIC VISION

by

M. H. H. Kemna

in partial fulfillment of the requirements for the degree of

Master of Science
in Aerospace Engineering

at the Delft University of Technology,
to be defended publicly on Tuesday September 1, 2020 at 01:00 PM.

Supervisor:	Dr. ir. D. M. Pool,	TU Delft
Thesis committee:	Prof. dr. ir. M. Mulder,	TU Delft
	Dr. ir. M. Wentink,	multiSIM Ltd.
	Dr. M. A. Mitici,	TU Delft

An electronic version of this thesis is available at `http://repository.tudelft.nl/`.

TU Delft Delft University of Technology

PREFACE

In loving memory of my dear mother

This document signifies the completion of my master thesis and the end of my journey as a student at the Faculty of Aerospace Engineering at Delft University of Technology, and what a journey it has been. I feel privileged to have had the opportunity to learn and work in the field of aviation with brilliant and kind people. I am thankful to many who have made this journey possible with their guidance and support.

This master thesis would firstly not have been written if it wasn't for the guys in Soesterberg. Bernd and Mark gave me the opportunity, time and space to work on a very interesting project for which I am most grateful. Together with Martijn, Joris and Stephen an experiment setup essential to this research was made and therefore to all you guys: thanks. In addition, I would like to thank the guys at the Royal Netherlands Air Force for their contribution to this project. The knowledge and expertise of Kees, Wilbert and Ton and the pilots and boom operators of Eindhoven Air Base's 334 Squadron helped me tremendously for which I want to thank you all.

The step of making a project for the Royal Netherlands Air Force into a scientific research project proved quite the challenge initially. Over the course of research, I experienced several difficulties that probably every new guy into scientific research should face. However, I never felt alone when sometimes progression got more difficult. I am thankful to Daan and Max for their exceptional professionalism, brilliance and kindness from start to end of this project. You guys helped and challenged me to make this research better for which I thank you both.

Lastly, I want to thank the people at home for believing in me and supporting me "in the big city". First of all, my parents and sister. We have been through much together and yet you have supported me and been there for me all the way to the end. Thank you, "pa, ma en Mau". I also thank you, Erik en Henriët, for your care and support to me. And of course you, Melissa. I know the patience you had to have and understand this has been a journey as much for you as it has been for me. Nonetheless, you were and still are always there for me. I want to thank you for all you have done and I love you.

Maarten Kemna
August 20, 2020

EXECUTIVE SUMMARY

Our society has reached a point where technology allows for the design and creation of many advanced systems. Since technology became ever more complex, system designers acknowledged the growing need for understanding human control behavior. Novel research was undertaking to find methods for the identification and modelling of human control behavior. Ultimately this would provide engineers with powerful tools to optimize the design of a human-machine system for enhanced safety and performance. However, with increasingly complex technology at one's disposal nowadays, the availability of sophisticated technology has surpassed our understanding of their impact on human adaption and behavior to these interfaces. A recent study by Mulder et al. (2018) stated the limited knowledge on the impact of stereoscopic displays as modern system interfaces in terms of human control behavior. Remarkably, numerous applications of control tasks with stereoscopic vision on depth are carried out every single day whilst their implication apparently is hardly understood from a cybernetics point of view.

This document considers a novel research envisioned as a first attempt to identify and model human control behavior in depth control tracking tasks with stereoscopic vision. The goal is to gain a basic understanding of currently lacking knowledge on the adaption and behavior of human operators with stereoscopic displays. The research comprises a number of different fields of study that together form the basis for human control behavior in depth control tasks.

A human being navigates through and interacts with the surrounding environment for many thinkable reasons. Perception of the environment comes for a great portion from the visual system. Visual perception makes a human aware of the surroundings by the sensing of visible light. The eyes create two unique images of the environment that are passed on to the brain. In the visual cortices complex and lesser understood processes develop a meaning to the images provided by the eyes. That is, visual perception is the result of a complicated system in the brain. Another high-order neural process generates depth perception which is crucial for three-dimensional world we live in. The construction of a three-dimensional single image of the environment in our brain follows from numerous sources of information about depth in the images called depth cues. Some may be perceived monocularly and some require binocular vision by sufficiently overlapping visual fields of both eyes. Moreover the strength of these depth cues vary and the sensation of depth may therefore alter depending on the (artificial) environment.

The adoption of binocular vision and third-dimensional depth in control tracking tasks renders a new concept for control tasks in cybernetics. Whereas classical control tracking tasks consider predominately flat-plane displays, alterations must be made to the well-known compensatory and pursuit displays to allow for depth control tasks. These changes may heavily affect the relation between display and human controller in the human-machine interaction diagram by McRuer and Jex (1967). The perception of system states is namely said to change from absolute perception from flat-plane displays to perceived distances in a stereoscopic display by Wickens (1990). That may imply that control behavior in depth control tasks is continuously dependent on and perhaps nonlinear to the distance to the target.

To identify such control behavior a research is proposed to conduct experiments with manual depth control tracking tasks with stereoscopic vision. From experimental data, an attempt will be done to model human control behavior with current identification and modelling methods. Such results may firstly determine whether current techniques can be used to analyse control behavior in depth control tasks and secondly indicate potential differences with classical flat-plane control tasks. Ultimately, the research could contribute to the knowledge and models on control behavior by inclusion of identified human control behavior in depth control tracking tasks with stereoscopic vision.

NOMENCLATURE

α	Visual angle
$\delta, \Delta d$	Objects' relative distance
ω_{nm}	Natural frequency of neuromuscular system
$\omega_{t,d}$	Forcing function frequency
$\phi_{t,d}$	Forcing function phase shift
ρ	Relative disparity
$\rho^2_{u_{t,d}}$	Correlation coefficient
$\sigma^2_{e,u}$	Variance
τ_p	Human controller delay time
ζ_{nm}	Damping ratio of neuromuscular system
a	Interpupillary or interocular distance
$A_{t,d}$	Forcing function amplitude
e	Control error
E, d	Object distance to observer
f_d	Disturbance signal
f_t	Target signal
H_c	Controlled element transfer function
H_p	Human controller transfer function
H_{nm}	Neuromuscular system transfer function
H_{OL}	Open-loop transfer function
K_p	Human controller gain
$K_{d_{rel}}$	Relative display gain
N	Measurement sample size
n	Human remnant
$n_{t,d}$	Forcing function base frequency multiplier
R^2	Coefficient of determination
u	Control input
x	Follower element position

ACRONYMS

AAR Air-to-Air Refueling.

ARO Air Refueling Operator.

DH Depth display condition with hyper-stereoscopic vision.

DM Depth display condition with monoscopic vision.

DS Depth display condition with stereoscopic vision.

E1 Environment 1.

E2 Environment 2.

FP Flat-plane display condition.

FRF Frequency response function.

HC Human controller.

HUD Head-Up Display.

IOD Interocular distance.

IOD Interpupillary distance.

LGN Lateral geniculate nucleus.

NMS Neuromuscular system.

PG1 Participant group 1.

PG2 Participant group 2.

RNLAF Royal Netherlands Air Force.

S3D Stereoscopic three-dimensional.

TRVS Tanker Remote Vision System.

CONTENTS

1

INTRODUCTION

Human beings have been interacting with their environment since the start of their existence. The tools and machines at our disposal used to accomplish an enormous range of tasks is increasing ever since, as is their complexity. This more and more drove the need to understand *how* human operators behave with and adapt to the systems they are coupled with. The interaction of a human being with the machine may, as a result, be optimized to enhance performance and safety. A noteworthy shortfall of these models is the inclusion of and adaptation to the three-dimensional environment we life in. The vast majority of current knowledge on manual control behavior comprises flat-plane two-dimensional representations of the surrounding environment. However, technology allows for new ways of presenting a three-dimensional environment on displays. One may argue that the insurmountable requirement to represent a three-dimensional world on a two-dimensional viewing plane disappears. Nonetheless, with new technologies and opportunities rise new uncertainties. How do human operators adapt to modern interfaces such as stereoscopic three-dimensional displays? Are our current models even applicable to systems that involve tracking tasks based on visual perception of depth? These questions impose that advancing technologies leave a gap in our understanding of human control behavior in a wider range of applications.

This document deals with the numerous steps taken to design and carry out a novel research to gain knowlegde on manual control of depth. It is the preliminary report of the master's thesis devoted to the identification of manual control behavior in depth control tracking tasks with stereoscopic vision. A literature study is presented encompassing various domains of research to explain the basis of manual depth control. The essence of current knowledge on manual control behavior is provided as well as information on physiology to define the concept and implications of depth control. Moreover, a description is given of the intended research experiment to gather experimental data during (depth) control tracking tasks with stereoscopic-enhanced vision. This research could ultimately extend our knowledge on human control behavior in a non-flat, three-dimensional environment.

1.1. MOTIVATION FOR RESEARCH

Understanding and modelling of manual control behavior is an essential research topic. It allows for better designed systems in human-machine interactions [1–4]. Although many environments wherein these interactions may take place can be represented in a two-dimensional flat plane, it is hardly thinkable all aspects of our environment can be displayed on a viewing plane without the element of depth. Nonetheless, one may question why one would implement a depth control task rather than a flat-plane control task given the potential difficulties with perception of depth [5, 6], i.e., distances along the third dimension? Such question is perhaps even more worth considering for situations where non-natural binocular vision is implemented.

With the development of certain task needs in numerous fields of industry, however, human operators have been performing control tracking tasks that are not visually represented as a two-dimensional image of a real three-dimensional environment. Many real-life operations thus step up from flat-plane control tasks as their three-dimensional worlds are not or cannot be represented in a flat-plane environment. This drove the need for better interfaces for non-flat control tasks and boosted implementation of stereoscopic 3D (S3D) displays [5, 6]. Wickens [7] wrote as a motivation for the adoption of stereoscopic 3D displays: "Two basic human factors arguments may be made for the implementation of 3D displays: (1) The visual scene of a 3D world is a more "natural,"

"ecological," or "compatible" representation than that provided by 2D displays; and, closely related, (2) A single integrated representation of one object, relation, or scene reduces the need for a mental integration of two or three representations." The use of 3D displays may thus be a valuable alternative for direct natural binocular perception of the environment. Alternatives to natural vision are for instance necessary in situation whereby the controller is physically separated from the actual direct environment. Such situation is often referred to as a teleoperation [8, 9]. In teleoperation "a machine enables a human operator to move about, sense and mechanically manipulate objects at a distance. Most generally any tool, which extends a persons mechanical action beyond her reach, is a teleoperator." [9]

An example of such teleoperation where the operator is heavily reliant on adequate depth perception for his control task is aerial refueling. Aerial refueling is performed by many nations around the globe and has been a field of study for control tasks in conjunction with depth perception [11–14]. Some nations even facilitate a stereoscopic 3D display and remote vision with increased interocular distance to enlarge the sensation of stereopsis [15, 16]. The human operator is shown in Figure 1.1. Using two control sticks, the refueling boom is steered towards the fuel tank opening of the receiving aircraft called receptacle. The human operator (or boom operator in this particular case) is placed in a teleoperative environment. The Royal Netherlands Air Force (RNLAF) operates tanker aircraft equipped with a stereoscopic camera and display setup. The boom operator therefore perceives the three-dimensional environment remotely through an artificial interface. The external view of refueling boom and receiving aircraft are shown in Figure 1.2.

Figure 1.1: A Royal Netherlands Air Force aerial refueling boom operator using a stereoscopic three-dimensional display as a remote vision interface [10]

Figure 1.2: A KDC-10 tanker aircraft of the Royal Netherlands Air Force aerial refuels an F-16 fighter jet [17]

The superiority of stereoscopic 3D displays over classical flat-plane displays for depth control tasks have been studied thoroughly. Researchers conducted numerous experiments to determine how performance indications like accuracy and completion time altered for tasks where visual depth played a major role [5, 6]. A deeper

understanding *how* such measures of performance are attained lacks, however. One may argue that our current knowledge and models fail to predict human adaption and behavior with new interface technology such as stereoscopic 3D displays [18]. Nonetheless, similar to classical manual control behavior on flat-plane tasks, it would be beneficial to be able to adequately describe human control behavior in depth control tasks with stereoscopic vision [18–21].

1.2. RESEARCH OBJECTIVE

This research sets out to develop a better understanding how human control behavior is affected by changing the control tracking task from flat-plane movement(s) to depth movement. Additionally, the implication of human binocular vision is adopted as a research parameter. The relation *flat-plane - depth* and *monocular - binocular* are the research's two main variables of interest. The research objective is presented as:

The objective of this research project is to develop an understanding of human control behaviour in manual depth control tracking tasks with stereoscopic vision *by* comparing pilot describing functions of flat-plane pursuit tracking tasks against pilot describing functions of depth pursuit tracking tasks with both monocular and binocular vision obtained through experimentally gathered data.

It is crucial to understand that the research will focus on manual control behavior of a human controller in a teleoperative setting with remote artificial binocular vision. Applicability to control behavior with natural binocular vision of identified possible alterations to current control behavior models requires investigation beyond the reach of this research.

The focus of this research is inferred by the main research question:

What effect does binocular vision in manual depth control tracking tasks have on human control behavior?

Due to the complex nature of manual control behavior studies -particularly in conjunction with distinct other fields of study- a distributed approach is anticipated. Informative answers are therefore sought in multiple subquestions as a division of the main research question:

1. **What are the key variables affecting depth perception in the human visual system?**
2. **What are the key variables affecting manual control behavior in flat-plane control tracking tasks?**
3. **What are the characteristics of manual control behavior in flat-plane control tracking tasks?**
4. **What is currently known about manual control behavior in depth control tracking tasks?**
5. **What is currently known about the effects of binocular vision in depth control tracking tasks?**
6. **What effect does control of pure depth in tracking tasks have on human control behavior compared to flat-plane control tasks?**
7. **What effect does binocular vision on depth control tracking tasks have on human control behavior compared to monocular vision?**
8. **What effect does visual complexity of surrounding environment have on human control behavior in depth control tracking tasks?**
9. **What effect does expertise with depth control tracking tasks on stereoscopic displays have on human control behavior in depth control tracking tasks?**

Subquestions 6-9 consider an element on the envisioned research that can only be answered after their effects have been analysed. The remaining first five subquestions shall be answered based on literature discussed in Chapters 2, 3 and 4. Firstly Chapter 2 on the visual system answers subquestion one. Subquestions two and three are answered by literature on manual control behavior in Chapter 3. The concept of depth control tracking tasks is introduced in Chapter 4 where subquestions four and five will be answered.

1.3. STRUCTURE OF REPORT

This document comprises three parts. Part I presents the scientific paper on this research's experiment and results. Part II is the literature study conducted prior to the experiment. Part III presents the book of appendices.

The purpose of Part II is threefold. Firstly the concept of manual control in depth control tracking tasks is defined and an explanation of the implications compared to flat-plane control tracking tasks is given. Secondly a more detailed analysis of real applications of depth control is given to motivate the significance of (extending) knowledge on depth control. Lastly a description is provided on the envisioned research and intended experiment to attempt provision of the aforementioned understanding.

The concept of depth control tasks and the human perception of depth requires some basic understanding of physiology of the human visual system. Part II starts off with a general introduction to the visual system and visual (depth) perception in Chapter 2. A similar introduction is given to manual control behavior in Chapter 3. Subsequently, Chapter 4 combines the theories of the two aforementioned chapters to define the concept of manual control in depth control tracking tasks and outlines the implications. Chapter 5 then provides insight into the envisioned research and experiment to extend knowledge on manual control behavior with depth control and stereoscopic vision. In Chapter 6 an analysis is given of some preliminary results of experiment test runs. Finally Chapter 7 concludes the report with a summation of research questions answered at this phase in the project.

I

RESEARCH PAPER

Identification and Quantification of Manual Control Behavior in Depth Control Tracking Tasks with Stereoscopic Vision

Author: Maarten H.H. Kemna*
Supervisors: Daan M. Pool†, Mark Wentink‡ and Max Mulder§

***Abstract*— The perception of visual information is essential in satisfactory control tracking performance. Modern interface technologies such as stereoscopic three-dimensional displays allow for a more ecological representation of three-dimensional surroundings in a remote teleoperative environment without natural binocular vision. However, the effects such interfaces have on current human controller models are unknown yet desired from a cybernetic perspective. This paper presents the work to identify and quantify manual control behavior in depth control tracking tasks with stereoscopic vision. A human-in-the-loop experiment was conducted in two different environments with twenty four participants of which sixteen were inexperienced stereoscopic display users and eight were frequent users. Four different display conditions were tested; a baseline flat-plane view and three depth views with varying stereopsis. The change of predominantly studied flat-plane control tasks to depth tracking tasks with the considered stereoscopic display settings was found to strongly affect human capability to perform control tasks. The control-theoretic approach to identify and quantify manual control behavior in depth tracking tasks helps to extend current knowledge by integration of human behavior and adaption to modern system's three-dimensional interfaces.**

***Index Terms*— Manual control behavior, human-machine interfaces, stereoscopic displays, depth tracking tasks, human controller model, model parameter estimation.**

I. Introduction

A variety of motivations have separated the human controller from the actual location of action in numerous applications in e.g. military and medical industries [1]. The operator often performs a crucial task in a remote, teleoperative environment without natural binocular vision of the surroundings and visual perception may therefore significantly differ from everyday actions within one's direct environment. Nonetheless, accurate perception of visual information is undeniably essential to effectively control in a teleoperation [1]. Feedback of visual information is often provided by simplification of a three-dimensional environment into a two-dimensional representation [2]. Although effective for numerous applications, many real-life operations comprise three-dimensional environments that dictate the essence of accurate visualisation of depth [3]. Tasks with such emphasis on depth perception require interfaces that commonly cannot be represented on a flat-plane display. For instance, military aviation increasingly facilitates modern remote stereoscopic vision interfaces for aerial refueling tasks [4]. However, the advancement and implementation of technologies such as these modern stereoscopic display interfaces have surpassed our understanding of and knowledge on their implication on human control behavior in these visual depth-reliant teleoperation tasks.

According to Wickens [2], the adoption of stereoscopic three-dimensional displays may prove superior over standard two-dimensional flat-plane displays: *"The visual scene of a three-dimensional world is a more 'natural', 'ecological', or 'compatible' representation than that provided by 2D displays"*. A large number of studies on the possible advantageous implementation brought forth a diverse set of experiment results of which over 180 are bundled by McIntire et al. [3]. For six distinct categories, the implementation of stereoscopic displays was graded as "stereoscopic display is better", "mixed" or "no statistical difference with non-stereoscopic display" [3]. Their study concluded stereoscopic displays seem to benefit tasks reliant on depth information in near-field environments most [3]. Furthermore, tasks categorized as "spatial manipulation of real or virtual objects" turned out to mostly benefit from stereoscopic displays out of the six categories. This category has also been extensively researched for manual control behavior and human adaptation to various display types for visual feedback in control tracking tasks [5], [6], [7], [8]. A cornerstone in this field are the operator models by McRuer et al. for compensatory displays [5] and models validated for pursuit displays [6], [7]. However, currently lacking knowledge on manual control behavior in depth tracking tasks would be beneficial to bridge the gap between cybernetic studies and the apparent advantageous stereoscopic displays as modern remote interfaces of three-dimensional visual environment [8], [9], [10], [11].

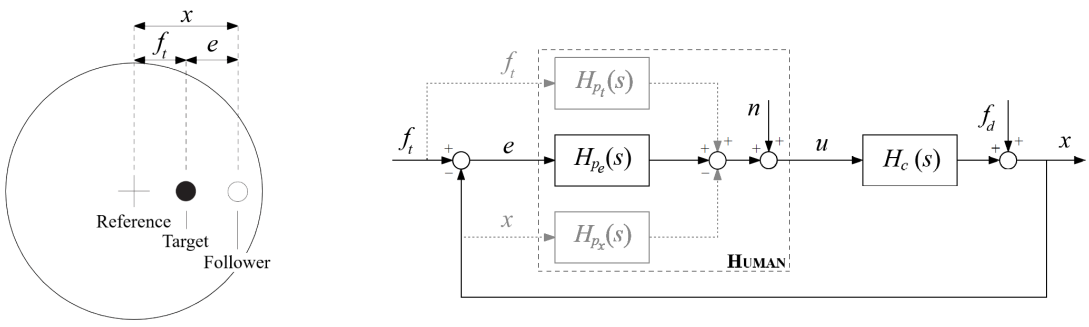
A noteworthy shortcoming of these latter research studies has namely been the inclusion of aforementioned depth tracking tasks with stereoscopic displays. The majority of research on manual control behavior comprises visual feedback of two-dimensional representations through flat-plane non-stereoscopic displays. Mulder et al. [8] state that *"cy-*

*M.Sc. student, Control & Simulation Section, Faculty of Aerospace Engineering, P.O. Box 5058, 2600GB Delft, The Netherlands; m.h.h.kemna@student.tudelft.nl

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

‡Technical director, multiSIM Ltd., Kampweg 55, 3769 DE Soesterberg, The Netherlands

§Professor, Control & Simulation Section, Faculty of Aerospace Engineering, P.O. Box 5058, 2600GB Delft, The Netherlands; m.mulder@tudelft.nl



(a) Pursuit display **(b) Pursuit tracking task**

Fig. 1: Illustrations of (a) Pursuit display and (b) Pursuit tracking task.

bernetics theory has also often been shown to be limited in capturing the full breadth of human cognition and control. Modern interface technologies, such as three-dimensional visual displays . . . are rapidly expanding the way humans can interact with dynamic systems." Although extensively used in numerous crucial applications, the implications of depth-reliant tasks with stereoscopic displays on current knowledge and understanding of manual control behavior remain uncertain.

The goal of this paper is to identify and quantify manual control behavior in depth control tracking tasks with stereoscopic vision from human-in-the-loop experiment data with a cybernetic approach. A paper was published earlier presenting preliminary results [12]. The experiment encompasses collection of control behavior of two participant groups in a pursuit tracking task [6], [13] for two visual environments and four unique display conditions based on realistic applications of depth tracking tasks with artificial stereoscopic vision found in the KDC-10's remote vision aerial refueling system of the Royal Netherlands Air Force (RNLAF) [14]. Participants are tasked with a target-following and disturbance-rejection task (allowing multi-channel human operator identification [8], [15]) on the 3D Aerial Refueling Simulator developed by multiSIM Ltd. as experiment setup. The setup comprises a stereoscopic three-dimensional display and force-feedback control sticks.

This paper firstly addresses in more detail the experiment setup, the human-in-the-loop experiment and data analysis of measured control behavior. These descriptions can be found in Section II. The results after data analysis of the experiment measurements are then presented in Section III. A discussion of the reported findings is subsequently given in Section IV. Lastly, in Section V, the paper is closed with the main conclusions.

II. Method

A. Control Task

The visualisation of a typical pursuit tracking display is given in Fig. 1a and a representation of the control scheme with human operator in a pursuit task in Fig. 1b. The human controller can explicitly perceive target signal f_t, controlled element state x and the tracking error e [6]:

$$e(t) = f_t(t) - x(t) \tag{1}$$

In practice, however, the human controller is found to adopt compensatory control behavior based on H_{p_e} and omits utilization of H_{p_x} and H_{p_t} if e.g. easier tracking tasks or first-order dynamics for the controlled element, i.e., $H_c = K/s$ is selected [6], [7], [13]. For this experiment, the controlled element dynamics are indeed designed as a rate-controlled system and f_t and x responses are thus grayed out in Fig. 1b.

The aforementioned military application of stereoscopic depth control tasks in aerial refueling [14], [16] can be simplified by considering solely third-dimension movements and represented by the well studied single axis pursuit tracking task [6], [7], [13], [17]. The human operator (called a refueling boom operator or simply boom operator) then controls purely the extension of the refueling boom as the *follower element* to the receiver aircraft as the *target element* of the pursuit task. The control task is a derivation of the aerial refueling task to develop a pure depth axis target-following and disturbance-rejection task in an isolated abstract experiment environment.

The external signals f_t and f_d represent the target and disturbance signals, respectively. The target signal f_t defines the position of the target to follow, i.e., the receiving aircraft and disturbance signal f_d adds to (together with human controller input) the position of the follower. The forcing functions f_t and f_d are constructed as random-appearing multisine functions. Three sets of phase shifts per signal are generated to produce three uniquely appearing functions to reduce the possibility of signal recognition. The parameters are shown in Table I. The availability of dual signal input ensures multi-channel frequency-domain identification methods could be used if the human operator was found to still adopt additional H_{p_x} or H_{p_t} control strategies [7], [15], [18] in case of the current depth control tasks.

The depicted pursuit display in Fig. 1a represents a typical flat-plane task. All information is perceived from a two-dimensional plane (often called viewing plane) that is approximately perpendicular to the observer's line of sight. Movement of the target and follower elements occur

TABLE I: Target and disturbance forcing functions characteristics.

Target signal f_t							Disturbance signal f_d						
k	n_t	ω_t, rad/s	A_t, m	ϕ_t, rad			k	n_d	ω_d, rad/s	A_d, m	ϕ_d, rad		
				I	II	III					I	II	III
1	5	0.383	0.920	3.042	0.217	6.150	1	6	0.460	0.956	1.293	5.882	2.323
2	11	0.844	0.695	4.499	2.851	3.298	2	13	0.997	0.669	4.454	5.624	5.493
3	23	1.764	0.346	5.980	0.911	0.323	3	27	2.071	0.302	2.632	3.489	6.010
4	37	2.838	0.176	1.269	1.129	0.336	4	41	3.145	0.162	4.173	0.654	2.639
5	51	3.912	0.107	2.110	0.877	2.899	5	53	4.065	0.109	1.037	0.820	3.771
6	71	5.446	0.065	0.645	3.996	2.674	6	73	5.599	0.068	5.492	1.370	5.449
7	101	7.747	0.041	0.435	5.621	1.399	7	103	7.900	0.043	4.946	4.811	4.877
8	137	10.508	0.029	0.073	1.516	2.049	8	139	10.661	0.031	2.961	1.520	6.079
9	171	13.116	0.025	3.318	3.352	1.274	9	194	14.880	0.024	3.068	2.955	5.615
10	226	17.334	0.021	3.082	4.610	2.080	10	229	17.564	0.022	2.833	3.515	6.273

within this viewing plane as either horizontal, vertical or a combination of both. This situation is shown in Fig. 2a. When considering *depth* control tracking tasks, the object no longer necessarily remains within the viewing plane and may move parallel to the depth axis presented in Fig. 2b. In fact, this research considers depth control tracking tasks as tasks in which both the target and follower elements exclusively *move parallel with* and *along* the line of sight. The pursuit display, shown for flat-plane tasks in Fig. 1a is again presented for depth tracking tasks in Fig. 2c with a top-down view of the same situation in Fig. 2d.

The altered axis of control imposes a difference on the perception of distances between elements, too. As information of a three-dimensional environment is shown on an image surface (irrespective of three-dimensional or flat-plane display interfaces), *"an inherent ambiguity"* arises according to Wickens [19]. For depth tracking tasks he says: *"the absolute distance represented by a point along the line of sight cannot be ascertained with high accuracy, compared to absolute distances parallel with the viewing image plane"* [2]. This paper focuses on the implication of altered control direction by identifying manual control behavior in depth tracking tasks and explicitly comparing it to the better studied flat-plane control behavior.

B. Experiment Conditions

The experiment comprises four experiment conditions each with a unique display configuration. The display conditions are shown in Table II. All display conditions provide a visual representation of the target and follower at a distance of 18 m from the point of observation. This distance was selected to comply with the average distance between the remote stereo-cameras, i.e., observation point of a RNLAF KDC-10 tanker aircraft and the refueling receptacle of a receiver aircraft [14], [16]. First of all, a baseline condition on flat-plane (FP) tracking tasks is included to allow for comparison with the better studied classical flat-plane tasks of previous studies [8], [20]. The control direction is represented by the perpendicular vertical axis in Fig. 2a.

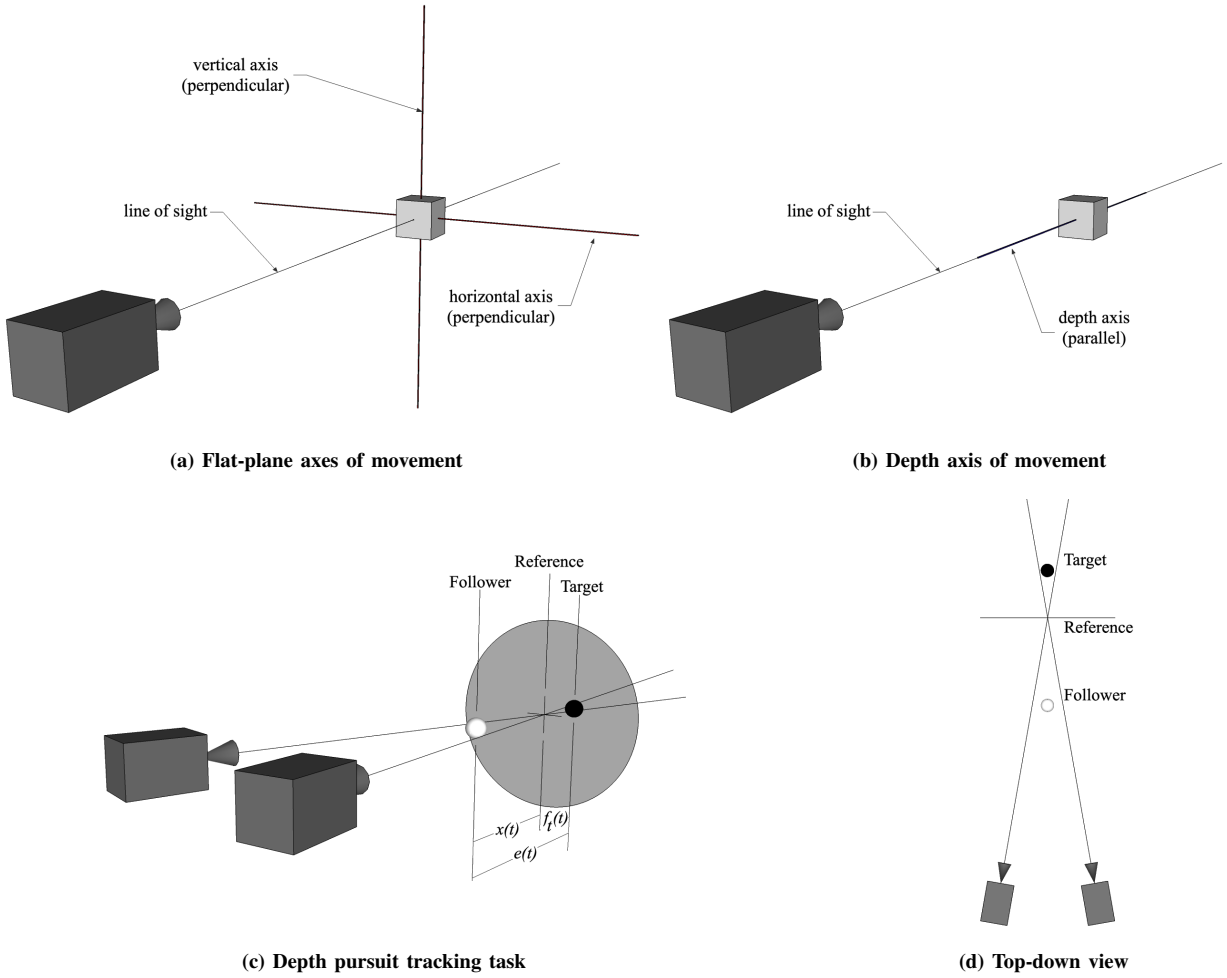
The remaining three display conditions comprise tracking tasks along the depth (D) axis as shown in Fig. 2b. They vary in interocular distance (IOD) as shown in Table II. The first depth display condition provides a monoscopic view (*D-Mono*), i.e., no interocular distance. The second condition provides stereoscopic vision (*D-Stereo*) with approximately human interocular distance of 6 cm. Lastly, a hyper-stereoscopic view (*D-Hyper*) is provided with interocular distane of 50 cm. This condition is based on the remote vision system characteristics of the RNLAF KDC-10 [14], [16]. The natural stereoscopic as well as the hyper-stereoscopic view (increased eye distance to provide higher-than-human threshold levels of stereopsis [14], [16]) provide the binocular depth cue of stereopsis (or binocular disparity) which is believed to be one of three most dominant depth cues [2], [3] that contributes greatly to the ability to distinguish depth between two objects at a certain distance [21].

TABLE II: Experiment conditions.

Condition	Axis	Description	IOD
FP	Vertical	Isometric (flat-plane) view	–
D-Mono	Depth	Monoscopic view	0 m
D-Stereo	Depth	Natural stereoscopic view	0.06 m
D-Hyper	Depth	Hyperstereoscopic view	0.5 m

C. Experiment Environments

The eventual perception of depth is a complex and complicated process made up of many possible cues of third-dimension distances [2]. Cues like binocular disparity, motion parallax and occlusion were found to contribute more significantly to the capability of judgement of distances in depth [2], [21]. The presentation of the pursuit task and display conditions of Section II-A and Section II-B is therefore believed to likely affect the results and eventual conclusions on depth tracking with (non/hyper-)stereoscopic vision. Environment *E1* is therefore designed to isolate as many possibly influencing parameters from the experiment measurements (see Fig. 3a). A simple visual representation of the target (red crosshair) and follower (green crosshair) are selected to omit possible additional depth cues through occlusion (which is believed to be one of three dominant depth cues [2]) of either crosshair. *E1* by design thus comprises no more than an all-gray environment which is not informative of perspective scaling like studies on the effect of perspective on control behavior [22], [23], and two visual representations of the target and follower elements.



(a) Flat-plane axes of movement

(b) Depth axis of movement

(c) Depth pursuit tracking task

(d) Top-down view

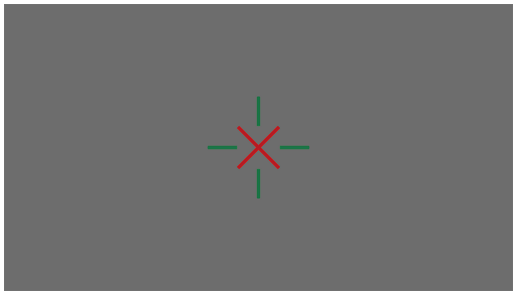
Fig. 2: Illustrations of (a) Flat-plane axes of movement and (b) Depth axis of movement (c) Depth pursuit tracking task with binocular vision and (d) Top-down view of a depth pursuit tracking task.

Given the contrast between basic appearance of experiment environment *E1* and the complicated nature of human depth perception [2], a second environment has therefore been included to assess the applicability of findings in highly simplified visual environments to more realistic environments. At present, the impact of visual information isolation from an environment in which depth tracking tasks are to be conducted is uncertain. Therefore, environment *E2* is tested to potentially provide initial insights in the effect of experiment environment selection on identification of manual control behavior in depth tracking tasks with stereoscopic vision. The remote vision system environment of the heavily depth-reliant aerial refueling boom operator [14] is selected as a basis for realistic environment *E2*. The environment is presented in Fig. 3b.

The control task described in Section II-A is kept identical for both environments. Fig. 4 firstly shows the target element as the receiver aircraft and the follower element as the refueling boom nozzle. The elements have been highlighted with equivalent green and red crosshairs as in *E1*. Both the target and follower element transition purely along the black line indicating the line of sight in Fig. 4. The orientations of the red and green crosshairs changed (see Fig. 3a-b) to omit the possibility of occlusion not only of either crosshair, but also of the receiver aircraft and refueling boom. The flat-plane display condition *FP* was not tested in the realistic environment *E2*.

D. Apparatus

The human-in-the-loop experiment was conducted at the Faculty of Aerospace Engineering at Delft University of Technology for *PG1* and RNLAF Eindhoven Air Base for *PG2*. Measurements were collected on the custom-built 3D Aerial Refueling Simulator (shown in Fig. 5) developed by multiSIM Ltd. in Soesterberg, the Netherlands. The experiment setup comprises the 3D PluraView 28" 4K Monitor as stereoscopic three-dimensional display. The display consists of two 28-inch monitors each with a resolution of 3840x2160 pixels at a 60 Hz refresh rate. A semi-transparent mirror allows simultaneous observation of each display. Passive po-



(a) Simplistic environment *E1*

(b) Realistic environment *E2*

Fig. 3: Illustrations of the developed experiment environments (a) Simplistic environment and (b) Realistic environment.

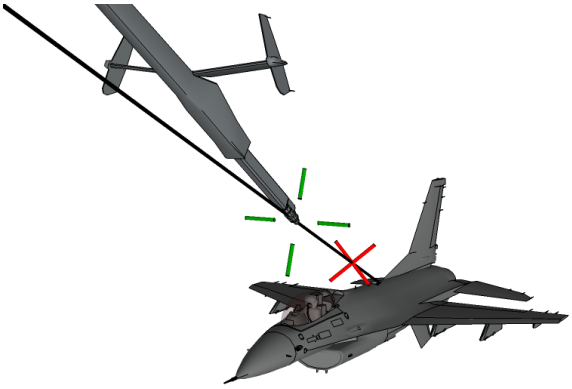


Fig. 4: A schematic perspective view on a boom and receptacle system in an aerial refueling scenario with highlighted follower (green crosshair on boom nozzle) and target (red crosshair at some distance to receptacle) and the line of sight is shown as a solid black line.

Fig. 5: An illustration of the 3D Aerial Refueling Simulator developed by multiSIM Ltd. used as a measurement setup for acquisition of human-in-the-loop experimental data.

larized glasses subsequently allow only one of two displays to be visible to each eye. At an approximate distance of 80 cm to the display, the stereoscopic three-dimensional display is capable of providing two half images with a minimum binocular disparity smaller than human threshold levels of stereoacuity identifiable with stereoacuity tests such as the "TNO Stereopsis Test" as well as the "Titmus Stereo Fly Test" [24]. That is, the experiment setup should have no effect on the minimally perceivable distance in depth as human thresholds are generally higher [25]. The *estimated* visual system delay time is approximately 50 ms based on considered hardware and software.

The 3D Aerial Refueling Simulator in Fig. 5 is equipped with dual CLS-E Active Force Feedback Joystick by Brunner Elektronik AG. For measurement of the human control input u only the right control stick was available during the experiment. Control inputs were given by excitation of the control stick in the pitch axis (aft movement steers follower up in *FP*, aft movement steers towards observer in *D-Mono*, *D-Stereo*, *D-Hyper*). The maximum allowed pitch axis travel was ± 20.5 degrees.

E. Participants & Procedures

Participants selected for the human-in-the-loop experiment (with approval by the TU Delft Human Research Ethics Committee, application number 862) were firstly students and staff members of the Delft University of Technology defined as Participant Group 1 (*PG1*) who mostly had prior experience with tracking tasks, but were inexperienced users of stereoscopic three-dimensional displays. On the other hand, RNLAF refueling boom operators as Participant Group *PG2* were selected as experienced users of these displays. All participants provided written consent and were tested for their capability to perceive stereopsis or binocular disparity (known as stereoacuity) as an inclusion requirement. The "TNO Stereopsis Test" as well as the "Titmus Stereo Fly Test" [24] where used to assure normal stereoacuity among all participants [25].

Prior to actual participation in experimental measurements, participants were briefed on the experiment and the task they are taking part in. Participants all voluntarily agreed to take part and were aware their participation could be stopped unconditionally at any point in time. Subsequently, all experiment conditions and environments were first shown to the participants to familiarize with the task, conditions and environments. After familiarization, all participants would do

several runs on a specific experiment condition for actual measurements of control behavior. The order in which experiment conditions were shown to participants during the measurements phase was randomized with the Latin square designs shown in Table III. As also shown in the table, due to restrictions in the time available, only *PG2* tested the realistic environment *E2*.

F. Data Analysis

A total of four dependent measures are presented as quantification of manual control behavior changes over the various experiment conditions in this paper. Each participant provided three runs of actual measurement. The method of averaging measurement data and analyzing mean values is indicated per dependent variable.

1) Task performance and control effort: The variance of the error signal $\sigma^2(e)$ as a measure for task performance and variance of the control input $\sigma^2(u)$ for control effort were measured. Spectral method analysis of the variances yields the individual contributions of the target signal f_t, the disturbance signal f_d and human control remnant n [26].

$$\sigma^2(x) = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2 \tag{2}$$

Eq. (2) shows the relation for measurement of variances where x represents either e or u for a measurement of sample size N. The error and control input signal variances were calculated for each participant's measurement run separately and averaged to mean variances per group per condition.

2) Control signal correlation coefficient: The linearity of the human controller's response with respect to the target and disturbance signals can be shown with correlation coefficients of control input u to the target signal f_t and disturbance signal f_d [5]:

$$\rho^2_{u_{t,d}}(j\omega_{t,d}) = 1 - \frac{\tilde{S}_{uu,n}(j\omega_{t,d})}{S_{uu}(j\omega_{t,d})} \tag{3}$$

Eq. (2), where $\tilde{S}_{uu,n}(j\omega_{t,d})$ is the contribution of remnant noise spectrum at the frequencies of ω_t and ω_d (shown in Table I) and $S_{uu}(j\omega_{t,d})$ the peak value, is a measure of the signal-to-noise ratio of correlation of u with f_t and f_d separately. The results are hereafter presented where the correlation coefficient was calculated for each individual measurement run and subsequently averaged to mean correlation coefficients per frequency of ω_t and ω_d per participant group and experiment condition.

3) Control error consistency: Analysis of the consistency of control error e with the location of the target f_t is done to reveal possible human adaption to perception of distances along the line of sight. In an objective way the achievable minimum error e between two objects increases with the distance of those two objects to the observer as the relative strengths of some dominant depth cues degrade [21].

However, recent studies have shown that the human controller may actively adapt the control strategy to compensate for scaled down visual stimuli on perspective displays [22], [23]. An additional effect of depth tracking in terms of visual perception is namely found in the scaling of visual objects on displays. Due to perspective, object sizes as well as distances between object visually scale with a display gain $K_{d_{rel}}$ over the range of movement of the target and follower element as shown in Fig. 6. Naturally, flat-plane control tracking tasks avert such effect due to different human perception of distances along the line of sight compared to distances perpendicular to the line of sight [2].

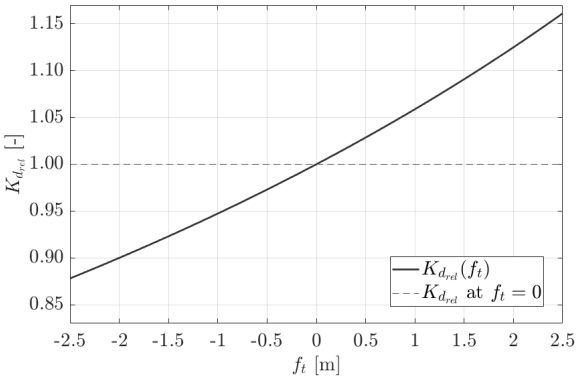


Fig. 6: An illustration showing the relative display scaling $K_{d_{rel}}$ over the range of f_t as a result of perspective. At positive values of f_t the target is in front of the convergence point, i.e., closer to the observer.

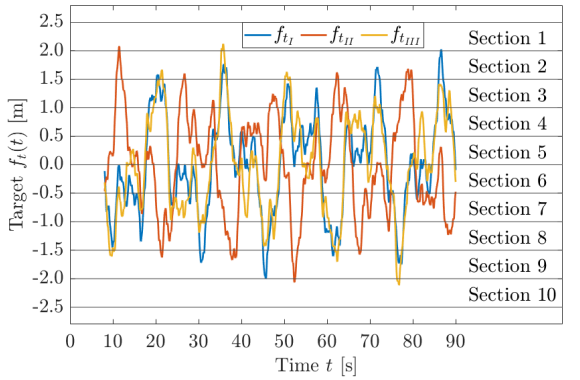


Fig. 7: An illustration showing the division of f_t into sections of 0.5 m distances. The Root Mean Square (RMS) of the control error is calculated per section of f_t.

The root mean squares of the error between the target and follower object are therefore graphically outlined per distance to the observer to reveal whether the human controller potentially adapts to and compensates for subjectively smaller appearing distances between objects at greater distances to those objects. The control error e is calculated as a function of the target position f_t to analyse error (in)consistency with distance between the point of observation and the target. Fig. 7 shows the three target signals of Table I and divides the range of values for f_t in

TABLE III: Experiment Latin Square design.

Participant	Group	Age	Condition						
			1	2	3	4	5	6	7
Student 01	1	25	E1-D-Mono	E1-FP	E1-D-Stereo	E1-D-Hyper	-	-	-
Student 02	1	24	E1-D-Hyper	E1-D-Stereo	E1-FP	E1-D-Mono	-	-	-
Student 03	1	26	E1-D-Stereo	E1-D-Mono	E1-D-Hyper	E1-FP	-	-	-
Student 04	1	24	E1-FP	E1-D-Hyper	E1-D-Mono	E1-D-Stereo	-	-	-
Student 05	1	24	E1-FP	E1-D-Stereo	E1-D-Mono	E1-D-Hyper	-	-	-
Student 06	1	25	E1-D-Mono	E1-D-Hyper	E1-D-Stereo	E1-FP	-	-	-
Student 07	1	25	E1-D-Hyper	E1-D-Mono	E1-FP	E1-D-Stereo	-	-	-
Student 08	1	25	E1-D-Stereo	E1-FP	E1-D-Hyper	E1-D-Mono	-	-	-
Student 09	1	25	E1-D-Mono	E1-D-Hyper	E1-FP	E1-D-Stereo	-	-	-
Student 10	1	24	E1-FP	E1-D-Stereo	E1-D-Hyper	E1-D-Mono	-	-	-
Student 11	1	24	E1-D-Stereo	E1-FP	E1-D-Mono	E1-D-Hyper	-	-	-
Student 12	1	25	E1-D-Hyper	E1-D-Mono	E1-D-Stereo	E1-FP	-	-	-
Student 13	1	27	E1-D-Stereo	E1-D-Mono	E1-FP	E1-D-Hyper	-	-	-
Student 14	1	25	E1-D-Hyper	E1-FP	E1-D-Mono	E1-D-Stereo	-	-	-
Student 15	1	22	E1-FP	E1-D-Hyper	E1-D-Stereo	E1-D-Mono	-	-	-
Student 16	1	27	E1-D-Mono	E1-D-Stereo	E1-D-Hyper	E1-FP	-	-	-
Boom operator 01	2	47	E1-D-Hyper	E1-FP	E1-D-Stereo	E1-D-Mono	E2-D-Hyper	E2-D-Stereo	E2-D-Mono
Boom operator 02	2	58	E1-FP	E1-D-Mono	E1-D-Hyper	E1-D-Stereo	E2-D-Mono	E2-D-Hyper	E2-D-Stereo
Boom operator 03	2	48	E1-D-Mono	E1-D-Stereo	E1-FP	E1-D-Hyper	E2-D-Stereo	E2-D-Mono	E2-D-Hyper
Boom operator 04	2	48	E2-D-Stereo	E2-D-Hyper	E2-D-Mono	E1-D-Stereo	E1-D-Hyper	E1-D-Mono	E1-FP
Boom operator 05	2	49	E2-D-Hyper	E2-D-Mono	E2-D-Stereo	E1-D-Hyper	E1-D-Stereo	E1-D-Mono	E1-FP
Boom operator 06	2	53	E2-D-Mono	E2-D-Stereo	E2-D-Hyper	E1-FP	E1-D-Mono	E1-D-Stereo	E1-D-Hyper
Boom operator 07	2	38	E1-D-Mono	E1-D-Hyper	E1-FP	E1-D-Stereo	E2-D-Hyper	E2-D-Stereo	E2-D-Mono
Boom operator 08	2	58	E2-D-Stereo	E2-D-Mono	E2-D-Hyper	E1-D-Stereo	E1-D-Mono	E1-D-Hyper	E1-FP

sections of 0.5 m for which the error is determined: closest to the observer is Section 1 within (2.5 m $\geq f_t > 2.0$ m), Section 2 within (2.0 m $\geq f_t > 1.5$ m) up until Section 10 where (-2.0 m $\geq f_t > -2.5$ m) and the target is thus furthest away from the observer. The possible scaling of error magnitude with increased observer-to-target distance is indicative of increased task difficulty and lower accuracy of absolute distance perception with depth tracking tasks [2].

$$e_{\text{RMS}} = \sqrt{\frac{\sum_{i=1}^{N}(f_{t_k} - x_k)^2}{N}} \tag{4}$$

Eq. (4) represents the Root Mean Square (RMS) of control error e defined by Eq. (1) as a measure of the mean error and is calculated for each individual measurement run. The results subsequently present the average (e_{RMS}) per participant per condition.

4) Human controller modeling: The measurements of the human controller's responses to the tracking tasks were evaluated with human control behavior identification and modeling techniques to compare manual control behavior between display conditions, environment and participant groups. Eq. (5) firstly represents the identification the human controller's frequency response function (FRF) of Fourier transformed control error e and control input u assessed at the frequencies of the target (ω_t) and disturbance (ω_d).

$$\hat{H}_{p_e}(j\omega_{t,d}) = \frac{U(j\omega_{t,d})}{E(j\omega_{t,d})} \tag{5}$$

The identification of the FRFs to both ω_t and ω_d should also point out any dissimilar trends between the responses to the target and disturbance signal. Although not expected [6], [7], [13], dissimilarity between the two trends indicates adoption of control strategies by the controller other than anticipated H_{p_e}. The model in Eq. (6) associated with rate-controlled systems, i.e., $H_c = K/s$ in previous studies was selected for parameter estimation [5], [8], [20].

$$H_{p_e}(j\omega) = K_p e^{-j\omega\tau_p} \frac{\omega_{nm}^2}{(j\omega)^2 + 2\zeta_{nm}\omega_{nm} j\omega + \omega_{nm}^2} \tag{6}$$

The model comprises four contributing parameters that were estimated by minimum realizations of sum-of-squared-errors calculations. The human controller's control gain K_p and τ_p are firstly indicative of aggressiveness of the response and delay of the response, respectively. The neuromuscular system characteristics are evaluated by its frequency ω_{nm} and damping ζ_{nm}. The four model parameters were calculated per participant per condition from the mean of three FRFs, which were identified for each individual measurement run.

$$R^2 = \max\left(0, 1 - \frac{\sum_{i=1}^{N} ||u[k] - \hat{u}[k|\hat{\theta}]||^2}{\sum_{i=1}^{N} ||u[k]||^2}\right) \tag{7}$$

The accuracy of the estimated model of Eq. (6) to the identified FRFs of Eq. (5) was determined by calculation of the coefficient of determination as shown in Eq. (7). Here $u[k]$ represents the actual measured human input and $\hat{u}$ the modeled response with model parameters $\hat{\theta}$ (see Eq. (6)).

5) Statistical analysis: Statistical significance of differences found in control behavior between conditions was assessed by a repeated-measures two-way ANOVA test, given the assumption of sphericity and therefore normal distribution of the measurement data is not violated.

G. Hypotheses

The design and execution of the experiments comprise four key parameters that are thought to affect manual control behavior in depth control tasks with stereoscopic vision. Four hypotheses are therefore formulated to investigate the implications of these four parameters in these experiments' pursuit tracking tasks. First of all, to test the reduced accuracy of distance perception along the line of sight compared to perception of distances within the viewing plane mentioned by Wickens [2] it is hypothesized that:

H1: Tracking performance degrades in depth control tasks compared to flat-plane tasks given identical control task parameters and visual characteristics, especially at lower and higher frequencies.

Given the potential superiority of stereoscopic displays over classic flat-plane displays in depth tracking tasks [2], [3], from a control theoretic perspective it is therefore also hypothesized that:

H2: Higher quantities of the stereopsis depth cue affect manual control behavior to improve tracking performance in depth control tasks with (hyper)stereoscopic vision.

The second participant group consisting of trained boom operators are believed to be used to perception of depth from hyper-stereoscopic imagery. Studies have shown the benefit of stereoscopic displays to task performance to be related to how well a tasked was learned [3], [27]. Given the more frequent use of stereoscopic displays by the boom operators, it is hypothesized that:

H3: Higher usage frequency of stereoscopic three-dimensional display interfaces improves tracking performance in depth control tasks with hyper-stereoscopic and, to lesser extent, stereoscopic vision.

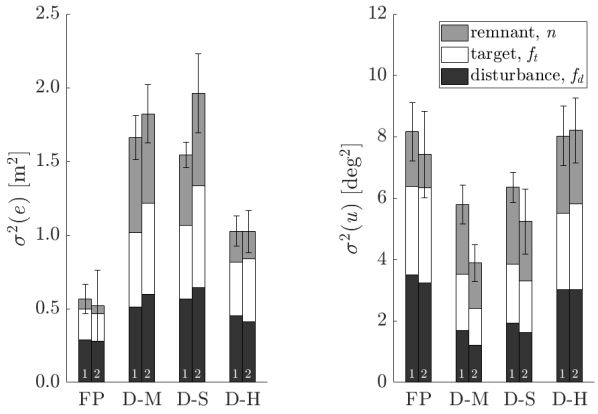
The human visual process of depth perception is complex and highly dependent on availability of various depth cues [2] as is the beneficial addition of stereoscopic displays to enhance task performance [3]. Given the possible influence of visual characteristics of the environment, it is hypothesized that:

H4: Control behavior in depth control tasks in simplified abstract environments is comparable to behavior in more sophisticated and visually realistic (background) environments given availability of sufficient effective depth cues and identical optical characteristics.

III. Results

A. Task Performance and Control Effort

The variations in task performance and control effort are presented in terms of average measured variances of error e and control input u for each specific display condition. Firstly, Fig. 8 shows the variances of e (Fig. 8a) and u (Fig. 8b) for *PG1* and *PG2* marked by "1" and "2" in the bar plots' bases, respectively, in the abstract environment *E1*. Fig. 9 subsequently shows the variances of e (Fig. 9a) and u (Fig. 9b) for *PG2* in abstract environment *E1* and realistic environment *E2* again marked by "1" and "2", respectively. The 95% confidence intervals are included for the total variances of e and u. The build-up of the total variances by contributions of the disturbance signal f_d is indicated in black, the target signal f_t in white and the human controller remnant n in grey. These contributions were isolated using spectral methods [26]. Lastly, the results of statistical analysis are presented in Table IV for both the variances of error and control input between participants group in the first row and variances of error and control input between environments (thus only for *PG2*) in the second row.



(a) Error variance **(b) Control variance**

Fig. 8: Task performance and control effort in variance of (a) Error and (b) Control input between *PG1* and *PG2* in *E1*.

1) Pairwise comparison of participant groups: First of all, the total variance of the error e of *D-Mono* presented in Fig. 8a is, compared to *FP*, significantly higher ($F(1, 22) = 161.9, p < 0.01$) by a factor of 2.8 for *PG1* and 3.6 for *PG2*. Additionally, the contribution of human remnant rises from 12% and 11% in *FP* to 39% and 33% in *D-Mono* for *PG1* and *PG2*, respectively. Simultaneously, the total variance in control input u in Fig. 8b reduces significantly ($F(1, 22) = 29.8, p < 0.01$) for *D-Mono* with relative smaller contributions by f_t and f_d. The portion contributed by n on the other hand increases to nearly 40% of the total variance in control input.

TABLE IV: Error and control input ANOVA results.

		error e			control input u		
		df	F	sig.	df	F	sig.
σ^2_{PG}	Display	(2.3, 50.4)	107.2	0.00	(2.3, 49.8)	19.8	0.00
	PG	(1, 22)	—	—	(1, 22)	—	—
	Dsp.*PG	(2.3, 50.4)	3.4	0.04	(2.3, 49.8)	1.5	0.24
σ^2_E	Display	(1.9, 27.0)	32.2	0.00	(1.7, 24.3)	24.8	0.00
	Env.	(1, 14)	182.9	0.00	(1, 14)	9.8	0.01
	Dsp.*Env.	(1.9, 27.0)	3.8	0.04	(1.7, 24.3)	1.5	0.25

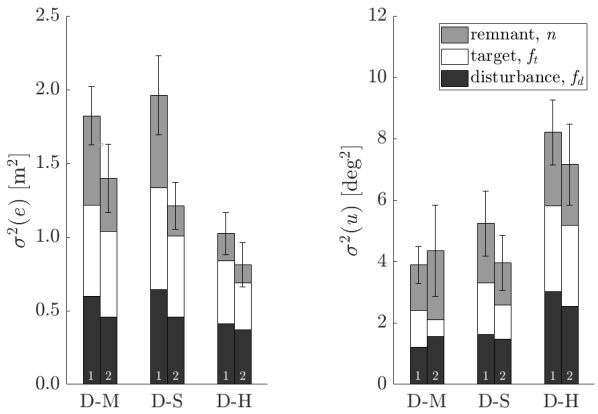
The total variances with subsequent display conditions *D-Stereo* and *D-Hyper* firstly show comparable task per-

formance in Fig. 8a. The control effort (Fig. 8b) between *D-Mono* and *D-Stereo* does significantly rise ($F(1,22) = 14.1, p < 0.01$). Task performance with subsequent *D-Hyper* does significantly increase ($F(1,22) = 101.6, p < 0.01$) as the variance of e in Fig. 8a decreases by between a third compared *PG1* in *D-Stereo* and a half compared to *PG2* in *D-Stereo*. Not only are the total variances lower compared to *D-Mono* and *D-Stereo*, the relative contributions of human remnant decrease as well from 31-39% to 18-21%. In addition, the control effort in Fig. 8b significantly rises ($F(1,22) = 20.7, p < 0.01$) for *D-Hyper* with again reductions of the relative portions of human remnant. The changes in variations of e and u as measures for task performance and control effort, respectively, indicate that the stereopsis depth cue is used in the tracking tasks. However, the equivalent error variances between *D-Mono* and *D-Stereo* in comparison with the significantly different variances between *D-Stereo* and *D-Hyper* suggest that merely the availability of stereopsis is insufficient to notably increase task performance. The quantity of stereopsis of the *D-Stereo* display condition namely appears to insufficiently affect control behavior to significantly increase task performance with respect to *D-Mono*. Furthermore, even though a significant decrease of error variance has been found for *D-Hyper* with respect to *D-Mono* and *D-Stereo*, the hyper-stereoscopic vision on the depth tracking tasks still results in a significantly lower ($F(1,22) = 85.5, p < 0.01$) task performance with respect to the flat-plane tracking task.

In terms of variations between *PG1* and *PG2* only the display conditions *D-Mono* and *D-Stereo* show some differences. Task performance for *D-Mono* in Fig. 8a between *PG1* and *PG2* is within each other's confidence intervals and with no significant difference found. Simultaneously, *PG2* was found to show significantly lower ($F(1,22) = 17.4, p < 0.01$) control effort in Fig. 9b. Subsequent task performance with *D-Stereo* in Fig. 9a was found to be significantly lower ($F(1,22) = 15.9, p < 0.01$) for *PG2* than *PG1*. The remaining display conditions *FP* and *D-Hyper* have similar variances between the participant groups for both task performance as well as control effort as shown by Fig. 8a and Fig. 8b, respectively.

2) Pairwise comparison of environments: The graphical representations of task performance and control effort in the realistic environment *E2* in Fig. 9a and Fig. 9b firstly show no dissimilar trends over the three display conditions compared to task performance and control effort in abstract environment *E1* (Fig. 8). Both task performance and control effort are found to be equivalent for *D-Mono* and *D-Stereo* with a significant reduction ($F(1,14) = 57.1, p < 0.01$) in task performance and significant increase ($F(1,14) = 34.3, p < 0.01$) in control effort for *D-Hyper*.

Comparison of the error variance between the two environments *E1* and *E2* in Fig. 9a clearly shows improved task performance in *E2*, especially for display condition *D-Stereo* with a significant decrease ($F(1,14) = 55.5, p < 0.01$) of σ^2 (2.0 m^2 in *E1* versus 1.2 m^2 in *E2*) and to lesser extent *D-Mono* also with significant decrease ($F(1,14) = 48.0, p <$



(a) Error variance **(b) Control variance**

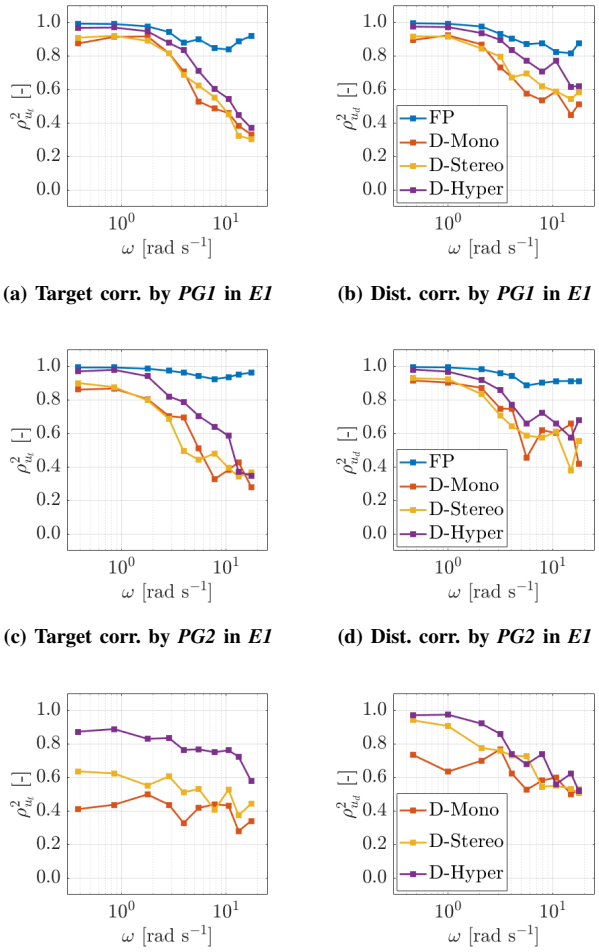
Fig. 9: Task performance and control effort in variance of (a) Error and (b) Control input between *E1* and *E2* for *PG2*.

0.01) of σ^2 (1.8 m^2 versus 1.4 m^2). Simultaneously, the total variances contain smaller parts of human remnant as n drops from 33% to 26% for *D-Mono* and 32% to 17% for *D-Stereo*. On the other hand, the variances in control input u in Fig. 9b show neither significant differences between the total variances nor an obvious reduction of the contribution of n. In fact, the segment of human remnant rises from 38% to 51% for *D-Mono* between *E1* and *E2*, respectively. Although a lower error variance for display condition *D-Mono* can be seen in Fig. 9a with a smaller relative portion of human remnant n (33% in *E1* to 26% in *E2*), an increase can be found in the relative human remnant in control effort. It seems that the contribution of the target signal to the control effort is only 13% with the *D-Mono* display condition in the realistic environment *E2*. The relative contributions of the target signal to the variances in u for *D-Stereo* and *D-Hyper* in the same environment are 28% and 37%, respectively. It thus seems that the control effort based on the target signal f_t increases more rapidly from *D-Mono* to *D-Stereo* and *D-Hyper* (13%, 28% and 37%, respectively) in the realistic environment *E2* than in the abstract environment *E1* (31%, 32% and 34%, respectively). On the other hand, the share of disturbance signal f_d in the total variances of u appears to be higher in *E2* than *E1* for *D-Mono* (31% to 36%) and *D-Stereo* (31% to 37%). The control effort of the boom operators thus seems less based on the target and more on the disturbance signal of the follower with display conditions *D-Mono* and *D-Stereo* in the realistic environment *E2*.

B. Control Signal Correlation

As a measure of linearity of the human controller's input signal u to the applied forcing functions for the target signal f_t and disturbance signal f_d, the average correlation coefficients of u to both f_t and f_d are presented in Fig. 10.

The correlation coefficients reveal the signal-to-noise ratio of the human controller's response at the frequencies of f_t and f_d for the investigated display conditions. Each graph shows the correlation to f_t or f_d for *PG1* and *PG2* in the abstract environment *E1* as well as the realistic environment *E2*.



(a) Target corr. by *PG1* in *E1* **(b) Dist. corr. by *PG1* in *E1***

(c) Target corr. by *PG2* in *E1* **(d) Dist. corr. by *PG2* in *E1***

(e) Target corr. by *PG2* in *E2* **(f) Dist. corr. by *PG2* in *E2***

Fig. 10: Correlation of control input signal u with (a & b) Target and disturbance signal for *PG1* in *E1*, (c & d) Target and disturbance signal for *PG2* in *E1*, (e & f) Target and disturbance signal for *PG2* in *E2*.

Observation of the baseline display condition *FP* in Fig. 10a-d shows a correlation of over 0.8 for control input u with both f_t and f_d for *PG1* as well as *PG2* for all target and disturbance signal frequencies. The human controller thus actively controls and responds to the full scale of frequencies in the target and disturbance signals in *E1*. The remaining depth display conditions *D-Mono* to *D-Hyper* show, on the other hand, a decline of coherence in the high frequency ranges of the target and disturbance signals. Although the correlation may not be as high as the *FP* for lower frequencies either, the difference in coherence between the flat-plane display condition and the depth display conditions is especially visible for higher frequencies.

Not only do the coherence coefficients of Fig. 10a-d support earlier findings on the results for task performance and control effort, they also indicate that the linearity of the human controller's response to both the target and disturbance signals decays with increasing signal frequency. Apparently the task to reduce the error between the target and follower elements in depth tracking task becomes increasingly harder with higher frequencies as suggested by the lower signal-to-noise ratios for *D-Mono*, *D-Stereo* and *D-Hyper*. Similar results to those found for task performance and control effort in Section III-A are again found concerning the different levels of stereopsis depth cue available. The display condition *D-Hyper* in Fig. 10a-d namely appears to have a consistently higher control correlation with f_t and f_d compared to *D-Mono* and *D-Stereo*.

In Fig. 10e-f the control correlations with f_t and f_d are presented for *PG2* in the realistic environment *E2*. It appears that coherence of human controller input u with target signal f_t diverges from the correlation coefficients found for *PG1* and *PG2* in *E1*. Earlier in the analysis of control effort of *PG2* in *E2* it was already shown that the contribution of human remnant n to the total variance with *D-Mono* rose from 38% to 51% with a smaller contribution (31% down to 13%) by the target signal f_t. It turns out that *PG2* responded different to f_t compared to the identical signal in *E1* as is deducible from Fig. 10c and Fig. 10e. The correlation coefficients of u with f_t in especially *D-Mono* and *D-Stereo* are not only lower, the coherence with the target signal appears to show less variations with frequency. Across the full range of frequencies the human controller responded less linear to f_t. This gives rise to the believe that, for the tasks and environments concerned with this research, the control behavior identified in a realistic environment as the RNLAF's KDC-10 TRVS cannot be compared to behavior in a much more simplified and abstract environment such as *E1*.

C. Control Error Consistency

Fig. 11 presents the root mean square of the error (e_{RMS}) as a means to test for consistency of control error e over the range of target distance f_t. Each column in Fig. 11 represents a specific display condition (*FP*, *D-Mono*, *D-Stereo*, *D-Hyper*). All combinations of display conditions, participant groups and environments are then made with *PG1* on the first row and *PG2* on the second (for abstract environment *E1*) and third row (for realistic environment *E2*). The mean e_{RMS} of each participant's three measurements within the considered participant group are presented per section of target signal f_t. Recall that the reference plane where $f_t = 0$ lies a distance of 18 m in front of the observer and that positive values for f_t indicate the target moving closer towards the observer. The full range of $f_t \approx \pm 2.3$ m is divided into ten sections with a range of 0.5 m per section. Each section contains the values for e_{RMS} attained within that specific section. Each plot indicates the linear correlation R of e_{RMS} with the sections of f_t and its p-value based on e_{RMS} within Sections 2-9. The Sections 1

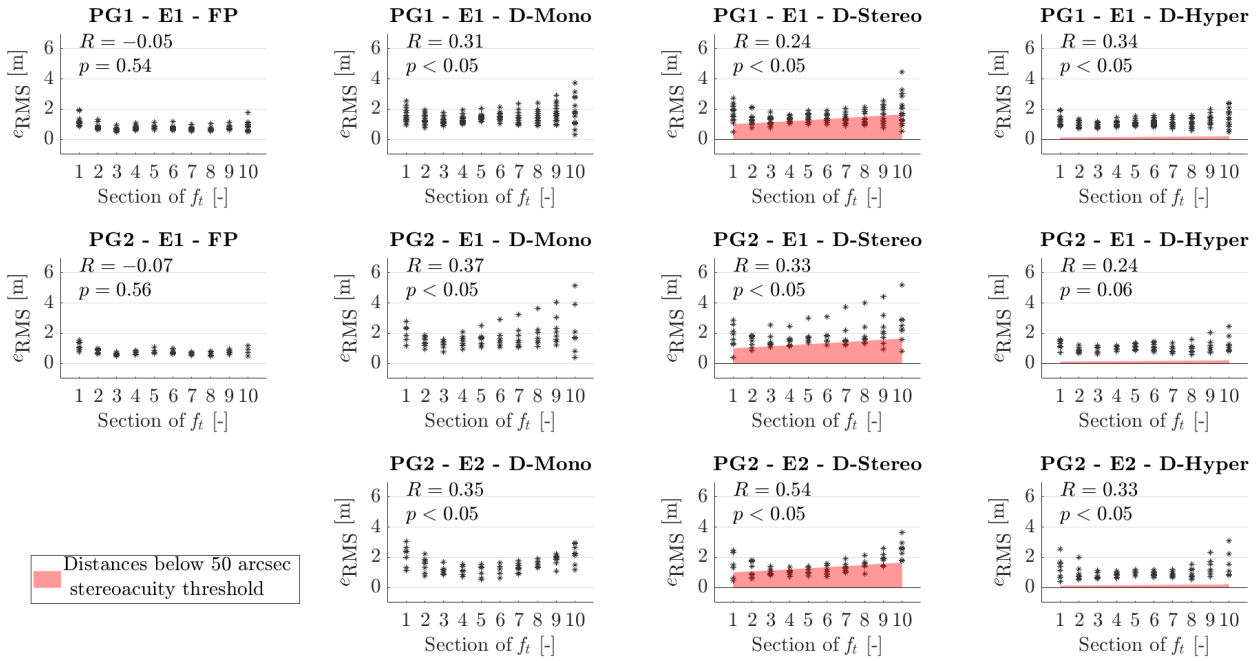


Fig. 11: Error consistency in e_{RMS} per Section of f_t for all conditions

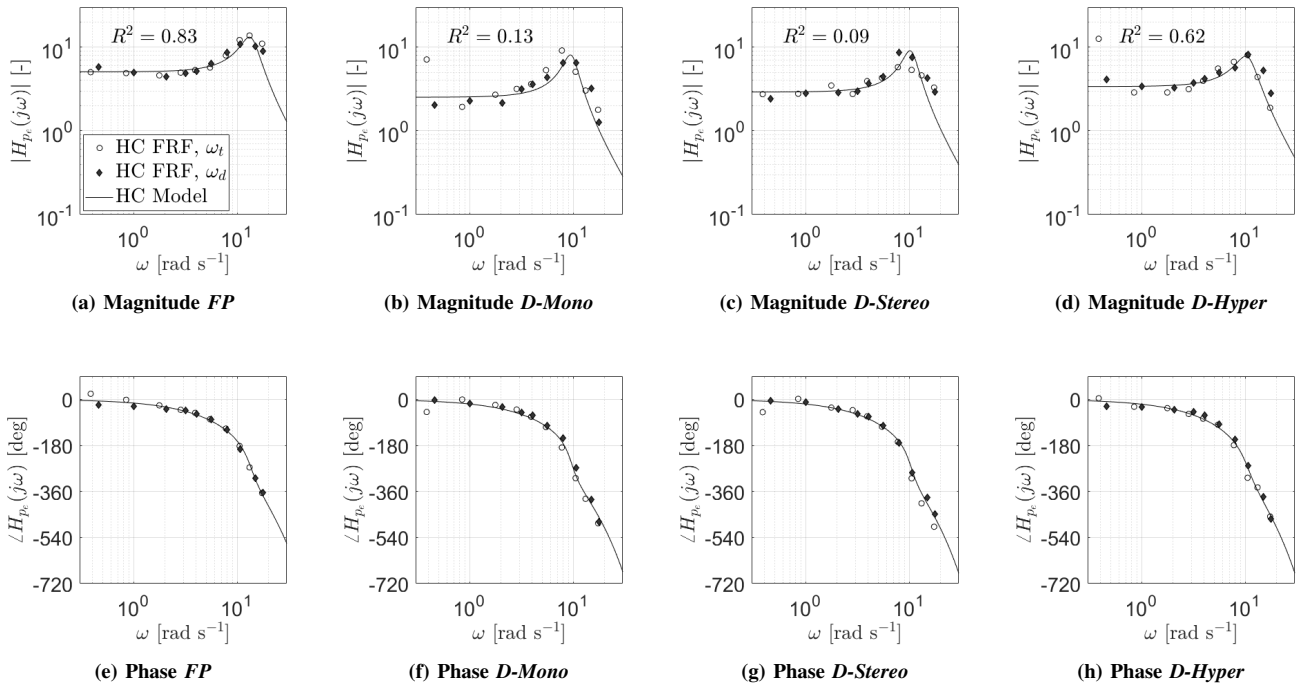
and 10 namely contain only the extreme target locations and corresponding e_{RMS} which are less often reached by f_t (see the division of f_t into sections in Fig. 7) and therefore have a relative low number of contained data points, i.e., N in Eq. (4) is relatively low in Sections 1 and 10. Finally, in the plots of display conditions *D-Stereo* and *D-Hyper* the areas highlighted in red represent distances, i.e., magnitudes of error that provide binocular disparity lower than an example stereoacuity threshold of 50 seconds of arc. This means found values of e_{RMS} within the red zones are magnitudes of error that are not perceived with *stereoscopic* vision.

In Fig. 11 the two figures in the first column show the quantities of e_{RMS} for Sections 1-10 with the flat-plane display condition by *PG1* and *PG2*. Firstly, the overall quantities of e_{RMS} appear to be the lowest in the graphs of *FP*. Smaller root mean squares of control error e indicate a higher control accuracy which is consistent with earlier statements on the nature of human perception of flat-plane versus depth distances [2]. Additionally, over the sections of f_t, the graphs in the most left column of Fig. 11 indicate correlations of -0.05 and -0.07 at non-significant p-values for both *PG1* and *PG2*, respectively. The results thus show no significant correlation of e_{RMS} with the sections of f_t for the *FP* display condition, i.e., the magnitude of error appears consistent over the range of positions of the target element.

The graphs in the remaining columns of Fig. 11 show the course of e_{RMS} with the sections of f_t for the three depth display conditions. Not only are the quantities of e_{RMS} generally higher compared to *FP*, the correlation variable R also reveals significant positive relations (except for *PG2 - E1 - D-Hyper* at $p = 0.06$) between the magnitude of error and the distance of the target to the observer for all display conditions concerned with depth tracking tasks. Apparently the participants could not (or did not) fully compensate for the subjectively smaller appearing errors at larger distances from the point of observation that inherently affect these depth control tasks [21]. These findings indicate that the inconsistent magnitude of errors over the range of possible locations of the target additionally yields inconsistent responses of the human controller to the control error e for different values of f_t. This in turn may result in a more time-varying control response towards control error e in the depth control tracking tasks with display conditions *D-Mono*, *D-Stereo* and *D-Hyper*.

D. Human Controller Modeling

A comparison of control behavior between the display conditions as well as the participant groups and the environments is firstly presented by means of the frequency response function (FRF) for which an example analysis of one participant in *PG1* can be found in Fig. 12. The FRFs describe the human controller's response dynamics to control error e as described in Section II-F.4. Secondly, the human controller's response dynamics are modeled by fitting Eq. (6) to the FRFs identified for all experiment conditions. The fitted controller model is included in the example analysis of Fig. 12 for visualisation of the precision of the fitted model with according coefficient of determination R^2 included. Fig. 13 shows the subsequent coefficients of determination as a measure for quality-of-fit for all participants in the abstract environment *E1*. The adjustable parameters of Eq. (6) to best fit the identified FRFs are presented in Fig. 14 for all display conditions and participant groups in the abstract environment *E1*. Fig. 15 shows the model parameters to best fit the control



(a) Magnitude *FP* **(b) Magnitude *D-Mono*** **(c) Magnitude *D-Stereo*** **(d) Magnitude *D-Hyper***

(e) Phase *FP* **(f) Phase *D-Mono*** **(g) Phase *D-Stereo*** **(h) Phase *D-Hyper***

Fig. 12: Illustrations of identified H_{p_e} frequency response functions and modelled H_{p_e} control dynamics for a single participant for all four display conditions (a & e) Flat-plane, (b & f) Monoscopic depth, (c & g) Stereoscopic depth and (d & h) Hyper-stereoscopic depth in the abstract environment *E1*.

dynamics of *PG2* in the realistic environment *E2*.

The frequency response functions are estimated at the frequencies of the target and disturbance signals, i.e., f_t (indicated with a white circle) and f_d (indicated with a black diamond), respectively. Any divergent response behavior with regard to either signal compared to the other may be pointed out by means of Fig. 12 . For every display condition, the identified FRFs are plotted by the magnitude (Fig. 12a-d) and phase (Fig. 12e-h). Subsequently, the determined model parameters to best fit the identified FRFs in Fig. 14 for *E1* and Fig. 15 for *E2* show the medians in red, the 25^{th} and 75^{th} percentiles as the solid bottom and top edges of the box while the dashed lines represent the ranges of most extreme values and outliers are marked by a red '+'. In Fig. 14 each pair of boxes per display conditions consists of the model parameter for *PG1* as the left box and parameters for *PG2* as the right box. Similarly, the pairs in Fig. 15 consists of model parameters in the abstract environment *E1* (left box) and the realistic environment *E2* parameters (right box).

Note from Fig. 15 that only the model parameters for *D-Hyper* were determined in the realistic environment *E2*. The other display conditions tested within *E2* were shown to result in lower correlations of control input u with the target and disturbance signals f_t and f_d, respectively. Since the response of u in *D-Mono* and *D-Stereo* to either f_t or f_d cannot be distinguished from noise as clearly as *D-Hyper* or especially *FP* given the lower signal-to-noise ratios, identification of the FRFs and subsequent parameter estimation to fit a model to those response functions is deemed inaccurate. Therefore only the model parameters for *D-Hyper* in *E2* are estimated and presented in Fig. 15.

In general the plots in Fig. 12 firstly show an accurate fit of the H_{p_e} model of Eq. (6) to the identified FRFs for all display conditions. Previous studies have already shown a human controller to not necessarily adopt pursuit behavior (the second level of Successive Organization of Perception [28]) in tracking tasks with pursuit display types comprising single integrator controlled element dynamics ($H_c = K/s$) [6], [7], [13], [17]. The absence of low-frequency phase leads among the FRFs in Fig. 12 as well as the capability to fit Eq. (6) to the FRFs in these experiments are thus in accordance with the findings of those studies. Additionally, the identified FRFs in Fig. 12 at the frequencies of f_t and f_d show no dissimilar trends in the dynamics of H_{p_e}. The similar characteristics of the FRFs therefore suggest that participants did not incorporate additional responses other than to the control error e and thus adopting a compensatory control behavior [7], [8].

The spread of identified FRFs at the frequencies of f_t and f_d, although their trend over the range of frequencies appears identical, seems to increase. This is for instance notable comparing Fig. 12a with *FP* to Fig. 12b with *D-Mono*. This development is likely linked to the reduced correlation between control input u and the target and disturbance

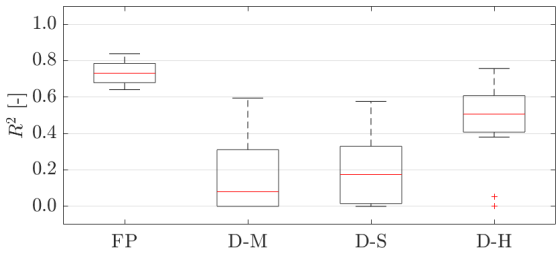
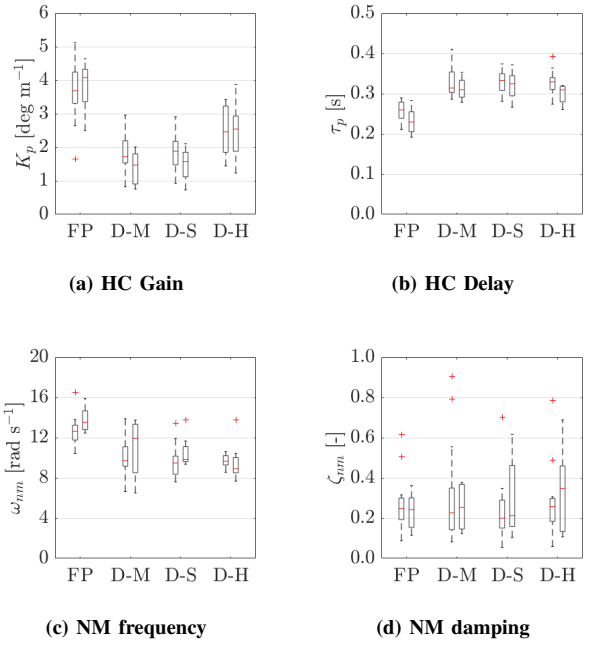


Fig. 13: Human controller modeling accuracy in coefficient of determination R^2 as quality-of-fit of modeled response to measurement data for the four display conditions of all participants in *E1*.

signals in *E1* already shown earlier in Fig. 10a-d. The criteria for inclusion of identified FRF to either f_t, f_d or both in the estimation of the model parameters is therefore based on adequate correlation of u ($\rho^2_{u_{t,d}}(j\omega_{t,d}) \geq 0.8$ at the lower frequencies) with f_t and/or f_d. Furthermore, as the signal-to-noise ratio degrades so does the accuracy of human controller modeling techniques as shown in Fig. 13. The modeling accuracy indicated by R^2 is highest for the classic flat-plane display condition at $R^2 \approx 0.75$. Subsequent depth display conditions show lower median modeling accuracy as well as an increase in spread of values, especially display conditions *D-Mono* and *D-Stereo*. In multiple occasions (29 fits for *D-Mono*, 27 fits for *D-Stereo* and 7 fits for *D-Hyper*) the minimum accuracy (see Eq. (7)) was attained.



(a) HC Gain **(b) HC Delay**

(c) NM frequency **(d) NM damping**

Fig. 14: Human controller modeling in model parameters (a) HC gain, (b) HC delay, (c) NM frequency and (d) NM damping for *PG1* (left bars) and *PG2* (right bars) in *E1*.

The human controller model included in Fig. 12 comprises four parameters that are estimated per display condition and participant group. The parameters human controller gain, delay time, neuromuscular frequency and damping are firstly presented for the abstract environment *E1* in Fig. 14a-d, respectively. Out of the four model parameters, only the neuromuscular damping appears equivalent (insignificant changes, see Table V) at an average value of $\zeta_{nm} \approx 0.25$ for all display conditions and both participant groups.

TABLE V: Human controller model parameters ANOVA results.

		df	F	sig.
K_p	Display	(1.9, 41.6)	128.2	0.00
	PG	(1, 22)	0.5	0.50
	Dsp.*PG	(1.9, 41.6)	2.0	0.15
τ_p	Display	(2.0, 44.8)	42.9	0.00
	PG	(1, 22)	1.3	0.27
	Dsp.*PG	(2.0, 44.8)	0.6	0.56
ω_{nm}	Display	(2.4, 53.4)	21.6	0.00
	PG	(1, 22)	5.6	0.03
	Dsp.*PG	(2.4, 53.4)	0.6	0.56
ζ_{nm}	Display	(2.7, 58.5)	1.3	0.29
	PG	(1, 22)	0.0	0.90
	Dsp.*PG	(2.7, 58.5)	2.3	0.09

The plot in Fig. 14a shows that *PG1* and *PG2* firstly appear to adopt equivalent control gains. Between the participants groups, no significant differences are found. Both participants controlled with the largest control gain K_p in *FP*. Participants achieved a median value of $K_p \approx 3.7$ to 4.1 deg/m. Differences in the magnitude of control gain K_p are notable with comparison of *FP* with the three depth display conditions. Firstly, on *D-Mono*, the identical task of *FP* as a depth tracking task with monoscopic display condition, both *PG1* and *PG2* reduce their magnitude of response significantly ($F(1, 22) = 235.9, p < 0.01$) by 53% and 64%, respectively. Given the visually scaled-down stimulus due to a change in control direction of *D-Mono*, the human controller appears to use a less aggressive control strategy which was also pointed out in previous studies [29], [30]. As stereopsis with human interocular distance at 18 m distance is introduced in *D-Stereo*, the control gains remain equivalent to *FP*. The hyper-stereoscopic display condition *D-Hyper* increases the gains compared to *D-Stereo* for *PG1* and *PG2* with significant increases ($F(1, 22) = 66.7, p < 0.01$) of 0.58 deg/m and 0.97 deg/m, respectively. Participants in both groups thus increase their control gain to reduce the error. The values for K_p have recovered significantly with *D-Hyper* from the gain reduction found when moving from flat-plane tasks to monoscopic and natural stereoscopic depth tracking tasks yet remain significantly lower ($F(1, 22) = 54.1, p < 0.01$) compared to *FP*.

The second parameter, the human controller delay time, is presented for all display conditions and participant groups in Fig. 14b. The delay time estimated of the FRFs with *FP* of $\tau_p \approx 0.23$ to 0.26 s was found to equal earlier estimated delay times in different experiments [5], [7], [31].

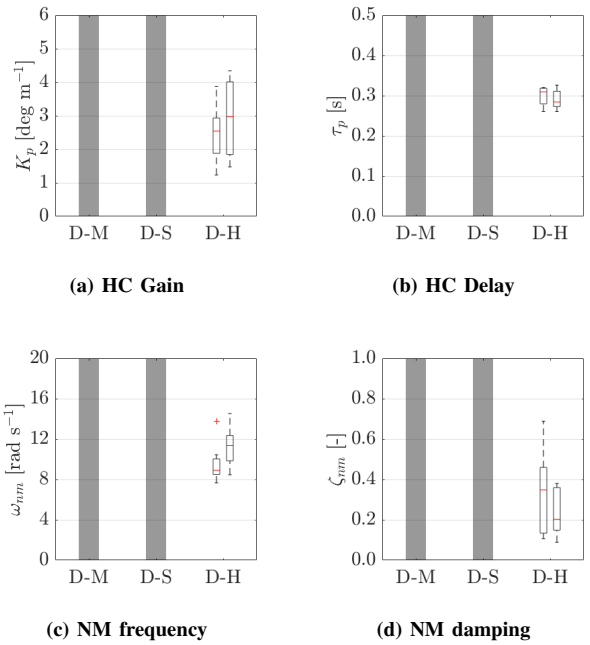
When participants performed the identical tracking tasks not in *FP* but in *D-Mono*, the delay time rose significantly ($F(1,22) = 129.2, p < 0.01$) for both *PG1* and *PG2* by 54 ms and 80 ms, respectively. However, contrary to the control gain K_p in Fig. 14a, the delay time appears to be rather constant at $\tau_p \approx 0.31$ to 0.33 s for the three depth display conditions for both participant groups. Although the quantity of stereopsis increases with *D-Stereo* and *D-Hyper* and therefore the minimum perceivable distance between the target and follower decreases, the delay time appears to be unaffected by the strength of the depth cue as no significant differences are found between *D-Mono*, *D-Stereo* and *D-Hyper*. These plots in Fig. 14a suggest stereopsis does not support the human controller in generating faster responses to either the target or disturbance signals yet also does not increase the required time to process the visual image. It appears that perception of more complex distances takes longer, as also demonstrated in previous research [31], [32].

The last parameter is the neuromuscular frequency presented in Fig. 14c and shows, much like the delay time in Fig. 14b, a significant difference ($F(1,22) = 77.9, p < 0.01$) between flat-plane and depth display conditions and no apparent trend between the three depth display conditions. The highest median value of ω_{nm} is found for *FP* at 12.6 rad/s by *PG1* and 13.5 rad/s by *PG2*. Subsequent depth display conditions then yield an average median value of $\omega_{nm} \approx$ 9.5 rad/s irrespective of availability/quantity of stereopsis (except for *PG2* with *D-Mono*). The reduced neuromuscular frequency appears to be linked to a difference in axis of movement much like the delay times in Fig. 14b.

The model parameters in Fig. 15 lastly present a visual comparison of the difference in human controller model between abstract environment *E1* in the left bar and realistic environment *E2* in the right bar. The non-modeled parameters for *D-Mono* and *D-Stereo* are grayed out. Firstly, although the spread of estimated values for K_p appears larger for *E2*, the median values differ insignificantly at $K_p \approx 2.5$ deg/m in *E1* and $K_p \approx 3.0$ deg/m in *E2*. Equivalent increase in spread and reduction in median value can be made for the delay time in Fig. 15b. Fig. 15c does present a notable difference between the spread of values as well as the median values of ω_{nm} between the abstract and realistic environment at $\omega_{nm} \approx 8.9$ rad/s and $\omega_{nm} \approx 11.4$ rad/s, respectively. Nonetheless, the estimated parameters in Fig. 15 provide no basis to believe a significantly different control behavior was adopted to complete the tracking tasks in the abstract and realistic environments.

IV. Discussion

In the experiment, the effects of control direction and stereoscopic depth cueing on manual control behavior was investigated from a cybernetics perspective. The goal of this paper was to identify, on one hand, the effect of a control axis shift from within the viewing plane to the depth axis and, on the other hand, the response of the human controller to stereoscopic three-dimensional display as modern interfaces within depth tracking control tasks.



(a) HC Gain **(b) HC Delay**

(c) NM frequency **(d) NM damping**

Fig. 15: Human controller modeling in model parameters (a) HC gain, (b) HC delay, (c) NM frequency and (d) NM damping for *PG2* in *E1* (left bars) and *E2* (right bars) for the *D-Hyper* display condition.

The alteration of control direction from flat-plane movement in display condition *FP* to depth axis movement in *D-Mono* was found to strongly affect the control behavior of both participant groups. It was hypothesized (**H1**) that, given identical control tracking task parameters and equivalent optical characteristics with flat-plane tasks, degradation of task performance would be found for a human controller tracking a target in depth. The variance in error indeed revealed a significant reduction in task performance. Estimation of the model parameters to fit the identified frequency response functions indicates that the human controller adopts a significantly lower control gain and increased delay time in depth tracking. Given the reduced display scaling of distances and lower accuracy of subjective distance perception along the operator's line of sight [2], the results adhere to earlier studies indicating a reduced gain and increased delay time for down-scaled visual stimili in compensatory tracking tasks [29], [30]. Also the control error over the movement range of the target was shown to increase with the target moving further away from the point of observation. Furthermore, the control behavior was found to vary with both position of the target signal and the frequencies of the target and disturbance signals. The correlation coefficient shows an overall lower coherence with the target as well as a decline at higher frequencies. The hypothesized degradation of task performance, especially at both lower and higher frequencies, can however only be confirmed for higher frequency ranges of the signal and therefore **H1** can only partially be accepted.

A comparison of the depth display conditions firstly reveals a significant increase in both task performance and control effort by both participant groups with hyper-stereoscopic vision in display condition *D-Hyper* with respect to monoscopic display condition *D-Mono*. Although the coherence of control input with the target and disturbance signals and inconsistency of error with target position remain, estimated model parameters indicate more aggressive responses by participants with hyper-stereoscopic vision available. Although performance increased significantly with only stereoscopic vision, task performance with hyper-stereoscopic vision is found to also significantly increase over natural stereoscopic vision. Superiority of hyper-stereoscopic vision on objects at beyond-natural stereoscopic range was also found in research focused on increasing control accuracy for beyond-natural stereoscopic ranges [16]. Given the viewing distance in this experiment of the observer to the target and follower elements reaching beyond the general effective natural stereoscopic range [33], the equivalent task performance between human interocular distance and no eye separation is in accordance with previous studies on stereoscopic depth perception [3], [33]. Simultaneously, the delay time remains equivalent independent of stereopsis depth cue strength. These findings comply with literature reporting not necessarily lower response times with stereoscopic versus non-stereoscopic displays [3]. The consistently higher delay times for tracking tasks with depth control therefore suggest the perception and process simply takes longer for depth tracking tasks in general, irrespective of (hyper)stereo-scopic vision. Such effects have for instance also been identified for more complex movement in multi-axis control tasks [31], [32]. The hypothesized task performance increase of **H2** is thus accepted for this experiment's hyper-stereoscopic vision condition by an overall higher coherence of control input with the target and disturbance and higher adopted control gains. Furthermore, even though the expected increase in task performance was found especially with the hyper-stereoscopic display conditions, the significant difference between flat-plane tracking tasks and depth tracking tasks remains irrespective of quantity of stereopsis provided. This is in agreement with Wickens' statement [2] on three-dimensional environment presentation on stereoscopic displays: *"3D displays create the potential for perceptual ambiguity which is not present in a set of orthogonal 2D displays."*

The frequent use of stereoscopic three-dimensional displays was hypothesized to increase task performance on depth control tracking tasks with (hyper)stereoscopic displays (**H3**). The results of task performance and control effort firstly indicate no difference was found between the two participant groups for depth tracking tasks with hyper-stereoscopic vision. In fact, at lesser quantities of binocular disparity the task performance was found to degrade more for the group frequently using stereoscopic displays compared to the group inexperienced with stereoscopic displays. Research into the beneficial addition of stereoscopic vision found the increase in task performance to heavily rely on the difficulty of task and whether a task was well-learned [3]. This experiment indicates a reverse analogy; frequent usage of stereoscopic displays appears to reduce task performance on depth control tasks without the availability of hyper-stereoscopic vision. It must be noted, however, that the comparison between the two groups must be considered carefully. The effect of task performance degradation with non-hyper-stereoscopic vision may emphasize the reduction of task performance due to higher age as the increased task performance due to experience with hyper-stereoscopic disappears. Future work with better age-matched groups may help better conclude on the effect of experience with stereoscopic three-dimensional displays. Nonetheless, based on the experiment results, **H3** is rejected.

To assess the applicability of results from theoretical experiment onto more practical applications, results in more visually advanced and realistic environment were considered and found to significantly differ for non-hyper-stereoscopic vision displays. However, the differences require careful consideration of the cause of differences in task performance and control behavior. It turned out participants showed different behavior primarily due to lower coherence with the target when low amounts of binocular disparity was shown or none at all. The considered control task to minimize the error between the target and follower therefore changed as participants no longer based control inputs on their target. In a pursuit display, uncertainty then rises on what is considered the new target to track. Coherence with the disturbance signal namely (although overall lower compared to findings in the abstract environment) persists. The availability of information on the refueling boom extension state on the TRVS HUD may have given participants a self-sought reference point to which to steer the follower element. However, the HUD only informs about x and not f_t. Nonetheless, it must be stated that the behavior differs not as a result of a more visually advanced and realistic environment, but rather the reduced possibility of parameter isolation in more practical applications. On the other hand, had the TRVS HUD been disabled in the realistic environment, participants would most likely control the follower at either limit of the extension range of the refueling boom omitting any possibility of control behavior identification. The inclusion of the realistic environment in this experiment, although providing no support for the applicability of results onto more practical applications (and therefore requiring rejection of **H4**), shows once more the difficulty of capturing realistic operational control behavior in isolated experimental environments. Additional research into the effects of depth cue isolation from visual environments on manual control behavior can proof useful for future experiment designs focusing on depth tracking tasks with (hyper-)stereoscopic vision.

The implication of studied circumstances in this experiment must be considered carefully to assess generalizability of the reported findings. It is important to realize the experiment was conducted in a teleoperative setting with artificial stereoscopic vision. First of all, the generalizability of results to applications exceeding this research's visual

parameters is uncertain due to optical characteristics such as object distances, vergence angles and possible stereo camera misalignments. These may seriously affect control behavior and adaptation and experiment variations to this paper's experiment could help provide better insight into the effects these optical characteristics may have. Furthermore, the implementation of artificial stereoscopy in itself may also seriously affect the perception of depth and subsequent control behavior in depth tracking tasks given the conflict of depth cues accommodation and convergence with artifical stereoscopic vision [34], [35]. The conflicting depth cues could potentially influence control behavior and adaptation to depth tracking tasks with artificial (hyper-)stereoscopic vision as tested in this paper's research. Additional experiments could be conducted to further extend the applicability of manual control behavior knowledge to natural stereoscopic vision depth tracking tasks, perhaps increasing interocular distances to create natural hyper-stereoscopic vision. A comparison of results between depth tracking tasks with natural and artificial (hyper-)stereoscopic vision may provide an understanding of the impact of the physiological phenomenon of conflicting depth cues on manual control behavior.

It must also be mentioned that the selected forcing functions may have provided a too oscillatory target to follow and object to control. This is thought especially true for the tracking task in a more realistic environment with more realistic tracking tasks based on e.g. aerial refueling. The design of this part of the experiment was based on earlier work with identical forcing functions to provide a basis for comparison. However, the results of this experiment in conjunction with literature on depth perception and depth tracking accuracy suggest the selection of forcing functions should potentially be adjusted to comprise less high-frequent sinusoidal signals when considering tasks concerned with depth tracking.

Lastly, the magnitude of error appears to significantly grow with increasing distance between the human controller and the target with the depth display condition, i.e., contrarily to tasks on classical flat viewing planes, the magnitude of error attained in depth tracking tasks is not consistent over the considered range of positions. The trend of error magnitude with distance to the observer complies with research into perception of distances along the depth axis [21]. Although studies on perspective scaling of distances have shown alteration of control behavior to compensate for its effect [23], [22], the results of this experiment show no support for adoption of control behavior that (fully) compensates for and reverses the effects of subjectively smaller appearing errors at larger distances from the point of observation. These results could potentially imply a non-consistent control response of the human controller to identical magnitudes of error at different distances between the operator and the target. The influence of participants' individual stereoacuity thresholds may turn out to be a dependent parameter. It may be useful for subsequent studies on control theory with depth tracking tasks to analyse e.g. task performance and control gain measures not as a function of objective distances but rather as a function of perceived visual angles. Although this experiment clearly shows an upward trend of control error with distance, future work into analysis of control gain K_p with varying distances could be fundamental for following steps into understanding and modeling human control behavior in depth tracking tasks.

V. Conclusion

With a cybernetics approach, this paper aimed to identify and model divergent manual control behavior in depth control tasks with (hyper-)stereoscopic vision by means of a two-part human-in-the-loop experiment. The results contribute to current knowledge on McRuer et al.'s operator model [5] to accurately describe and predict manual control behavior of a human controller tasked with depth tracking to ultimately provide a better understanding of the human controller in a teleoperative environments with artificial (hyper-)stereoscopic vision [17].

A change of movement orientation from within a two-dimensional orthogonal viewing plane to the depth dimension was proven to strongly affect human capability to track a target. Significant degradation of task performance in depth tracking tasks was verified as a result of altered control behavior with less aggressive responses and increased delay times. Hyper-stereoscopic vision on depth tracking tasks at ranges beyond natural stereoscopic vision was found to be superior to monoscopic and stereoscopic vision, although the implication of visual perception of distances along the third dimension remains and degradation of task performance due to altered control direction therefore cannot be completely overcome. The comparison of control behavior in a visual highly isolated environment with a more realistic and practical environment moreover emphasized the delicate human adaption of behavior and response to visual stimili to be highly dependent on the extent of visual isolation. Overall, the experiments gave initial yet fundamental insights into human manual control to ultimately understand from a cybernetics perspective the behavior in depth tracking tasks and adaptation of human controllers to nowadays' modern stereoscopic three-dimensional interface technologies.

References

[1] T. B. Sheridan, "Teleoperation, Telerobotics and Telepresence: A Progress Report," *Control Engineering Practice*, vol. 3, no. 2, pp. 205 – 214, Feb. 1995.

[2] C. D. Wickens, "Three-dimensional Stereoscopic Display Implementation: Guidelines Derived from Human Visual Capabilities," in *Stereoscopic Displays and Applications*, J. O. Merritt and S. S. Fisher, Eds., vol. 1256, International Society for Optics and Photonics. SPIE, 1990, pp. 2 – 11.

[3] J. P. McIntire, P. R. Havig, and E. E. Geiselman, "Stereoscopic 3D Displays and Human Performance: A Comprehensive Review," *Displays*, vol. 35, pp. 18 — 26, 01 2014.

[4] M. D. Winterbottom, C. Lloyd, J. P. Gaska, S. Wright, and S. Hadley, "Stereoscopic Remote Vision System Aerial Refueling Visual Performance," *Electronic Imaging*, vol. 2016, pp. 1–10, Feb. 2016.

[5] D. T. McRuer, D. Graham, E. S. Krendel, and W. J. Reisener, "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 (OH), Tech. Rep. AFFDL-TR-65-15, 1965.

[6] R. J. Wasicko, D. T. McRuer, and R. E. Magdaleno, "Human Pilot Dynamic Response in Single-loop Systems with Compensatory and Pursuit Displays," Air Force Flight Dynamics Laboratory, Tech. Rep. AFFDL-TR-66-137, Dec. 1966.

[7] M. C. Vos, D. M. Pool, H. J. Damveld, M. M. van Paassen, and M. Mulder, "Identification of Multimodal Control Behavior in Pursuit Tracking Tasks," in *Proceedings of the 2014 IEEE International Conference on Systems, Man, and Cybernetics, San Diego (CA)*, Oct. 2014, pp. 69–74.

[8] M. Mulder, D. M. Pool, D. A. Abbink, E. R. Boer, P. M. T. Zaal, F. M. Drop, K. van der El, and M. M. van Paassen, "Manual Control Cybernetics: State-of-the-Art and Current Trends," *IEEE Transactions on Human-Machine Systems*, vol. 48, no. 5, pp. 468–485, Oct. 2018.

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

[10] A. J. Grunwald and S. J. Merhav, "Vehicular Control by Visual Field Cues," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 6, no. 12, pp. 835–845, Dec. 1976.

[11] M. Mulder and J. A. Mulder, "Cybernetic Analysis of Perspective Flight-Path Display Dimensions," *Journal of Guidance, Control, and Dynamics*, vol. 28, no. 3, pp. 398–411, May-June 2005.

[12] M. H. H. Kemna, D. M. Pool, M. Wentink, and M. Mulder, "Manual control behavior in stereoscopic vision-enhanced depth control tasks," in *AIAA Scitech 2020 Forum: 6-10 January 2020, Orlando, FL*, 01 2020.

[13] R. A. Hess, "Pursuit Tracking and Higher Levels of Skill Development in the Human Pilot," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. SMC-11, no. 4, pp. 262–273, 1981.

[14] J. C. P. Bol, L. van Breda, and M. Ringers, "Tanker Remote Visual System (TRVS)," TNO, Tech. Rep. DenV S070047, 2007. [Online]. Available: https://www.tno.nl/en/focus-areas/defence-safety-security/roadmaps/information-sensor-systems/tanker-remote-vision-system/

[15] M. M. van Paassen and M. Mulder, "Identification of Human Operator Control Behaviour in Multiple-Loop Tracking Tasks," in *Proceedings of the Seventh IFAC/IFIP/IFORS/IEA Symposium on Analysis, Design and Evaluation of Man-Machine Systems, Kyoto Japan*. Pergamon, Sep. 1998, pp. 515–520.

[16] P. Bijl, S. De Vries, and J. Kalthof, "Upgrading the KDC-10 RARO Refueling Vision System: WP3b," TNO, Confidential Internal Report TNO-DV3 2005-A 161, 2005.

[17] M. Mulder, D. M. Pool, K. van der El, F. M. Drop, and M. M. van Paassen, "Manual Control with Pursuit Displays: New Insights, New Models, New Issues," in *Proceedings of the 14th IFAC Symposium on Analysis Design and Evaluation of Human Machine Systems Tallinn, Estonia*, 2019, pp. 139–144.

[18] A. van Lunteren, "Identification of Human Operator Describing Function Models with One or Two Inputs in Closed Loop Systems," Ph.D. dissertation, Delft University of Technology, Faculty of Mechanical Engineering, Jan. 1979.

[19] C. D. Wickens, S. Todd, and K. Seidler, "Three-Dimensional Displays: Perception, Implementation, and Applications," University of Dayton Research Institute, Tech. Rep. CSERIAC-SOAR-89-001, Oct. 1989.

[20] D. T. McRuer and H. R. Jex, "A Review of Quasi-Linear Pilot Models," *IEEE Transactions on Human Factors in Electronics*, vol. HFE-8, no. 3, pp. 231 – 249, Sep. 1967.

[21] S. Nagata, *How to Reinforce Perception of Depth in Single Two-Dimensional Pictures*. USA: Taylor & Francis, Inc., 1991, p. 527–545.

[22] K. van der El, D. Pool, M. M. Van Paassen, and M. Mulder, "Effects of linear perspective on human use of preview in manual control," *IEEE Transactions on Human-Machine Systems*, vol. PP, pp. 1–13, 08 2017.

[23] B. Sweet, "A model of manual control with perspective scene viewing," in *AIAA Modeling and Simulation Technologies (MST) Conference*, Aug. 2013.

[24] J. Lee and A. McIntyre, "Clinical Tests for Binocular Vision," *Eye*, vol. 10, p. 282–285, 1996.

[25] S. Lee and N. Koo, "Change of Stereoacuity with Aging in Normal Eyes," *Korean journal of ophthalmology : KJO*, vol. 19, pp. 136–9, 06 2005.

[26] H. R. Jex, R. E. Magdaleno, and A. M. Junker, "Roll Tracking Effects of G-vector Tilt and Various Types of Motion Washout," in *Proceedings of the Fourteenth Annual Conference on Manual Control*, 1978, pp. 463 – 502.

[27] J. A. Karasinski and S. K. Robinson, "Evaluating Augmented Reality in a Three-Axis Manual Tracking Task," in *Proceedings of the AIAA Modeling and Simulation Technologies Conference, San Diego (SA)*, no. AIAA-2019-1227, Jan. 2019.

[28] E. S. Krendel and D. T. McRuer, "A Servomechanics Approach to Skill Development," *Journal of the Franklin Institute*, vol. 269, no. 1, pp. 24–42, 1960.

[29] W. H. Levison and R. Warren, "Use of linear perspective scene cues in a simulated height regulation task," in *Proceedings of the 20th Annual Conference on Manual Control*, 1984.

[30] S. W. Breur, D. M. Pool, M. M. Van Paassen, and M. Mulder, "Effects of displayed error scaling in compensatory roll-axis tracking tasks," in *AIAA Modeling and Simulation Technologies Conference*, 2010.

[31] S. Barendswaard, D. M. Pool, M. M. van Paassen, and M. Mulder, "Dual-Axis Manual Control: Performance Degradation, Axis Asymmetry, Crossfeed, and Intermittency," *IEEE Transactions on Human-Machine Systems*, vol. 49, no. 2, pp. 113–125, 2019.

[32] W. H. Levison and J. I. Elkind, "Two-Dimensional Manual Control Systems with Separated Displays," *IEEE Transactions on Human Factors in Electronics*, vol. 8, no. 3, pp. 202–209, Sep. 1967.

[33] B. Sweet and M. Kaiser, "Depth Perception, Cueing, and Control," *AIAA Modeling and Simulation Technologies Conference*, 08 2011.

[34] J. P. Wann, S. Rushton, and M. Mon-Williams, "Natural Problems for Stereoscopic Depth Perception in Virtual Environments," *Vision Research*, vol. 35, no. 19, pp. 2731–2736, 1995.

[35] D. M. Hoffman, A. R. Girshick, K. Akeley, and M. S. Banks, "Vergence Accommodation Conflicts Hinder Visual Performance and Cause Visual Fatigue," *Journal of Vision*, vol. 8, no. 3, pp. 33–33, 03 2008.

II

LITERATURE STUDY

2

Introduction to Visual Perception

The nervous system is a fundamental part of many living organisms' bodies comprising numerous neural systems. A neural system is categorized as one of three types: sensory system, motor system or associational system [22]. Sensory systems and motor systems serve "baseline" functions of information representation and action generation, respectively. The associational systems, on the other hand, link the aforementioned two systems for "higher-order" functions like perception. Perception is fundamental to interacting with the environment. For human beings, the five basic senses (somatic sensation, vision, audition, vestibular sensation and chemical senses) perceive various forms of information through stimili either from within themselves or their surroundings [22]. The human visual system is such an integral neural system and quite essential to a human being's perception of its surroundings.

Figure 2.1: The visual system in the central nervous system [23]

This chapter describes the necessary process for a human being to be able to see. The visual system is taken into special consideration as the envisioned research assumes visual stimili as the single input to the human controller. Section 2.1 first briefly describes the flow of signals from the source of vision on one end towards visual interpretation of the external surroundings on the other. However, the eventual visual interpretation is predominantly based upon sensing stimili with two eyes in case of studying human beings as well as many animals. Section 2.2 therefore addresses the physiological phenomenon of binocular vision. Lastly, Section 2.3 focuses on the aspect of vision that allows for perceiving visual depth, i.e., third-dimensional information about the surrounding environment.

2.1. Breakdown of the Visual System

In 1950, James J. Gibson wrote a number of prerequisites on the visual system. These have to be met by an individual before it can see. Needless to say, the eye is inseparably linked to anyone's knowledge on the visual system, but Gibson's stated conditions reach further than merely the eye [24]:

- The environment must have light to see by;
- The eyes must be open;
- The eyes must focus and point properly;

- The sensitive film at the rear of each eyeball must react to light;
- The optic nerves must transmit impulses to the brain.

The first condition refers to the very source of the visual system. Light, which in terms of the human visual system should refer to visible light, lies at the absolute start of the process of seeing. On the other hand, the last condition refers to provision of signals to the brain before the eventual spatial perception as representation of the surrounding environment. The complete flow of signals within the visual system (see Figure 2.1) is a complex process involving a number of key aspects. Gibson refers to this process as the 'sequence of events in vision' [24]. This section will break down the visual system into three main components and elaborate on those key parts.

1. **The eye**: capturing and converting the surrounding's visible light to neural signals;
2. **The visual pathways**: distributing sensory and motory neural signals throughout the visual system;
3. **The cortical areas**: interpreting and integrating neural signals to visual perception and vision.

2.1.1. THE EYE

The eye, as the first medium in the visual system, serves two main purposes. The eye should of course provide its information to further sections of the visual system by converting visible light from one's surroundings to neural signals [24] . However, the visible light rays reflecting from physical surfaces in the outside world must properly project onto the retina. The eye should therefore firstly transmit and refract incoming light before useful transduction into neural signals takes place [22, 25]. Note that the eye is treated as a single instance in this section. Section 2.2 focuses on the implication of having two eyes (binocular vision).

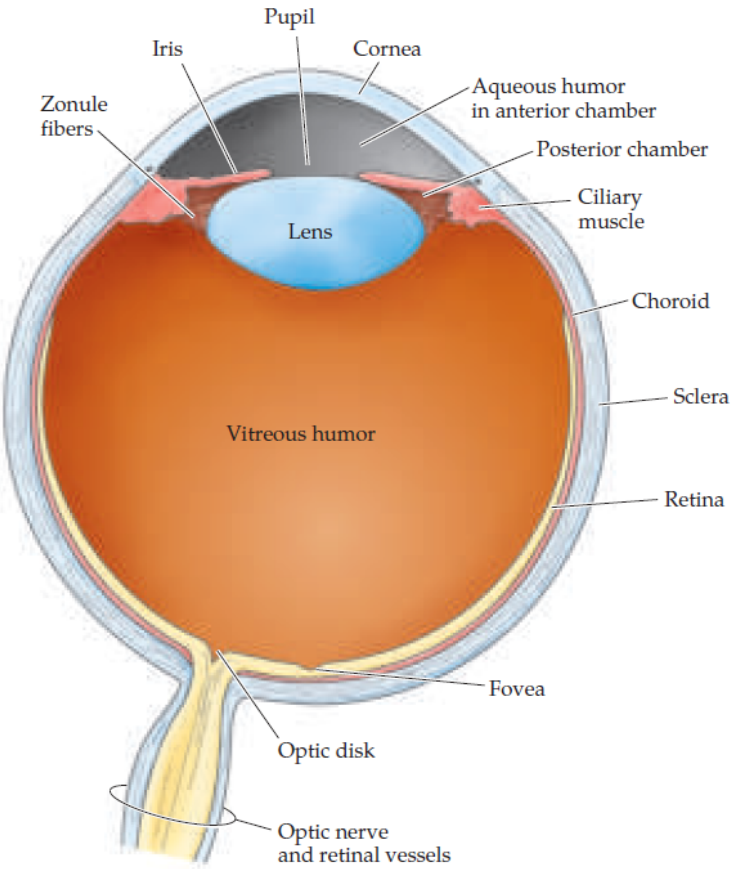


Figure 2.2: A schematic cross-section of the eye. The front of the eye is shown at the top of the figure, indicating e.g. the anterior and posterior chambers and optical components [22]

Anatomy of the Eye
The eye is an approximately sphere-shaped sense organ and comprises three distinct layers of tissue that define its particular volumes [22, 26]. Figure 2.2 provides a cross-section of the human eye with all its layers and volumes [26]:

- The outermost layer of tissue forms the cornea and sclera;
- The middle layer is composed of the choroid, the iris and ciliary muscle;
- The innermost layer at the rear of the eye is the retina;
- The anterior chamber;
- The posterior chamber;
- The vitreous chamber.

The outermost layer of the eye is partly seen when looking into the eyes. Its surface consists mainly of an opaque tissue called sclera; the white of the eyes [22]. The cornea at the frontal side of the eye (top side of the cross-section in Figure 2.2) is only a minor section of the outer layer. Contrary to the sclera, it is transparent and therefore permits light to enter [25]. Together, the cornea and sclera form the outermost layer of the eye and enfold three specific volumes. The first smaller volume is the anterior chamber and lies between the cornea on the one side and the eye's lens and iris on the other [22]. The second one, called the posterior chamber, is the cavity between the iris, the lens and the vitreous chamber. The choroid then covers the majority of the eye's inside surface at the sides and rear. Together with the ciliary body containing the ciliary muscle at the front, it forms the middle layer. This layer contains the largest of volumes called the vitreous chamber. The last layer is the innermost layer of the eye adjacent to the choroid called the retina and lies at the rear side of the eye [22, 24–26].

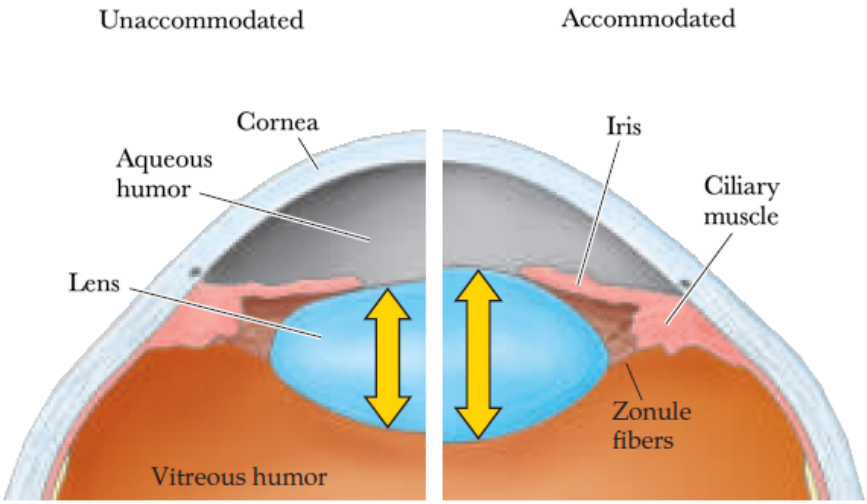


Figure 2.3: Schematic cross-sections focused on the front of the human eye. The fixed cornea is shown at the top of both cross-sections and the different shapes of the lenses for unaccommodated in the left and and accommodated in right cross-section [22]

Optical Pathway
One of the key purposes of the eye is to not only transmit but also properly focus light on the retina [24]. Its instruments are the cornea, the iris and the lens and together they form the optical components to achieve that purpose [22, 26]. The optical pathway starts with visible light from one's surroundings firstly passing through the cornea at the front of the eye (see Figure 2.2). Its tissue is transparent and manipulates rays of light due to its curved shape. The cornea is the primary component to refract light; approximately two-thirds of the total refraction. Note, however, that its refractive power is fixed and forms the image of the environment only roughly onto the retina [25]. On the other hand, the optical characteristics of the lens can be altered to very precisely refract the new bundle of light to focus the image clearly onto the retina [22, 24, 25]. Changes to the shape of the lenses are called accommodation and a cross-section of the different lens shapes can be seen in Figure 2.3. The corrective shaping of the lens is induced by balanced pairing of the contrary lens' natural elasticity and tensile forces excited by the ciliary muscle [22, 25]. More specifically, upon contraction of the ciliary muscle various

changes occur to increase lens power. For a start, the thickness of the lens increases and its width decreases. However, not only its shape changes. The front surface of the lens also moves forward, decreasing the anterior chamber's depth [26]. Figure 2.3 moreover shows how the iris covers the zonule fibers and part of the lens itself. The iris is commonly known to adjust its radial size as a means to control the light intensity within the eye. As the iris expands or contracts, the pupil increases or decreases in surface area to ultimately allow more or less light into the posterior chamber of the eye [25]. Although light intensity control is essential to the functionality of the retina, it is not the only purpose of pupillary light reflex. Controlled regulation of the size of the pupil by the iris adds to the eye's available set of instruments to clearly form an image on the retina [22]. Not only are the undesirable effects of spherical and chromatic aberrations reduced by covering the outer curvatures of the lens, also the depth of field can be increased with reduced pupil size. Therefore, a concession must be made to maximize image quality whilst assuring sufficient illumination of the environment.

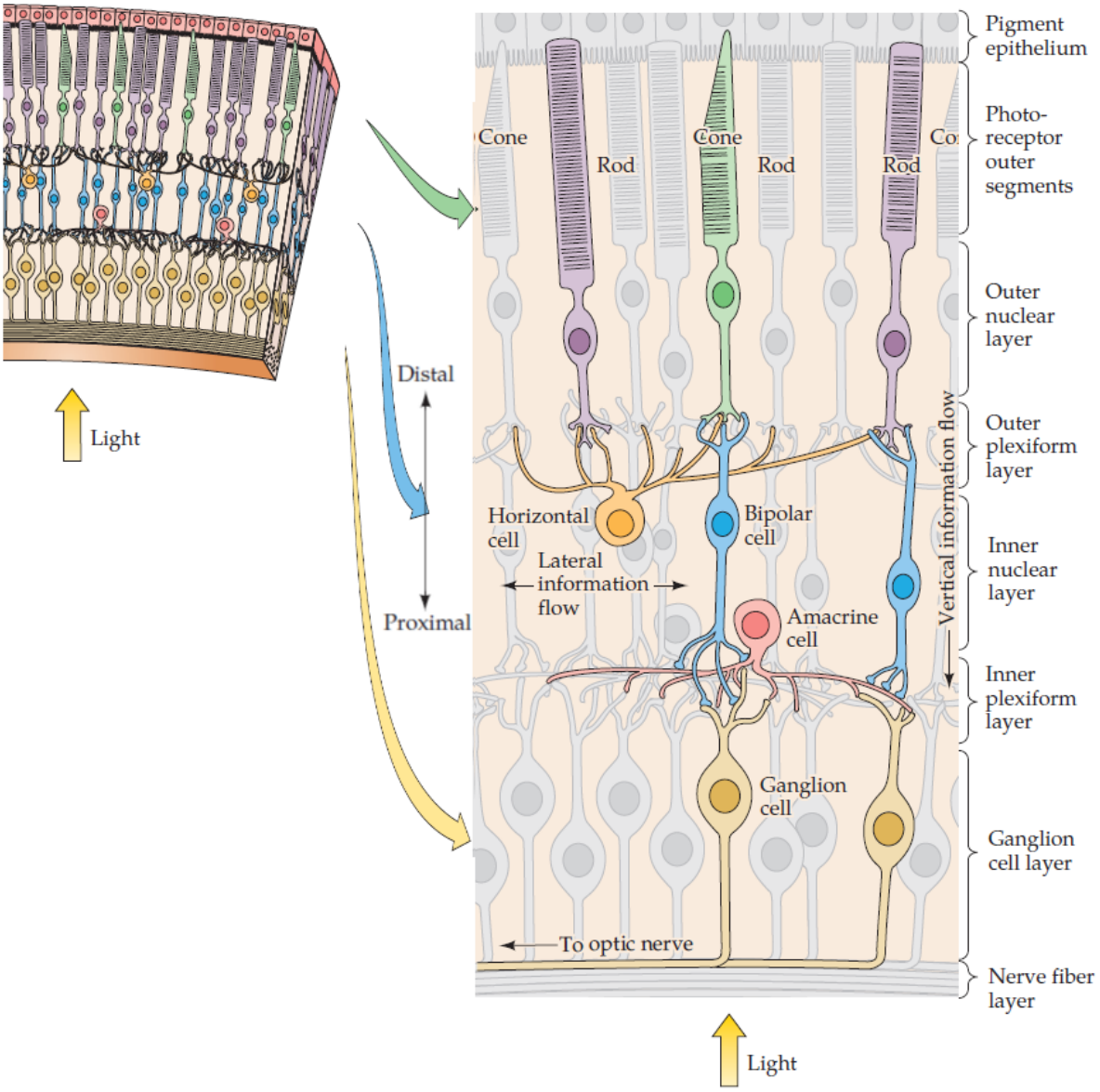


Figure 2.4: A schematic cross-section of the retina as the innermost layer of the eye. Light rays fall upon the retina at the *bottom side* of the sketch passing through various layers of the neural layer. The five types of neurons are highlighted with different colors [22]

The Retina
Light from the environment has at this stage been altered to form a clearly focuses image on the retina. All processes described before are an outcome of optical components in the optical pathway of the eye. Only now begins the actual neural system as the retina is the only component within the eye that is part of the central nervous system [22]. The neural layer converts energy of incoming light rays by means of complex biochemical processes into neural signals [26]. The composition of the retina is shown with a cross-sectional sketch in Figure 2.4. Visible light reaches the retina from the lower side of the figure and passes numerous types of neurons: ganglion cells, amacrine cells, bipolar cells, horizontal cells and photoreceptors [22, 26]. The photoreceptors, bipolar cells and ganglion cells form the main supply of information towards the optic nerve and further inward to the central nervous system, specifically in that order [22]. Indeed the energy firstly passes the ganglion and bipolar cells as light energy before reaching the photoreceptors of the retina. A distinction can be made between the neuron chain of photoreceptor, bipolar cell and ganglion cell as vertical information flow on the one hand, and the horizontal and amacrine cells as lateral information on the other (see Figure 2.4 for the information flow directions).

At the start of flow of information in the neural domain, the photoreceptors serve as the source of neural signaling. Photopigments absorb photons of light that eventually alter the rate of neurotransmitter release [22, 24–26]. The retina holds two sorts of photoreceptors. Rod and cones differ by, among others, the photopigment they contain and their distribution across the retina's surface. Rods have low resolution densities yet are highly sensitive photoreceptors whereas cones have a very high resolution density for a loss of sensitivity [22]. The retina contains rods distributed only in the peripheral zones and cones most highly concentrated in the fovea. The total amount of rods greatly exceeds the amount of cones [22, 25, 26]. The density of cones in the fovea ranges between 100,000 and 324,000 per mm^2 whilst further into the peripheral regions the concentration drops to 6,000 per mm^2. The opposite holds for rods; no rod-type photoreceptors appear in the fovea and their density rises to roughly 160,000 per mm^2 in the peripheral retina [25]. The fovea is thus a special section of the retina of only 1.5 mm in diameter (see Figure 2.2) and is especially well suited for distinction in detail or high visual acuity [22, 25, 26]. In fact, motivation for accommodation follows from retinal blur. The lens' shape changing is thus initiated by a stimulus following primarily from cones [26]. However, as only such a small part of the retina may achieve such high visual acuity, eye movements are constantly required to maintain a clear image of a particular object in the environment [22, 27]. Parts of eye movements, especially those associated with binocular vision, will be discussed in Section 2.2.

The remaining sequence of information transmission passes the bipolar cells and ganglion cells. However, information flows not only strictly towards the optic nerve and onto the brain itself. Section 2.1.2 addresses the visual pathways behind the eye and it will become clear information signals run back and forth between the eye and the brain. Already in the retina itself such signalling takes place [22, 26]. The interconnection of horizontal, bipolar, amacrine and ganglion cells enable the retina to regulate and maintain various visual system's aspects and distribute information within the retina itself [22, 25, 26]. Neural signals that form the base for vision and visual interpretation of surroundings environment leave the eye through the optic nerve (see Figures 2.2 and 2.4). The bundling of all ganglion cell nerve fibers before the optic nerve takes place as indicated by the optic disk in Figure 2.2. There are no photorecepters in this point of the retina and is therefore often known as blind spot [22].

2.1.2. THE VISUAL PATHWAYS

The optic nerves leave the eyes at the rear and slightly nasal sides. Figure 2.5 shows a cross-sectional sketch and the continuation of the optic nerves with regard to Figure 2.2. The visual pathways moderate various key functions running both from and towards the eyes. Figure 2.5 for instance indicates the travelled path of neural signals running from the retina towards various destinations in the midbrain. On the contrary, pupil reflexes are triggered in the midbrain and run back towards the eyes as shown in Figure 2.6 [22]. Some visual pathways and various targets are hereafter briefly enumerated and described.

Starting just aft of the rear side of the eye, the optic nerves from both eyes run towards the optic chiasm (see Figure 2.5). The optic chiasm can be seen as a crossroads of nerve fibers. The optic tract, running deeper into the midbrain contains fibers from both the left and right eye as these have been distributed in the chiasm [22]. Figure 2.5 merely indicates the paths taken by the eye on the left side of the figure, but a similar opposite path can be drawn for the other eye. The visualised path indicates four target structures. The most predominant target is

the lateral geniculate nucleus (or LGN); around 90 % of the ganglion cell nerve fibers from both eyes terminate in one of two LGN and information is relayed to higher cortical areas through the optical radiation capsules [22, 25, 26]. Further elaboration on the cortices can be found in Section 2.1.3.

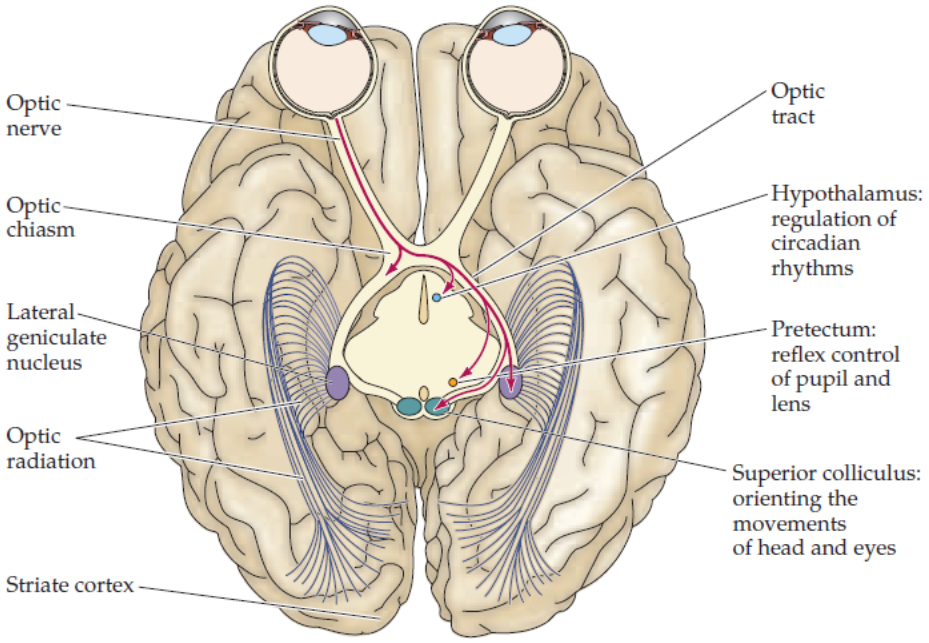


Figure 2.5: A cross-sectional sketch of the central visual pathways between the eyes and various regions of the human brain [22]

Other targets of the ganglion cell nerve fibers are the hypothalamus, pretectum and superior colliculus [22]. The superior colliculus regulates movements of both the eye and head to visual objects in the environment [27]. Some of those movements will be covered more thoroughly in Section 2.2. The hypothalamus is a structure involved with, in terms of the visual system, the regulation of circadian rhythms, i.e., the cycle of day and night by means of fluctuation in light intensities [22, 26]. The last structure, the pretectum, is likewise involved with monitoring and regulating light intensities, however with a much quicker response time. The pretectum regulates the pupillary light reflexes. The diameter of the pupil can be altered as discussed in Section 2.1.1 and the various steps can be found in Figure 2.6. Ganglion cells' nerve fibers carry information from the retinas in both eyes to the pretectum. Its neurons eventually trigger, through the Edinger-Westphal nucleus and the ciliary ganglion, the muscle in the iris to ultimately constrict the pupil and govern light intensity on the retina [22].

2.1.3. THE CORTICAL AREAS

Figure 2.5 shows the transfer of information (optic radiation) from the lateral geniculate nucleus onward to the cerebral cortex [25]. The axons of the optic radiations terminate in the striate cortex, also called primary visual cortex or V1 [22]. The path from the retina to the primary visual cortex is called the primary visual pathway of which the biggest portion was already revealed in the section before. However, it is only at the striate cortex where the information of the eyes starts to lead to actual visual perception of the environment. The striate cortex namely integrates the visual information conveyed by the lateral geniculate nucleus and passes its work to other visual areas in the extrastriate cortex. The integration moreover covers summation of information from both eyes, for which a more elaborate description will be given in Section 2.2. Essential visual processing takes place in those areas, which encompasses the cortical areas V2, V3, V4, and V5 [26].

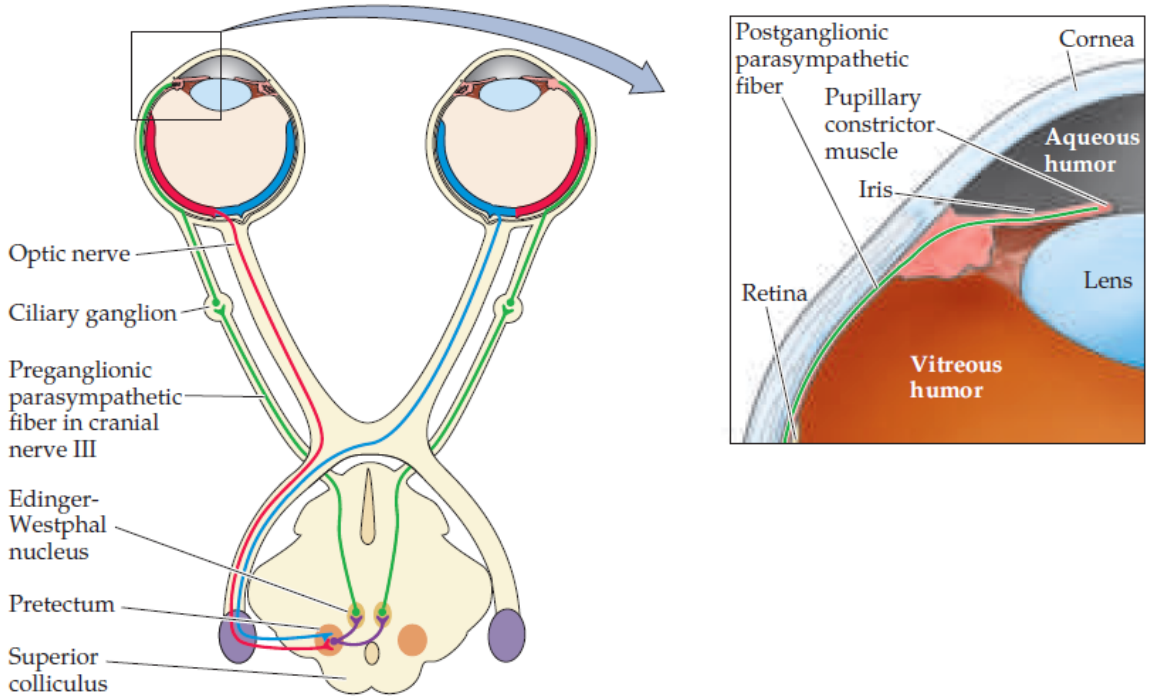


Figure 2.6: A sketch revealing the loop from retinal information to pupillary light reflex to regulate light intensity of the retinas of the eyes [22]

To understand what makes a human being perceive its surrounding environment visually requires analysis of very complex systems and structures in the central nervous system. The very processes in the various cortical areas are often not fully (if at all) understood and remain key topics for research in neurology [22, 24–26]. However, what is essential to adopt from these pieces of literature is that what is ultimately seen, i.e., visually perceived is in fact a creation of the cortical areas of the brain, not so much the eyes.

2.2. BINOCULAR VISION

The previous section listed various stages in the visual system from light in one's surroundings to visual perception in the cortices. However, nearly all flows of visual information are addressed irrespective of the number of eyes. In reality though, many living creatures visually perceive their surroundings through two eyes [25]. Neuroscientist Fahle outlined six advantageous reasons for having more than one eye [28]. Particularly interesting and within the scope of this research project is the third mentioned motivation: " *Stereoscopic depth perception is possible on the basis of binocular disparities*". Fahle mentions a specific phenomenon of visual perception called depth perception. The importance of depth perception to the project described in this document cannot be overstated. Its concept is clarified in the section hereafter (Section 2.3), but an understanding of having binocular vision first is therefore essential. This section will thus bridge the gap between (general) visual perception explained in previous Section 2.1 and visual *depth* perception in the following Section 2.3.

Binocular vision by definition implies being able to perceive the visual world through two eyes. However, its functionality demands more than solely having two eyes [25]. In fact, the difference between having two eyes and having binocular vision may be found in two groups of animals. Prey animals require by nature a different situation awareness through visual perception than predator animals. Their eyes are often placed at the side of their heads to extend their visual field as mentioned by Fahle. Predator animals, on the other hand, usually have a more focused visual field. The extent of their visual field is somewhat sacrificed for increased visual acuity and (significantly more) binocular vision. This phenomenon is given by Fahle as the third point in the aforementioned list. A schematic representation of the different visual fields of prey and predator is given in Figure 2.7. Note the smaller total visual field for predator animals as well as their increased span for the binocular visual field. The binocular visual field is the part in the full visual field that is overlapped by two eyes. Having binocular vision therefore implies a (significant) overlap of the fields of view from both eyes. Not only must one *have* two eyes, one must be able to *use* two eyes *together* [25].

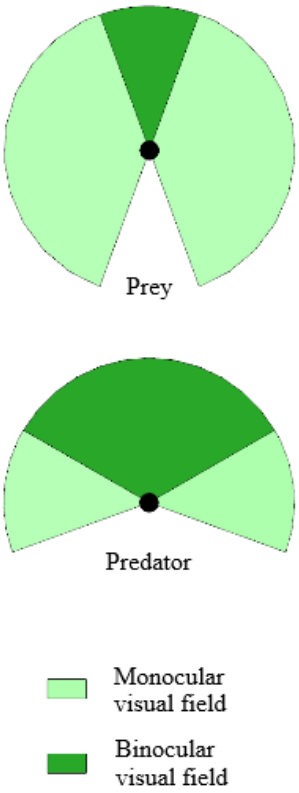


Figure 2.7: A sketch of monocular and binocular visual fields for two distinct groups of animals

As stated before, this section serves to bridge the gap between general visual perception of one's surroundings and eventual visual depth perception. An understanding of the latter is essential to the concept of the envisioned project stated in this document. With regard to binocular vision, this section will have its emphasis on the perception of three-dimensional or visual depth (Fahle's thirdly mentioned reason in the above list) and comprises two parts. Firstly, Section 2.2.1 picks up on the formation of retinal images in the eye and describes the necessary eye movements for proper retinal images within binocular vision. Subsequently, Section 2.2.2 addresses the main implication of binocular vision (in scope of this document as mentioned before). Stereopsis is one of the most dominant depth cues to human beings where this section will elaborate on. It serves as a basis for Section 2.3 where more depth cues will be analysed and the formation of eventual visual depth perception as a whole.

2.2.1. REQUISITE EYE MOVEMENT

In Section 2.1.1 the anatomy of the eye was exposed and the functionality of the retina explained. In order for a clear and focused image to be perceived, the specific point of interest in the surrounding environment should be refracted onto the fovea. This particular section of the retina is capable of producing an image with the highest visual acuity possible. In theory, one would be able to 'steer' the eye towards the required target by pure head movements if only one eye had to focus on that particular target. However, should both eyes focus on an object that is not infinitely far away (hence parallel viewing angles of the two eyes), the angle between the eyes must be altered in order for both eyes to have the object focused on the foveae. The movement of eyes in opposite directions is called vergence [22, 24–27].

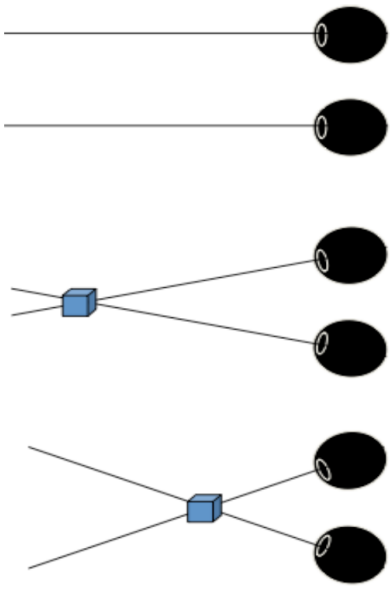


Figure 2.8: A depiction of converging eye movement with varying object distance [29]

The formation of an image on the eye's retina is schematically represented in Figure 2.9. An object is placed at a certain distance from the eye. With proper accommodation of the lens, the object forms a clear visual image on the retina. The center of the object is focused on the fovea of both eyes and a large portion of the visual fields of the left and right eye overlap. This is considered the binocular visual field [22, 25]. However, if the lines of sight of both eyes are considered parallel, the given images in Figure 2.9 could only be true if the eyes are at exactly the same position. Needless the say, they are not. Human beings have some interpupillary distance of 6.4 cm on average separating the two eyes [30]. This is fundamental for binocular disparity (outlined in the section hereafter) [25]. However, this also implies that the orientation of the eyes must be altered. Eventual fusion of the input from both eyes requires images of a particular focused object on corresponding points of the retina.

Bringing an object into focus close to the observer requires a convergent eye movement [27]. A schematic of such eye movement can be found in Figure 2.8. The top situation shows parallel sight lines for an object at infinite distance. As the object approaches the observer, the orientations of the eyes must be controlled towards the new object location. This is achieved by extraocular muscles. One group has already been covered in Section 2.1.1 and is called the intrinsic muscles located within the eye. Convergence is placed under a second group of muscles called the extrinsic muscles. These connect to the outer surfaces of the eye and control its movement as a whole [26].

In order for both foveae to be aligned with the fixation point in Figure 2.9, the eyes must thus converge. The convergence movement is shown in Figure 2.10. In accordance with the binocular visual field shown in Figure 2.9, the figure displays two eyes now directed towards a particular fixation point at some distance from the observer. The images of the fixation point are formed at the corresponding relative locations on both retinas. The ability to focus both eyes on the fixation point is essential for binocular disparity to be eventually a valid depth cue.

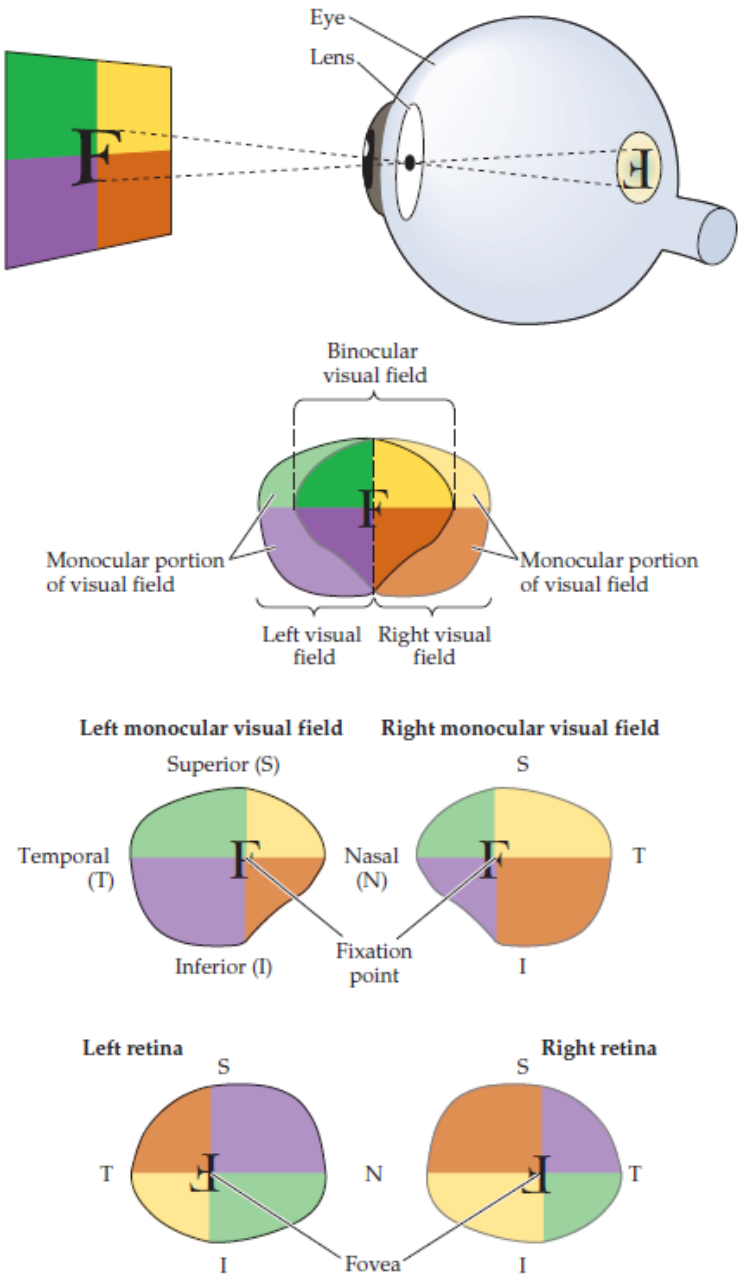


Figure 2.9: A sketch showing the formation of monocular retinal images for the left and right eye as well as a fused binocular visual image [22]

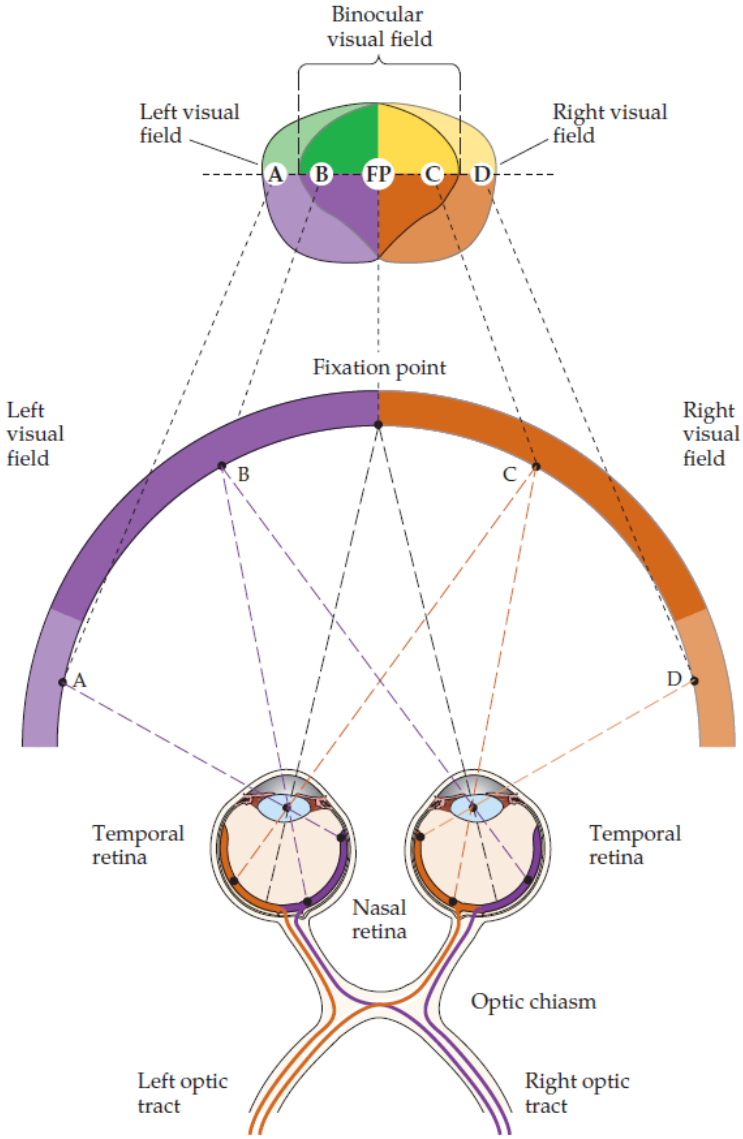


Figure 2.10: A sketch showing the monocular and binocular visual fields in binocular vision. The eyes convergence to a fixation point some finite distance from the observer and form an image on corresponding retinal points [22]

2.2.2. BINOCULAR DISPARITIES

In Section 2.3 following this current section an elaboration of visual signals that enhance three-dimensional vision, i.e., depth perception is given. However, a cue to depth results from binocular vision which has been studied more than any other depth cue [30]. This section will therefore focus specifically on perception of depth through binocular disparities preparatory to discussion of depth cues and depth perception in Section 2.3.

A shortest and simplest description of binocular disparity, also stereopsis, is the difference between what is seen with the left and right eye. The two eyes namely see the surrounding environment from slightly different perspective [22]. Given the converging orientation of the eyes to some fixation point at a finite distance from the observer, objects appearing in front or behind the fixation point form an image at non-corresponding points of the retinas [25].

The concept can be grasped from the sketch in Figure 2.11. Both eyes in the figure converge at some fixation point *F*. With the information from previous section on converging eye movements and corresponding retinal points, the image of *F* must therefore be at the same corresponding retinal points of the left and right eye. This can be seen in Figure 2.9 and Figure 2.10. In other words, should the two half images from both eyes be fused into a single frame, the location of *F* is identical for both eyes.

If point *F* is subsequently held as the fixation point and object *A* is considered, the fundamental concept of binocular disparity is noticeable. Whilst looking at *F*, point *A* appears at different locations relative to point *F* for the left and right eye. Taking the left eye into consideration, point *A* appears to the left of fixation point *F*. Similarly, point *A* is perceived to be right of point *F* for the right eye. Consider now the two half-images obtained in both eyes. Starting with the left eye, the fixation point *F* is projected onto the fovea and therefore placed in the center of the image. As stated before, point *A* will appear to the left hand side of *F*. This point is labelled a_1 in the figure. Should one do a similar analysis for the right eye, point *F* will again be found in the center of the image with *A* off to the right. For the right eye, this point is labelled a_r. Should both obtained half images from the eyes be overlaid into a *single* but *unfused* image, one will find the image as indicated by the box in the figure. Point *F* has identical locations for the left and right eye, but the projections of point *A* on the left and right eye appear offset to the left and right side of *F*, respectively. Roughly speaking, this offset between projections in the half images is known as disparity and eventually computed into depth by the brain [30].

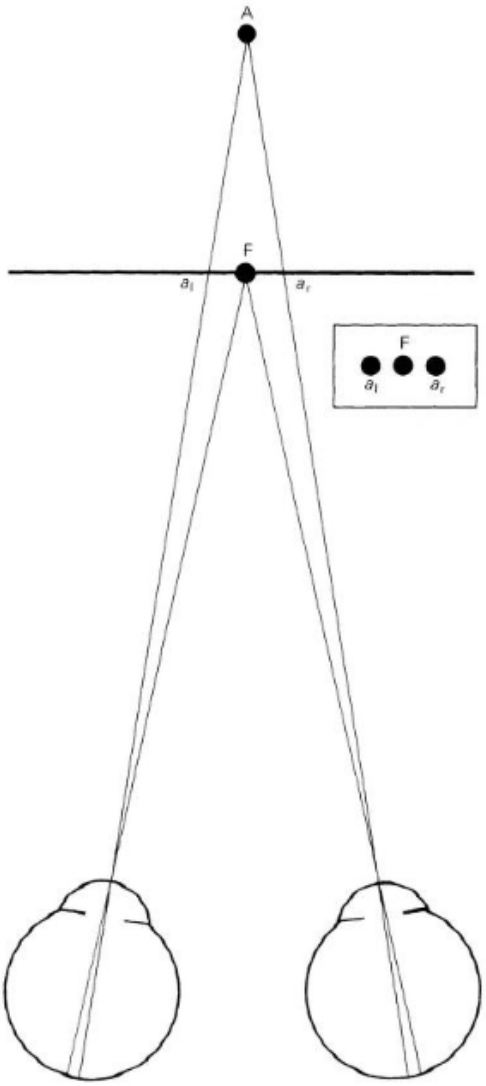
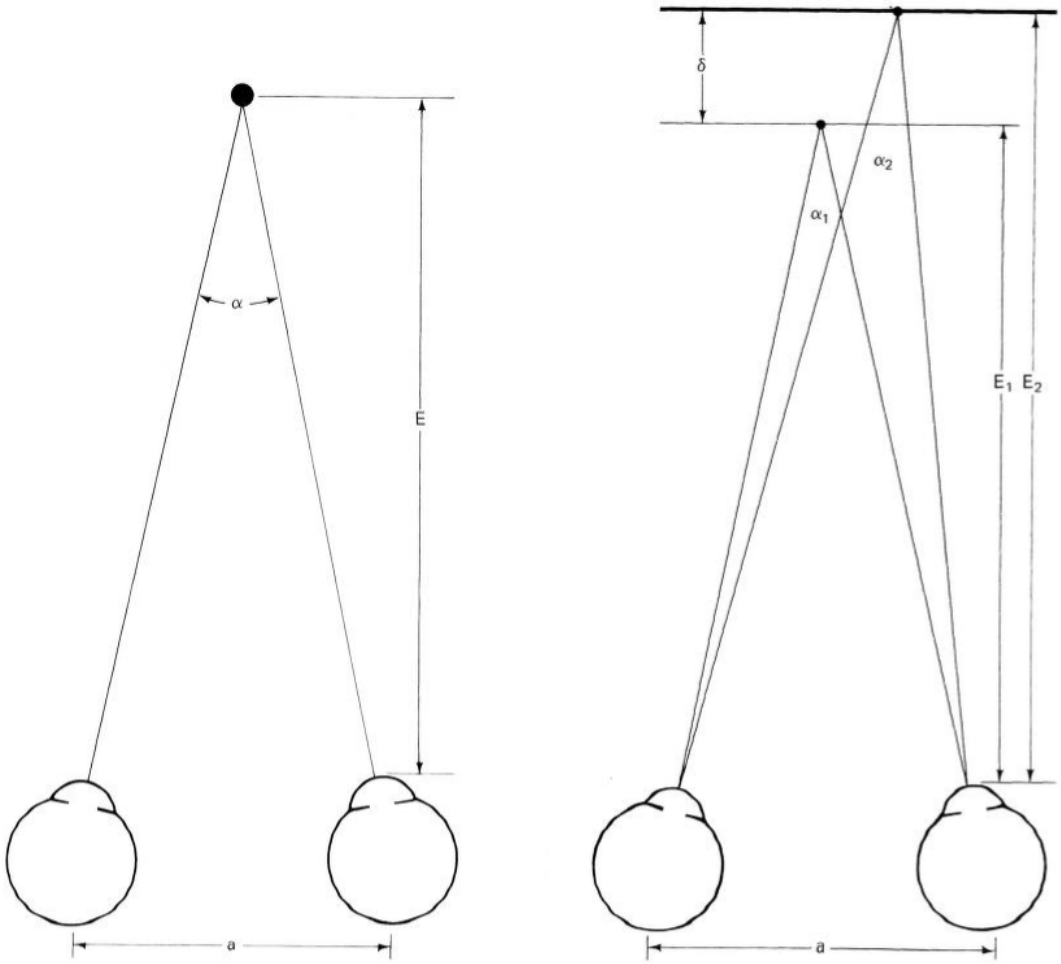


Figure 2.11: Fixation Point and Disparities [30]

Several types of disparities can be identified when considering optics in human vision. The aforementioned disparity is called horizontal disparity [27, 28]. Types like vertical disparity arise from e.g. non-centered frontal vision where an object closer to one of both eyes appears slightly larger to that eye. This may serve as a cue to how the observer is oriented to the object [28]. However, binocular disparity types other than horizontal disparity are out of the scope of the project described in this document and therefore omitted from this section.

The horizontal disparity is expressed in terms of an angle rather than a distance. Earlier the briefly described offset between two points on the retinas of both eyes was mentioned. The proper realization of binocular disparity and its computation may be deduced from Figure 2.12.



(a) Absolute parallax **(b) Relative parallax**

Figure 2.12: The illustrations show two different forms of parallax. (a) The absolute parallax is essentially identical to required eye convergence to focus on an object at distance E given interocular distance a. (b) The relative parallax is the binocular disparity to distinguish between two objects in depth [30]

Firstly the required inward angle of the eyes can be computed as shown in Figure 2.12a. Elementary geometry dictates that the angle between the two lines of sight known as absolute parallax angle α can be found by [30]:

$$\tan \alpha/2 = \frac{a}{2\mathrm{E}} \tag{2.1}$$

where a is the interpupillary distance or interocular separation and E the distance between the observer and the target. If the value for angle α is small, then by small angle approximation the relation may be reduced to $\alpha = a/\mathrm{E}$ [30]. In fact, this relation describes an aspect which has already been covered in Section 2.2.1 on convergent eye movements. The angle α found with the aforementioned equation really only describes the required angle to focus the target on both foveae at a distance E from the observer with a given interocular separation a. It is actually the difference between two convergence angles required for different targets that imposes the binocular disparity. Taking into consideration Figure 2.12b, the relative disparity ρ may be found by [30]:

$$\rho = \frac{a(\mathrm{E}_1 - \mathrm{E}_2)}{\mathrm{E}_1 \cdot \mathrm{E}_2} \tag{2.2}$$

where again a is the interocular separation and E_1 and E_2 are the distances from the observer to targets 1 and 2, respectively [30]. A simplification may be proposed if the relative distance δ between targets 1 and 2 is relatively small compared to the distances E_1 and E_2 between the observer and the targets [30]:

$$\rho = a\delta/E^2 \tag{2.3}$$

where E is the average of distances E_1 and E_2 [30]. A relation has thus been found that determines the quantitative binocular disparity in a situation where two targets are placed a distance δ relative to one another at an average distance E away from the observer with an interocular separation of a. One should take note that the given equations hold for *forward* looking situations. More complex geometry situations require adaption of the given equations that is not within the scope of this research.

For a given situation, one is now capable to compute the available binocular disparity. Take for instance a situation where two objects are placed 10 cm apart at an average distance of 5 m to the observer. The observer has average interocular seperation of 6.4 cm. Computation of binocular disparity ρ yields an angle of $2.56e^{-4}$ rad. For convenience, the value is often expressed in seconds of arc [25]. For the situation above, that is a binocular disparity level of about 53 seconds of arc.

Although the presence of binocular disparity can usually simply be approximated, the observer must also be able to sense the disparity to eventually perceive a sensation of depth through binocular disparity. The degree of sensitiveness to binocular disparity is known as stereoacuity [25]. Whereas regular visual acuity determines how well an image can be perceived with an eye, stereoacuity determines the minimum separation in depth that is distinguishable between two objects at a certain distance. More specifically, it is a measure of the smallest horizontal disparity that can be perceived by an individual in seconds of arc [31].

Researchers and medical personnel have various clinical examination possibilities at their disposal to establish one's stereoacuity. Some well known examples are the *TNO Stereo Test* and *Titmus Stereo Test*. The former comprises a red and green random dots stereogram capable of testing stereoacuity from 480 to 15 seconds of arc. The *Titmus Stereo Test* combines vectograph tests and passive polarized glasses to test stereoacuity all the way from 3000 seconds of arc to 40. All in all, the stereoacuity tests are great tools to qualitatively and quantitatively assess binocularity [31]. A selection of stereoacuity testing possibilities can be made based on aforementioned literature as a method of verification of adequate depth perception through stereopsis by experiment participants.

2.3. DEPTH CUES AND PERCEPTION

The visual perception of our surroundings allow for instance navigation through and interaction with the environment. The fundamentals of vision in general have been covered in Section 2.1. Section 2.2 built upon that basis by including the aspect of binocular vision. These steps have been taken to eventually move towards the concept of depth perception. This section addresses the construction of a three-dimensional visual perception or visual depth perception. Firstly, in Section 2.3.1 the physiological depth cues of binocular vision, accommodation and convergence will be stated. These have already been introduced in earlier sections. Section 2.3.2 and Section 2.3.3 will introduce several new depth cues that are categorized as pictorial depth cues and motion depth cues, respectively. Lastly, the summing and inclusion of depth in visual perception will be explained in Section 2.3.4

2.3.1. PHYSIOLOGICAL DEPTH CUES

The first category of depth cues are known as physiological depth cues or primary depth cues [29]. Different from the category pictorial depth cues discussed in the section hereafter, these depth cues are (not only) sensed by the retina and interpreted in the visual cortices. Accommodation, convergence and binocular disparity arise from physiological signals within the visual system. All concepts have already been introduced before. A concise description is given below.

Accommodation The elaboration on the visual system discussed firstly the eye itself. It senses visible light from the environment and refracts rays of light onto the retina to form a clear image of the surrounding. Light will firstly pass through the cornea [26]. It is the eye's major refracting element, but cannot be adjusted in refractive power. The lens within the eye can be changed in curvature to alter its refractional characteristics [22]. The ciliary muscles connected to the lens must therefore either contract or relax. The accommodation that is necessary to clearly define the image of an object at some distance to the observer serves the visual cortices as information about the required state of the lens and by optic principles thus the distance to that object [24, 25].

Convergence The second source of information was introduced with binocular vision in Section 2.2.1. Anatomy of binocular vision implies some eye movement is required to form an image of the object at some finite distance at both foveae [25, 27]. Elementary geometry dictates the required converging angles of the eyes to be inversely related to the distance between a focused object and the observer [24]. The converging eye movement inward to some object is produced by extrinsic muscles connected to the sclera at the outer surfaces of the eyes [25]. Like accommodation, convergence serves as a physiological depth cue through the state of muscles.

Binocular disparity The introduction of binocular disparity also followed the inclusion of binocular vision. Contrary to accommodation and convergence, binocular disparity follows not from some muscles' internal state. However, it cannot be categorised as a pictorial depth cue either. It is very much a binocular depth cue that requires analysis of differences between the two half-images of both eyes [24, 30]. Binocular disparity namely, like convergence, finds its basis in (human) anatomy. The eyes are separated on average by 6.4 cm and thus have a slightly different perspective on the same focused object [30]. The separation of the two images' retinal points of the same object point is referred to as disparity. The disparity is interpreted in the visual cortices and eventually yields a single, three-dimensional vision from two flat half-images [25].

Over recent years, technology has allowed implementation of these depth cues in artificial environments [7]. The functioning and details of these possibilities will be discussed at a later stage in this document, but it is essential to understand the coupling of accommodation and convergence in a natural environment. As it turned out, the eyes turn inward to align their foveae with the desired object. Simultaneously, the eyes individually adjust their lenses to provide the clearest image possible. Should the object move closer to or further away from the observer, both the accommodation as well as the converge will have to adapt to the new situation accordingly. At this point, a mismatch is introduced between the two cues when observing that same object from a stereoscopic 3D monitor for example (more information on display technology follows in Chapter 6). Although the converging movement of the eyes can be triggered in an artificial environment, the eyes will inevitably remain identical accommodation at the display screen. The direct movement of an object towards or away from the observer in an artificial environment causes a mismatch between two depth cues [29, 32, 33].

2.3.2. PICTORIAL DEPTH CUES

The art of expressing depth in paintings was already a highly sought-after skill many centuries ago. Artists began to develop an understanding how the illusion of a three-dimensional scene or objected could be put on a flat piece of material. [30]. Even though it may not be straightforward to isolate every single visual 'trick' to encourage such illusion, it needs no expertise to perceive some sensation of depth in a flat-surfaced painting. Our brains are very capable of registering various sources of information that assist in building up a three-dimensional visual perception. These sources are known as pictorial, secondary or monocular depth cues [24, 25, 29, 30, 34]. Find a list of pictoral cues below.

Perspective The visual effect of perspective demands explanation of the formation of images of an object on the retina as a result of the distance to that object [30]. This is illustrated in Figure 2.13. If firstly an object with length h is placed close to the observer, the object is perceived as visual angle α_1. The object thus covers a portion α_1 of the total visual field (for instance denoted as α_{total}). Should the object now be moved further away from the eye, its length would obviously remain the same. However, the observer will have a different perception of the very same object based on its new distance. As the object moved further back, the visual angle reduces to e.g. α_2 and thus covers less of the total visual field compared to the object closer to the observer. The object at distance is said to 'fill' a lesser portion of the total visual field and is therefore perceived differently.

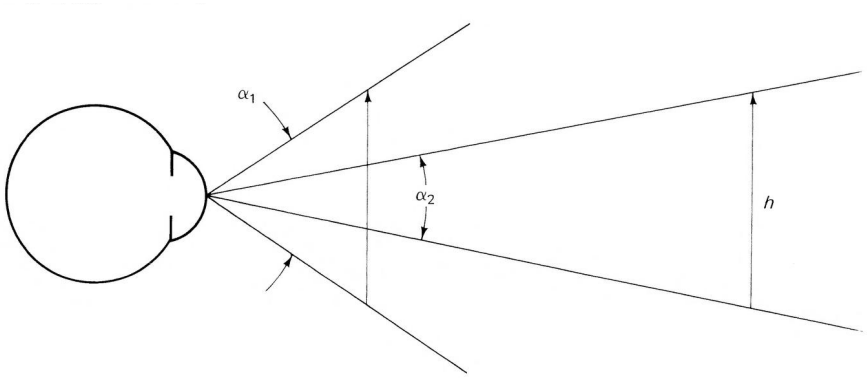


Figure 2.13: A sketch indicating a change of visual angle α by increasing the distance to some object of length h [30]

The formation of a visual image of the object may end up at two extremes. Firstly, if length h at some close distance results in a certain α larger than the total visual field, the object is larger than what may be seen and exceeds the visual field limits. On the other hand, for any object with finite length h, the visual angle perceived approaches zero if the distance between object and observer reaches infinity.

The effect of perspective can subsequently best be shown in some scene with identical objects and large viewing distances. The perceived sizes of all objects in that scene vary linearly with their distance to the observer. This perspective is therefore often referred to as linear perspective [24, 25]. Eventually, the objects at great distances can hardly be seen as their perceived size approaches zero visual angle. Perspective can thus be defined as the ensemble of perception of sizes of objects at various distances to the observer.

Texture gradient The principle of perceived sizes as a result of perspective implies another cue that is said to serve as information about distance [24]. Not only is the perception of size proportional to visual depth, also the detail of the object that is perceived [30]. This is a monocular depth cue known as texture gradient [24, 25] or detail perspective [30]. The way surface details such as grass, sand or water are perceived vary as their textural density increases with distance [24, 30]. It is believed that a change in detail perception with distance may tell an observer much about distances in the environment [24].

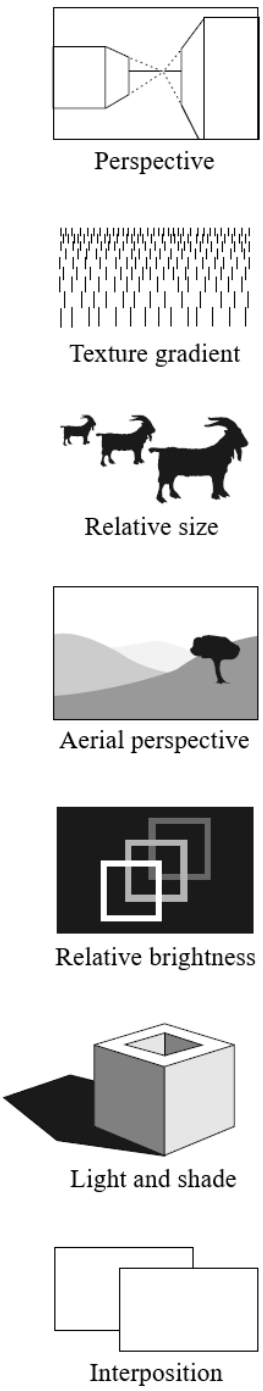


Figure 2.14: A list of illustrations of pictorial depth cues [34]

Relative size The previously described Figure 2.13 reveals a relation between object size and object distance. Given a specific value for the object's size h, one may argue the distance can be deduced of a certain visually perceived scene. However, memory of a familiar object's size or knowledge of its size in general help interpret size and distance more progressively than illustrated in Figure 2.13 [24, 25, 30]. If a particular object is known to be large yet is visually perceived small, the observer may judge that the object is placed far away from the observer. Moreover, if more than one similar object is found in a specific scene with varying image sizes of those objects, the impression of different distances towards the objects arises [30]. The knowledge and/or memory of sizes of objects thus contributes to the judgement of distances and is called visual size or relative size [24, 25, 30].

Aerial perspective The perception of distance in a far-stretched scene is encouraged by a visual effect referred to as aerial perspective [24]. Objects and landscapes at great distances appear different in color and clearity [24]. The explanation lies in the fairly large distances that must be travelled by light through the atmosphere. Its effect is therefore highly dependent on atmospheric conditions, although the visual system does expect aerial perspective to affect the visual representation of large distant scenes [29].

Relative brightness In certain scenes the luminance of objects may be informative of the object's distance to the observer [30]. The depth cue called relative brightness was often assumed to illuminate objects closer to the observer more than distant objects. However, this conception is only valid if point sources are considered, not surfaces reflecting light [24]. The amount of luminance has no relation to the distance between the object and the observer [30]. Relative brightness may nonetheless inform the observer about distances in certain scenarios. Objects at various distances to a point source of light e.g. a light bulb or fire will appear differently illuminated and can therefore be informative about relative distances between objects as a result of relative brightness [30].

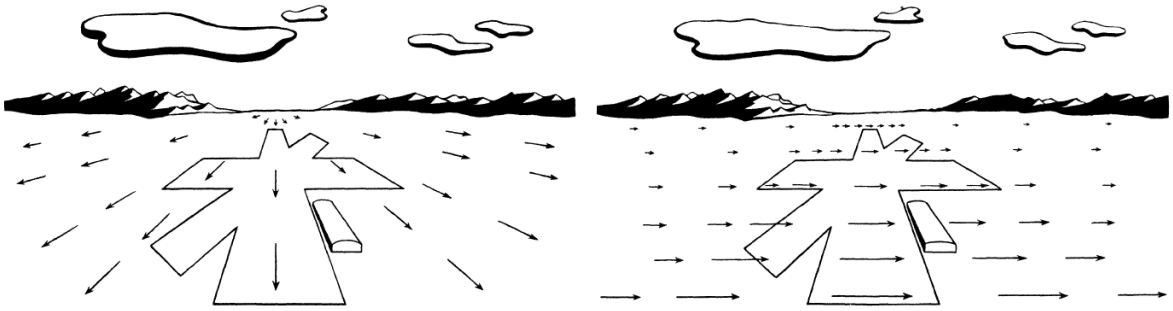
Light and shade The effect of light and shade may greatly enhance the depth of any flat-surfaced scene [24, 30]. For a given object, the difference between lighted and shaded parts of an object reveal the depth of an object [24]. The casting of shadow of a particular object may moreover reveal some relation of its distance to the surface on which is shadow is projected [30]. Light and shade thus enhance not only the distance of an object, but also the sensation of depth of a single object.

Interposition Images of objects are only formed if that object can be placed on the retina. If for some reason the object is (partially) concealed from the eye, it is occluded from view. Occlusion, or interposition, is a depth cue that reveals some relative distance of one object visually blocked by another object or surface [24, 25, 30]. The knowledge of relative position of an occluded object is said to be a rather powerful depth cue [30].

2.3.3. MOTION DEPTH CUE

The depth cues stated up until this point, whether physiological or pictorial, all considered visual observations possible from stationary position. A motion depth cue is, as Gibson states, only perceivable by "the active observer" [24]. The motion depth cue, or motion parallax, requires a relative apparent motion between the observed and the observer [24, 25, 30]. In essence, the functioning of motion parallax is somewhat like binocular parallax (also binocular disparity). Whereas binocular disparity is based on a difference between images from two different perspectives, motion parallax is based on a difference between two different images from slightly different points in time [29].

Two different sorts of motion are said to trigger motion parallax [7]. Firstly, objects at varying distances viewed from a motion relative to those object by the observer appear to move or 'flow' at different rates proportional to the distance between those objects [25]. The motion relative to the object may be along the line of sight (shown in Figure 2.15a) or perpendicular to the line of sight (Figure 2.15b) [29]. The effect may for instance be noticed looking out the side window of a moving vehicle; light posts and trees closer to the road appear to pass by much quicker than distant buildings or mountains. It may moreover be demonstrated by lateral head movements. Objects closer to the observer tend to shift more in the visual field than objects at larger distances [24].



(a) Flow of motion parallax with the line of sight parallel to the line of movement [24]

(b) Flow of motion parallax with the line of sight perpendicular to the line of movement [24]

Figure 2.15: The illustrations show two forms of motion parallax. (a) The observer's movement is towards the horizon, as is the line of sight. With subsequent observations, the environment 'flows' under the observer at rates dependent on distance (b) The observer's movement is now towards the left and the line of sight towards the horizon. Observations are thus perpendicular to movement. With subsequent observations, the 'flow' is in a uniform direction, but the magnitude once again depends on distance

Additionally, the second form is the rotation of an object itself which may give an impression of its three-dimensional shape [7]. This form does not necessarily require a movement by the observer as Gibson stated. Imagine several two-dimensional static images of some three-dimensional object. Even if the object is completely unlit and merely its silhouette can be grasped from a single image, its three-dimensional shape may be understood if a sequence of several images is shown. Note, however, that this form of perception considers the depth of an object as an aspect of its three-dimensional shape rather than a distance in an environment along the third dimension.

Motion parallax may thus advantageously contribute to visual depth perception by indication of distances between objects through their flow in the visual field, or judgement of depth of the object itself. Contrary to the earlier discussed depth cues by physiological or pictoral source, however, motion depth cues require a dynamic environment that cannot be displayed on a painting or stereogram. On the other hand,

2.3.4. DEPTH PERCEPTION

The environment surrounding the observer is found to be filled with sources of information that somehow assist with judgements on distances. It may result from what is seen or how it is seen depending on the type of depth cue. In Section 2.1 the flow of information was already described for any general visual image. To attain depth perception rather than visual perception, the depth cues obtained from either pictorial or physiological sources must likewise be integrated. The information passes the neural circuits to be brought together in the striate cortex [22]. This section will however focus solely on the construction of visual depth perception with depth cues as building blocks and their roles in the three-dimensional world.

For numerous reasons, researchers have been interested in the role and significance of various depth cues. A noteworthy mention is the testing of depth perception precision in displays. Depth perception from display is discussed later in this document, but the significance of studies on this matter greatly contributes to the understanding of depth cue usage and dominance [29]. The concept of depth cue usage and dominance is included in the weighted additive model of depth perception. It states that the amount of depth sensation linearly depends on which cues are available in the scene and their strength or effectiveness [35]. The relation between two or more depth cues was analysed by two different sorts of studies. Firstly, depth salience studies determined which depth cue or depth cues in a certain scene resulted in the highest perception of depth in that scene. Secondly, in depth cue conflict studies, scenes were generated with unique opposing depth cues to analyse which depth cue proved a stronger source of information [35]. The survey firstly concludes that the weighted additive model is generally accurate, but does state some exceptional situations. Moreover, the conclusion is drawn that binocular disparity, motion parallax and interposition are depth cues with relative high weights, i.e., dominant depth cues [35].

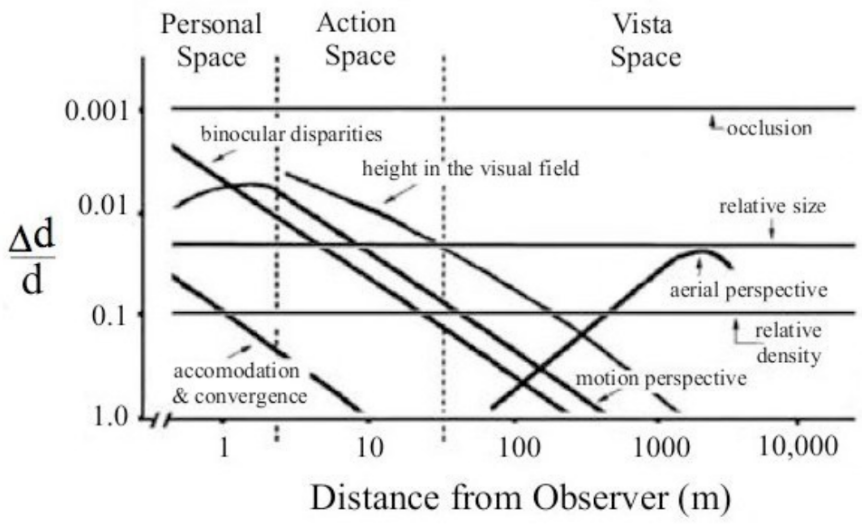


Figure 2.16: Build up of stereoscopic vision [29]

The conclusions on the weighted additive model can be supported with somewhat similar results found in a series of other experiments. These have been carried out to analyse the accuracy of judgements on distance for various depth cues and the influence of distance between object and observer [36]. The results are illustrated in Figure 2.16. The graph firstly categorises three distinct observation ranges [37]. Personal space is within arm's reach wherein manipulation of objects takes place. Action space is within close distance wherein self-movement takes place. Vista space is within a larger environment. These spaces determine the distance from the observer to the object(s). The graph subsequently indicates the depth sensitivity for numerous depth cues across the three distinct spaces [36]. An indication is thus given for each depth cue how small a distance Δd can still be distinguished between two objects at a distance d from the observer. One could argue the graph presents the 'accuracy' of various depth cues at a given distance.

Results such as the illustration in Figure 2.16 are of great importance to our understanding of depth perception. Knowledge not only on the existance of depth cues, but also their coherence across the observation range yields a fundemental understanding in the adaptive and dynamic nature of the visual system [29].

3

INTRODUCTION TO MANUAL CONTROL BEHAVIOR

The human pilot is "*a multimode, adaptive, learning controller capable of exhibiting an enormous variety of behavior*" [2]. Nothing is better indicative of the complexity of human control behavior than this statement by McRuer and Jex in 1967. However, as a result of brilliant researchers and their great dedication, numerous novel research studies helped to discover that human controllers do respond quite identical in specific situations. Of course, these specific situations cover merely a fraction of the full range of human control capabilities, but with ever ongoing research into new aspects of manual control a better understanding is gained for more and more applications.

This chapter introduces manual control behavior in terms of requisite knowledge for the continuation of this document. Over the years, numerous great pieces of literature on manual control behavior are published focusing on many different aspects and new ways to understand and model human control behavior. This chapter will, however, cover only the fundamentals of manual control behavior that play a role further down this document for the introduction of depth control and initial attempt to model depth control behavior. Section 3.1 firstly puts the human controller in a control system to discover the fundamentals of manual control behavior. In Section 3.2 two distinct control behavior types are discussed that may take a part in this research.

3.1. HUMAN CONTROL IN CONTROL SYSTEMS

The start of research into understanding human behavior during manual control tasks is found in studies by McRuer [1], Tustin [38] and Elkind [39]. As World War II rapidly drove the need for more complex technology, the ideology of human-machine interactions changed [40]. Knowledge on the dynamic nature of a human pilot was deemed more essential to design a system that could be controlled with the highest performance and safety. McRuer and Jex have therefore identified several variables that affect the overall system [2]. These are schematically presented in Figure 3.1.

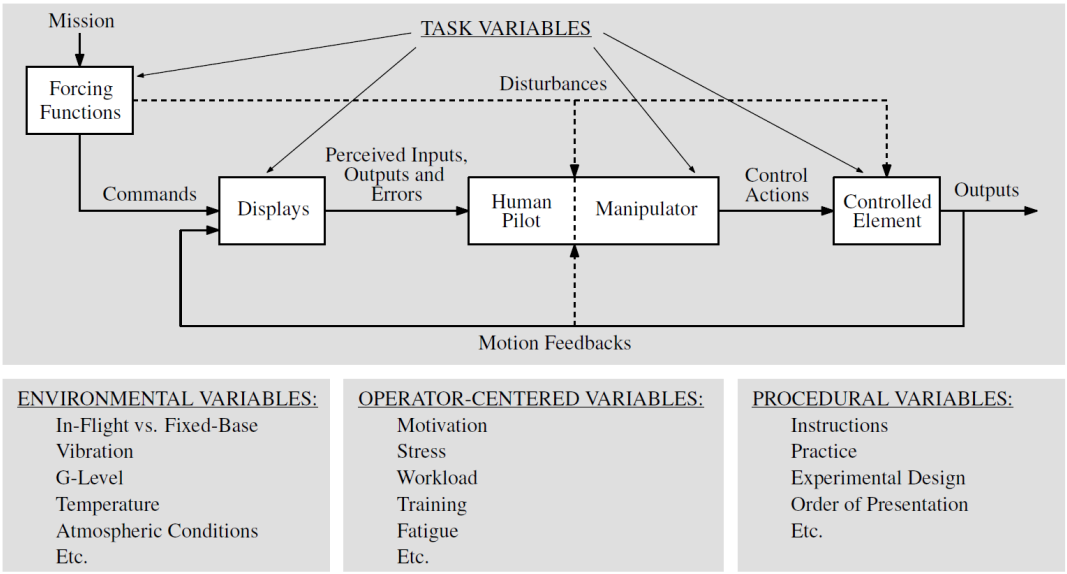
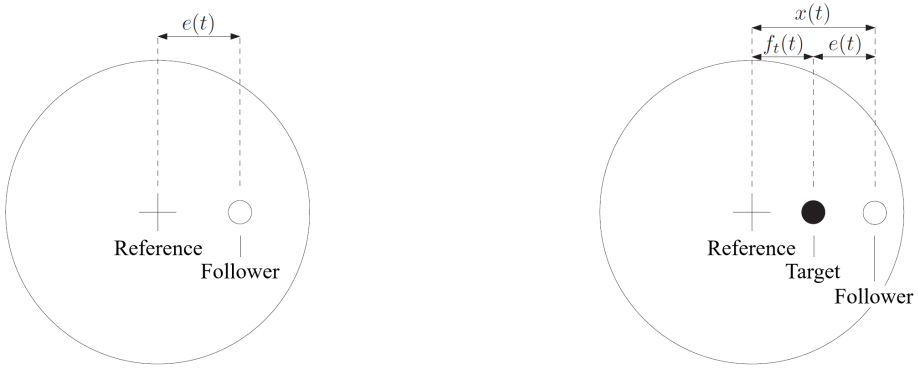


Figure 3.1: An illustration of the variables defined by McRuer and Jex [2] affecting the pilot-vehicle system [41]

The four categories of variables are task variables, environmental variables, operator-centered variables and procedural variables. For the basis of research into human adaption to and behavior with depth control tasks and stereoscopic displays the emphasis of this research is on the task variables. More specifically, the relation of *Displays* to *Human Pilot*. A full explanation of the changes required for depth control tasks and adoption of stereoscopic displays will be given in Chapter 4, but the classical displays found in many cybernetic research studies are briefly introduced in this section.



(a) Compensatory display **(b) Pursuit display**

Figure 3.2: Schematic representations of the different display types. (a) A compensatory display revealing solely system signal $e(t)$ between the reference and the follower. (b) Pursuit display revealing system states $f_t(t)$, $x(t)$ and $e(t)$ between the reference, and the target and follower [1]

Over the years two main display types have been the cornerstones for human control behavior in manual control tracking tasks. They are the compensatory display (Figure 3.2a) and the pursuit display (Figure 3.2b). At a first glance they appear quite similar. A major difference may however be found in the way system information is provided to the human operator. Note the flow of information in Figure 3.1 from *Displays* to *Human Pilot*: perceived inputs, outputs and errors. Whether all of these signals are available to the operator depends on the display provided. The compensatory display namely only reveals the error signal $e(t)$ as shown in Figure 3.2a. On the other hand, the pursuit display indicates all three system signals $x(t)$, $f_t(t)$ and $e(t)$. Not only do these

display types represent a distinct way of presenting system feedback to the operator, they also represent two unique control behavior types. These are described in Section 3.2.

The remaining task variables -*Forcing Functions*, *Manipulator* and *Controlled Element*- are aspects that will play a role in the eventual experiment design discussed in Chapter 5. They are shortly explained here to highlight their effect on the experiment. Firstly the forcing functions are the driving element behind the control tracking task. They are to induce a responsive action by the human operator and often constructed of quasi-random appearing multi-sine signals [1, 42]:

$$f_{d,t}(t) = \sum_{k=1}^{N_{d,t}} A_{d,t}(k)\sin(\omega_{d,t}(k)\,t + \phi_{d,t}(k)) \tag{3.1}$$

The manipulator is considered the means for a human operator to generate system input. As shown in Figure 3.1, the *Human Pilot* and *Manipulator* are mostly considered lumped as only the input to the human operator and his manipulator output can be objectively measured in a standard control tracking task experiment. The manipulator can be thought of as a control stick or steering wheel to control the system. The system's response is then defined by the last task variable *Controlled Element*. Its dynamics define how a system responds to the human output. As the *Manipulator* and *Controlled Element* have no explicit part to play in the experiment variables, these are not further considered in this document.

3.2. CYBERNETIC MODELS OF MANUAL CONTROL BEHAVIOR

In Chapter 2 several statements are already given on the difference between what the eyes see and how eventual visual perception is attained. A similar difference between displayed information and perceived information can be identified. Krendel and McRuer determined the Successive Organization of Perception (SOP) [43]. A human controller is said to perform a control task at one of three different levels: compensatory, pursuit and precognitive behavior. With increasing knowledge about the system, the operator starts to control with different use of available information. This section will elaborate only on compensatory (Section 3.2.1) and pursuit behavior (Section 3.2.2). Precognitive behavior namely considers a human controller fully aware of the system and its trajectory. For the envisioned experiment, random appearing forcing functions are being used which most likely omits the possibility of precognitive behavior.

3.2.1. COMPENSATORY BEHAVIOR

Compensatory behavior is the first level of SOP [43]. The human operator controls only the system error signal $e(t)$ between $f_t(t)$ and $x(t)$. This is schematically represented by Figure 3.3. For this SOP level the compensatory behavior can be modelled with the well known crossover model [1]. The model describes the operator with a linear response to the error signal e with additional remnant n accounting for non-linear control behavior. These signals can be found in the figure.

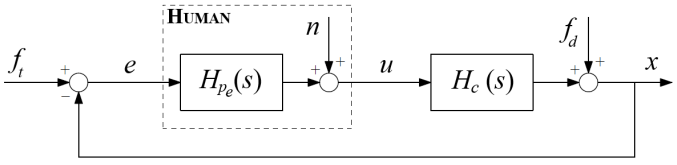


Figure 3.3: Compensatory behavior as a first level of skill-based manual control behavior in the Successive Organization of Perception defined by McRuer and Jex [2]

The adaptive nature of a human controller can be shown by the fact that the open-loop transfer function has the form of single integrator around crossover frequency ω_c. That implies that the dynamics of the human controller depends on the dynamics of the *Controlled Element*:

$$H_{OL}(j\omega) = H_p(j\omega)H_c(j\omega) = \frac{\omega_c}{j\omega}e^{-j\omega\tau_e} \tag{3.2}$$

In the crossover frequency region, the pilot describing function has the general form:

$$H_p(j\omega) = K_e \frac{(T_L j\omega + 1)}{(T_I j\omega + 1)} e^{-j\omega\tau_i} \tag{3.3}$$

where the fraction shows the equalization term that forces the open-loop transfer to have a single integrator form around the crossover frequency. The term can thus be neglected for a single integrator dynamics *Controlled Element*.

The applicability of the crossover model may be extended with the addition of a compensatory term for the neuromuscular system [44, 45]. This allows for better modelled high-frequency behavior that is largely due to the human neuromuscular dynamics. This term may be added to the pilot describing function:

$$H_{nm}(j\omega) = \frac{\omega_{nm}^2}{(j\omega)^2 + 2\zeta_{nm}\omega_{nm} j\omega + \omega_{nm}^2} \tag{3.4}$$

where ω_{nm} represents the neuromuscular frequency and ζ_{nm} the damping factor.

The methods for identification of manual control behavior, parameter estimation of the pilot describing functions as well as actual results of identification of a preliminary experiment are presented in Chapter 6.

3.2.2. PURSUIT BEHAVIOR

The second stage of SOP is pursuit behavior where possibly all three inputs are used by operator [43]. This is shown in Figure 3.4. However, since the signals $f_t(t)$, $x(t)$ and $e(t)$ are related by $e(t) = f_t(t) - x(t)$ the system is overdetermined [46, 47]. What is more, the pursuit behavior is less studied and due to this lack of research on this specific topic, relatively little is known about pursuit behavior [18].

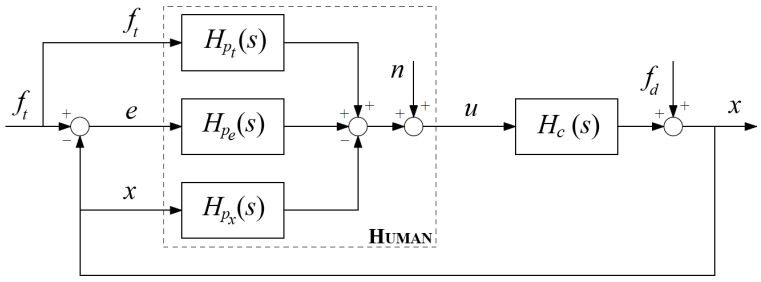


Figure 3.4: Pursuit behavior as a second level of skill-based manual control behavior in the Successive Organization of Perception defined by McRuer and Jex [2]

What must be taking into consideration is the fact that pursuit-displayed information to the human operator does not necessarily induce pursuit control behavior. This is quite essential to the research described in this document. In fact, a compensatory control strategy may be identified on a control task with pursuit displays. For rate-controlled dynamics for the *Controlled Element*, that is, $H_c(s) = K/s$ the human operator exerts a error-reducing control strategy based solely on H_{p_e}. This compensatory control behavior thus neglects additional responses of H_{p_x} and H_{p_t}. With an experiment settled on single integrator control dynamics for the *Controlled Element*, the actual identification and modelling of behavior may therefore be done with better studies compensatory models even though a pursuit display type is used.

4

INTRODUCTION TO MANUAL DEPTH CONTROL

The fundamentals of the visual perception of system feedback in flat-plan control tracking tasks have been given in Chapter 3. However, an alteration to current control tracking task schematics is necessary to define the concept of manual depth control. The basis of visual perception and particularly depth perception was therefore firstly discussed in Chapter 2. This chapter combines the information of Chapters 2 and 3 to define and elaborate on the concept of manual depth control tracking tasks.

The chapter firstly shows the differences with flat-plane control and the implication of addition of stereoscopic vision in Section 4.1. Various illustrations are provided to visually show the various steps from monoscopic flat-plane control tasks towards stereoscopic depth control tasks. Section 4.2 will thereafter show in more detail the already mentioned practical application of depth control in aerial refueling and discuss numerous studies on depth control tasks.

4.1. CONTROL SYSTEM ALTERATIONS

The steps to move from better studied flat-plane tracking tasks to lesser understood depth tracking tasks are outlined in this section. Several illustrations show the required adjustments for which the starting point is already established in the previous chapter with Figure 3.2a and Figure 3.2b. A distinction is made between the change of flat-plane towards depth control and monoscopic towards stereoscopic vision. These changes can be found in Sections 4.1.1 and 4.1.2, respectively.

4.1.1. MOVEMENT AXIS FROM FLAT-PLANE TO DEPTH

Lets start off by defining a flat-plane control task as found in many manual control tracking task experiments as these -in many variants of course- have been more or less the standard for cybernetics research. Figure 4.1a and Figure 4.1b illustrate the established definitions for flat-plane and depth control tasks, respectively.

Flat-plane control tasks (Figure 4.1a) are considered with tracking tasks where the follower (and possibly target for pursuit display tasks) moves in a plane that is *perpendicular* to the line of sight of the observer. It is generally said that objects then move within the viewing image plane. This is shown in Figure 4.2a and Figure 4.2b for a compensatory and pursuit display, respectively. The displays are identical to Figure 3.2a and Figure 3.2b from the previous chapter, but shown in addition with the perspective of the observer to the display to elucidate the orientation of relative follower/target movement. Note that the movement may be either horizontal, vertical or a

combination of both.

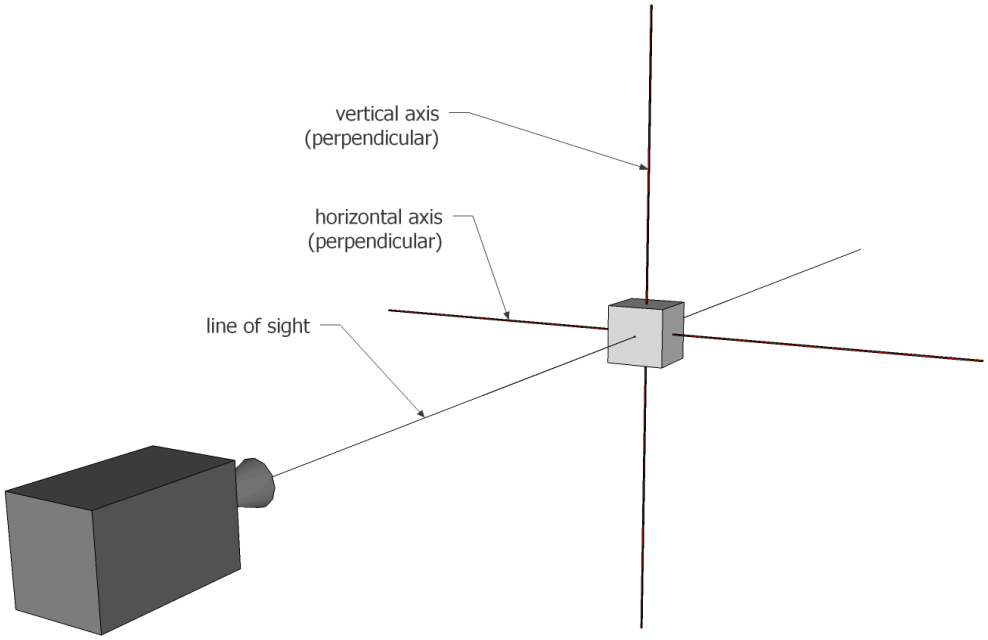
A control tracking task where the object is solely moving towards or away from the observer is defined as a depth control task. However, for a *pure* depth control task, the object may not translate along either the horizontal or vertical axes. The axis of movement for a depth control task must therefore not only be *parallel* to the line of sight, it must also *coincide with* the line of sight. Were the axis of movement parallel but offset from the line of sight, the tracking task would not be dealing with pure depth but rather a perspective view of the situation with different implications to pure depth control [48].

Upon changing the axis of movement from a perpendicular horizontal or vertical axis to a parallel coinciding depth axis, the perception of system signals found in e.g. a compensatory or pursuit display change accordingly. Although the graphical representation may feel quite obvious, various aspects of the perception of those signals alter thoroughly. The way a human controller adapts or behaves to these changes may significantly affect the applicability of current models to control behavior in these depth control tasks [18].

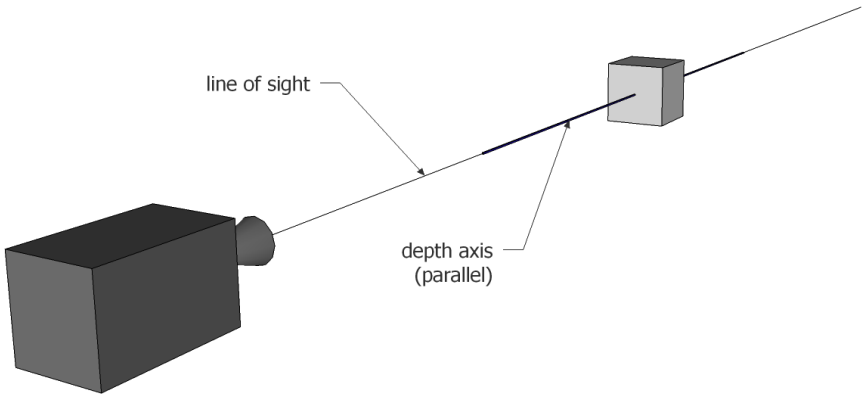
The difference in perception between flat-plane and depth control is shown in Figure 4.3 for a pursuit display. Here a pursuit display is chosen as an example, but a similar comparison can be made with e.g. a compensatory display. The pursuit display is once more shown with a perspective view on observer-display situation for a flat-plane control task in Figure 4.3a. One may note the target and follower move in the plane perpendicular to the observer's line of sight. As such, the perception of $f_t(t)$, $x(t)$ and $e(t)$ are perpendicular to the observer. This of course characterizes the flat-plane control task. Figure 4.3b subsequently reveals the new orientation of system signals $f_t(t)$, $x(t)$ and $e(t)$ upon change of the axis of movement.

The new display in Figure 4.3b has an impact on the perception of $f_t(t)$, $x(t)$ and $e(t)$ worth noting. The explanation for this has already been treated in Section 2.3. The pictorial depth cues perspective and relative size may together allow the observer of some arbitrary scene to predict the distance to some (known) object by its size rather than perceive merely its size. Essential to recall is that the perception of object size is by retinal size rather than true object size (graphically shown in Figure 2.13). The retinal size is an angular measure of an object with size h at some distance to the observer. The retinal size of that object follows from geometric relation between the true object size and its distance. Should one translate this characteristic to the pursuit display in Figure 4.3b, some judgement is possible for the distance from the observer to the follower and target.

A comparison between flat-plane and depth control tasks, and the judgement of distances ,i.e., system signals $f_t(t)$, $x(t)$ and $e(t)$) yields a serious consequence. Wickens produced a number of guidelines for implementation of sterescopic 3D (S3D) displays based on human visual capabilities [7]. For the perception of distances along the depth axis, Wickens says: "*Any representation of a 3D world on a 2D image surface produces an inherent ambiguity. The absolute distance represented by a point along the line of sight cannot be ascertained with high accuracy, compared to absolute distances parallel with the viewing image plane (the plane orthogonal to the line of sight). (This limitation is also characteristic of direct viewing.) Thus, 3D displays create the potential for perceptual ambiguity which is not present in a set of orthogonal 2D displays*". For Wickens' statement in terms of cybernetics, the perception of system signals $f_t(t)$, $x(t)$ and $e(t)$ along the depth axis is not constant for all distances between the follower/target and the observer. The non-linear effect on depth perception may grow even stronger with the addition of stereopsis (discussed hereafter). An analysis of sensitivity to system signal $e(t)$ as a function of object distance is done in Section 6.2.

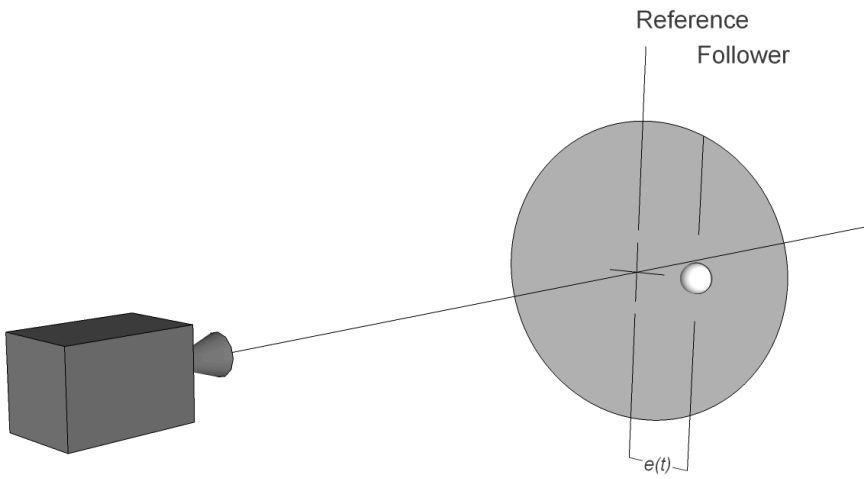


(a) Axes of movement defining flat-plane manual control tasks

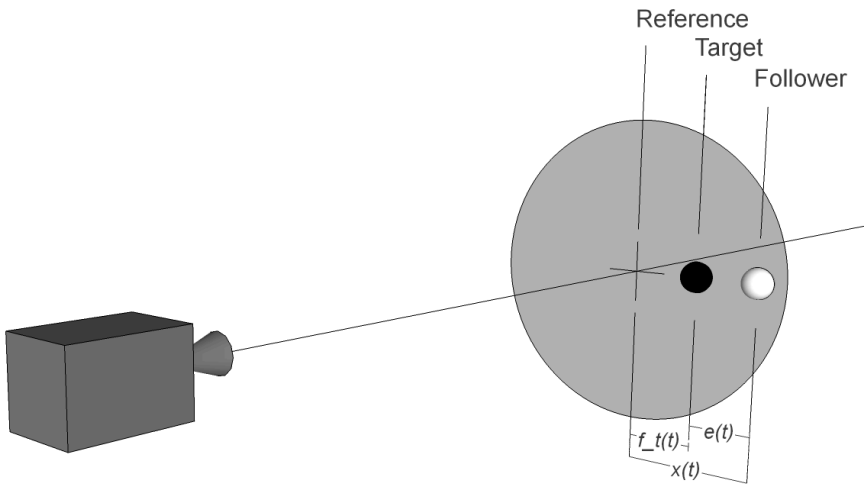


(b) Axis of movement defining depth manual control task

Figure 4.1: A schematic representation for the definition of flat-plane and depth control task planes of movement. (a) The object moves along the horizontal and/or vertical axis relative to the observer. Both axes are *perpendicular* to the line of sight. The axes typically define the classical viewing image plane (b) The object moves straight towards or away from the observer where the axis of movement is *parallel to and coinciding with* the line of sight.

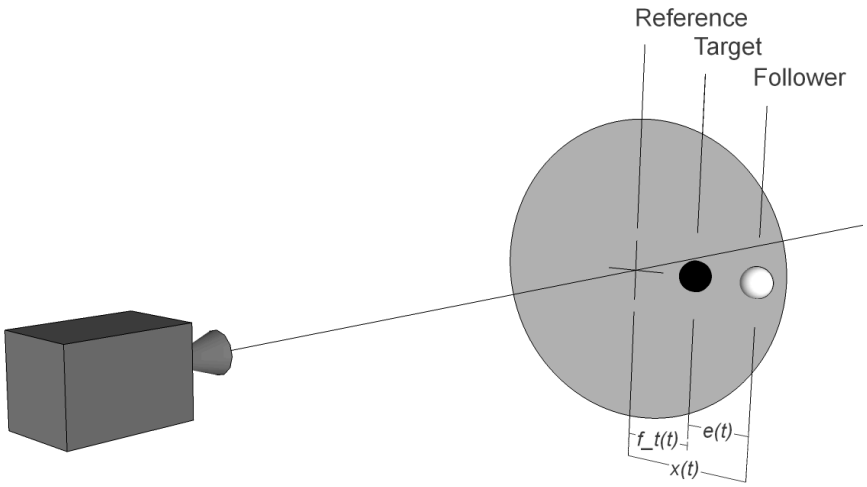


(a) Observation of a flat-plane compensatory display

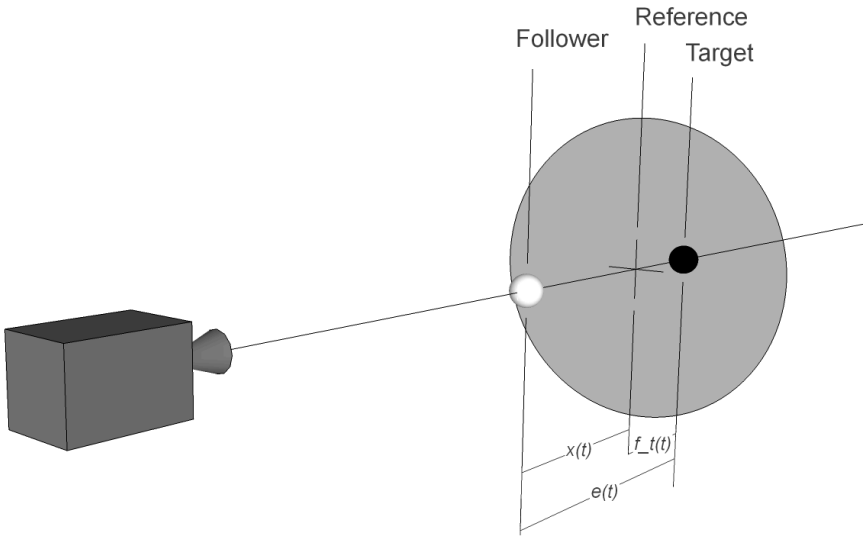


(b) Observation of a flat-plane pursuit display

Figure 4.2: The illustrations represent compensatory and pursuit display types. (a) The observer perceives a compensatory display at some distance with the system input $e(t)$ moving along an axis perpendicular to the observer's line of sight. (b) The observer perceives a pursuit display at some distance with the system inputs $f_t(t)$, $x(t)$ and $e(t)$ moving along an axis perpendicular to the observer's line of sight.



(a) Observation of a flat-plane pursuit display



(b) Observation of a depth pursuit display

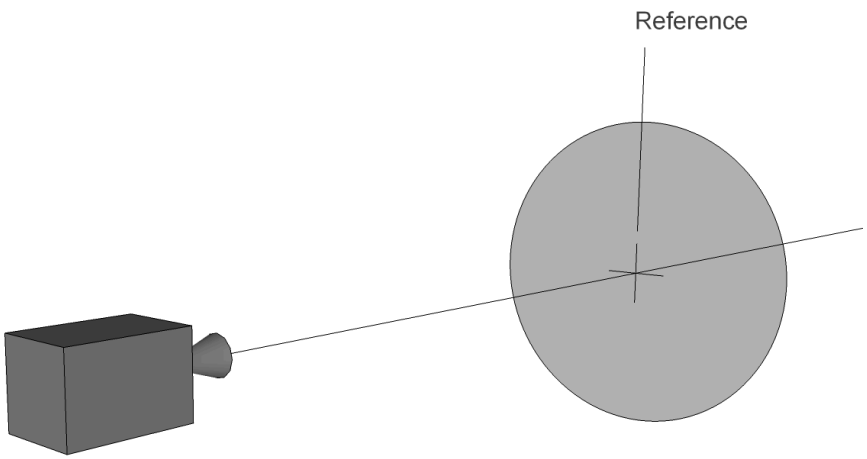
Figure 4.3: The illustrations represent pursuit display types for flat-plane and depth manual control tasks. (a) The observer once again perceives a pursuit display at some distance with the system inputs $f_t(t)$, $x(t)$ and $e(t)$ moving along an axis perpendicular to the observer's line of sight (which is identical to Figure 4.2b). (b) The observer perceives a pursuit display at some distance with the system inputs $f_t(t)$, $x(t)$ and $e(t)$ now moving along an axis parallel to and coinciding with the observer's line of sight

4.1.2. DEPTH CONTROL WITH BINOCULAR VISION

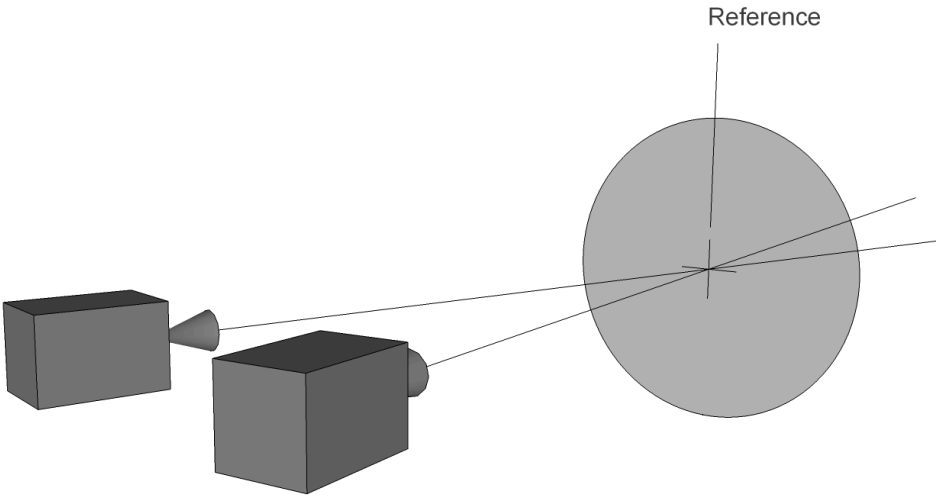
Knowledge on binocular vision which was previously covered in Chapter 2 can additionally be adopted to represent depth perception with two eyes. Consider firstly Figure 4.4. The alternations to the pursuit displays from flat-plane to depth control task settings have up until this point considered monocular vision (Figure 4.4a). The observer perceives a single image and has only one line of sight perfectly perpendicular to the image plane in gray. With the addition of another visual sensor (e.g. an eye or camera) focusing on the image plane, the observer now has binocular vision. The requisite inward eye movement, i.e., convergence was discussed in Section 2.2.1 and must be applied to the depth control task. The separated points of observation by some interocular distance and converging orientation of the lines of sight are shown in Figure 4.4b.

If the situation of a depth pursuit display is then supplemented with binocular vision, the true pursuit depth control task with stereoscopic vision is obtained. An illustration of the situation can be found in Figure 4.5a. Additional views in Figures 4.5b, 4.5c and 4.5d clarify the relative orientation of the target and follower to the observer. Note that target and follower end up at different orientations to the lines of sight of the left and right camera. With the addition of binocular vision, the observer senses binocular disparity as an new depth cue. However, like the non-linear image size scaling due to perspective, binocular disparity ,i.e., stereopsis might pose another non-constant factor in the perception of system signals with distance. Binocular disparity is the visual offset between two half images of the same object. Earlier in Section 2.2.2 a detailed analysis was already given on how the offsets are formed and how they assist a human being in distinguishing distance between two objects. The disparity is an angular measure for this offset and can be computed as given by Equation (2.3). Reversely analyzing, suppose a particular observer has a given sensitivity to binocular disparity also called stereoacuity. Of course, the interocular distance is fixed. By Equation (2.3), one may state that the distinguishable distance between two objects δ changes with the squared average distance to those objects. The minimum distinguishable distance between objects thus rapidly rises with increasing distance. Once again reflecting this aspect to the system signals, one may argue that the minimum perceivable system signals $f_t(t)$, $x(t)$ and $e(t)$ vary with their distance to the observer.

The implication of non-constant scaling of image size and perceivable depth difference on the control task is an important element to consider. These changes potentially affect the relation Display to Human Operator in the figure on Human-Machine cycle by McRuer (Figure 3.1) and with that our possibility to identify manual control behavior on depth control tasks. An analysis is therefore made in Section 6.2 to initially determine the sensitivity of object distance to system error signal $e(t)$.

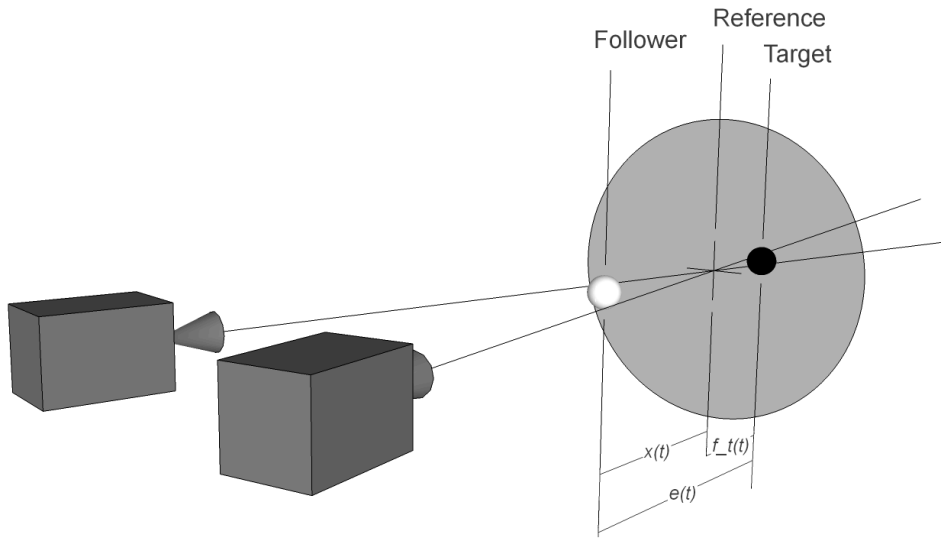


(a) Monocular vision on a reference point

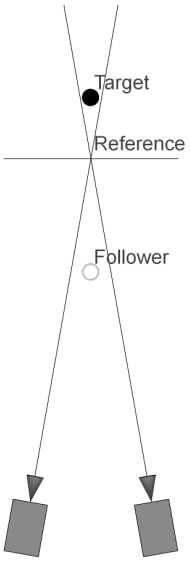


(b) Binocular vision on a reference point

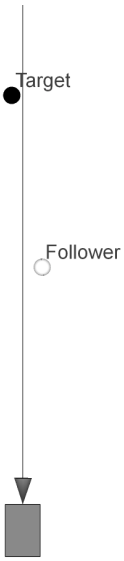
Figure 4.4: A graphical representation of monocular and binocular vision. (a) A monocular vision setup can have its line of sight exactly perpendicular to the plane's reference point. (b) The two lines of sight in a binocular vision setup requires convergence to focus both image central point (e.g. human eyes' foveae) to focus on the reference point.



(a) Observation with binocular vision of a depth pursuit display



(b) Top-down view of converging cameras on depth pursuit display



(c) Isolation of left optic's line of sight

(d) Isolation of right optic's line of sight

Figure 4.5: The illustrations show the concept of binocular vision on a depth control task with pursuit display. (a) A general overview of binocular vision on a display with system inputs $f_t(t)$, $x(t)$ and $e(t)$ moving directly toward or away from the average of both observation points. (b) The top-down view reveals the relative positions of the follower and target to both lines of sight. Note that the follower and target move toward or away from the system of cameras, not one camera specifically. (c & d) These isolation illustrate the positions of the follower and target relative to the lines of sight of the left and right points of observations. The perceived positions of the follower and target relative to the lines of sight differ and binocular disparity arises between the half images.

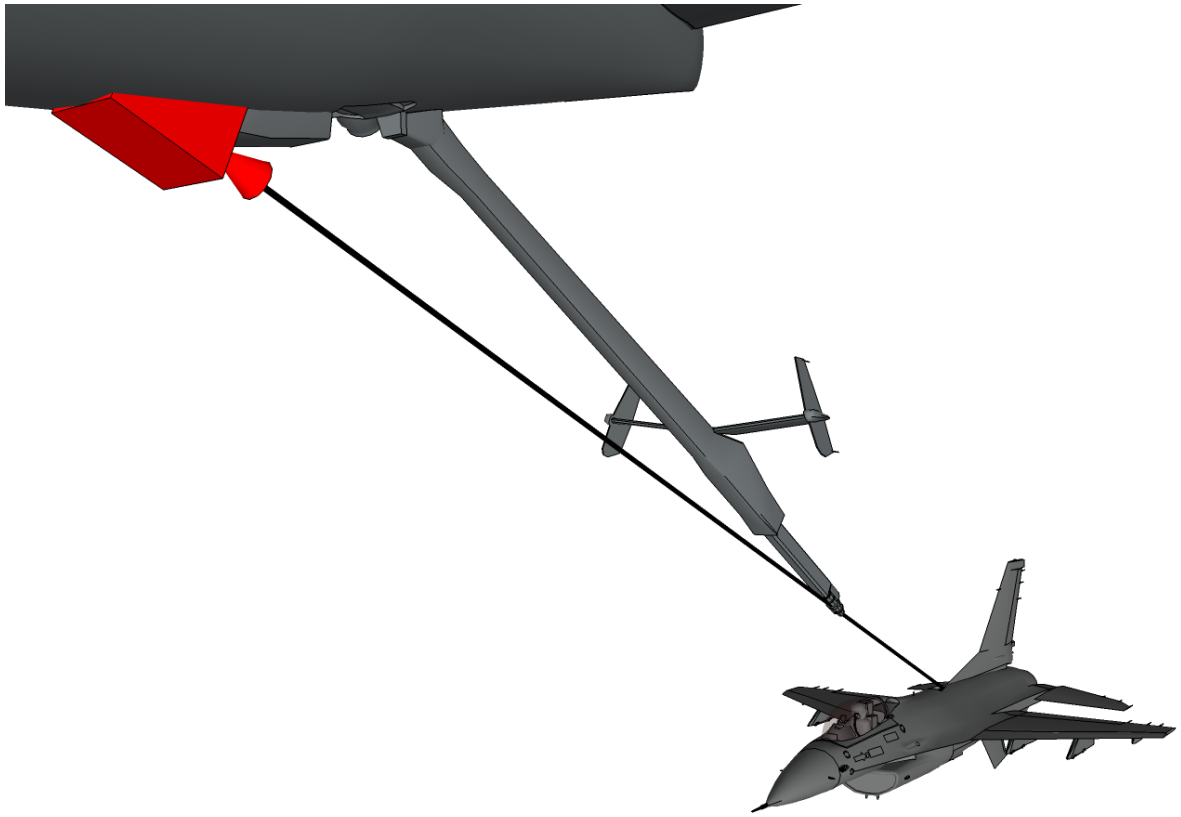
4.2. SIGNIFICANCE OF DEPTH CONTROL TASK

With the concept of depth control tasks now defined, lets reconsider the aerial refueling task of the boom operator introduced in Chapter 1. The boom operator controls a suspended refueling boom with rudder and elevator control surfaces. The refueling boom may be steered within its flight envelope by pitch and roll movements as well as extension of a telescopic piece. A receiver aircraft is holding position aft and slightly low of the tanker aircraft within reach of the refueling boom. The connection part of the refueling boom called the boom nozzle is ultimately connected to the receiver fuel opening known as the receptacle.

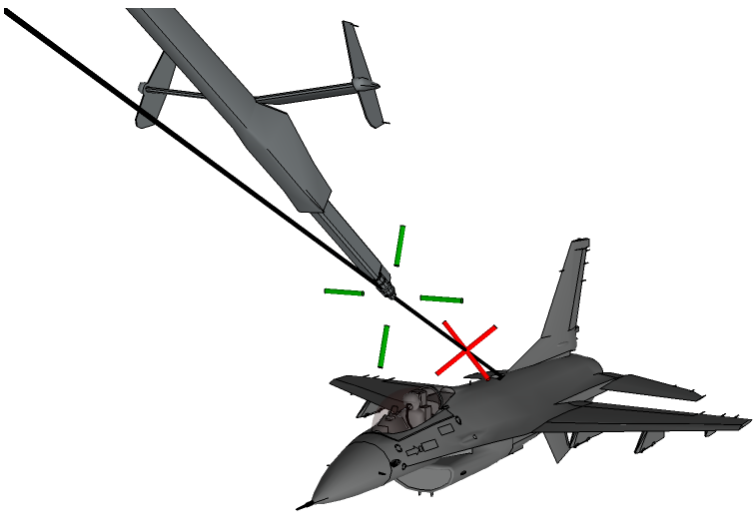
The Royal Netherlands Air Force deploys two KDC-10 tanker aircraft equipped with the Tanker Remote Vision System (TRVS) [15, 16]. The boom operator has therefore no direct vision on the environment where the refueling boom is coupled with the receiver. Instead, a stereoscopic camera and display system is installed to provide the boom operator with stereoscopic vision of the outside world directly behind the tanker aircraft flight deck. The TRVS stereoscopic cameras replace the human eyes as visual sensors of the environment. However, their interocular distance is increased to approximately 0.5 m and converge at the boom nozzle in neutral position (30° pitch below the body axis and 0° roll with 12 ft extension of the telescope). The extended interocular distance between the stereo cameras results in a increased stereoscopic vision referred to as hyper-stereoscopic vision. It essentially shifts the curve of binocular disparities in Figure 2.16 so small depth differences remain perceivable at distances further away from the observer. As a result, the fine depth distinction possible with binocular disparities in personal space can be extended to action space due to the hyper-stereoscopic camera system of the TRVS in Dutch KDC-10s.

The stereo cameras' position is highlighted with red color in Figure 4.6a. The line of sight is moreover included. For the situation shown in Figure 4.6, both the refueling boom nozzle and receiver receptacle are positioned exactly on the line of sight. These are practical representations of the earlier introduced follower and target, respectively. Figure 4.6b shows the highlighted follower and target. The boom operator's point of view is subsequently given in Figure 4.6c. One may deduce that the nozzle is placed perfectly on top of the receptacle in the flat viewing image plane, yet some distance remains between the nozzle and receptacle along the camera line of sight.

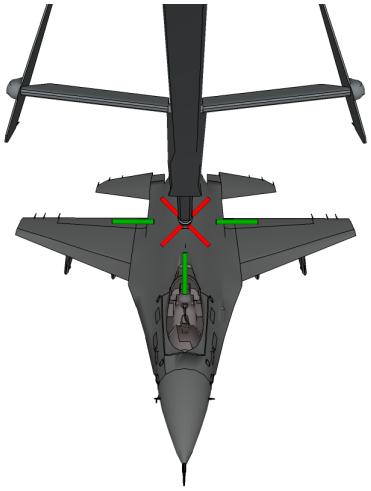
Although it must be noted that the aerial refueling task is by no means a *pure* depth control task, i.e., all three dimensions must be controlled by the boom operator, but the task certainly comprises control of the third dimension. As the perception of third dimensional signals was already found to differ from absolute flat-plane signals, much research has been conducted on the aspect of stereoscopic vision and depth perception in aerial refueling tasks [11–14]. However, it must be noted that these studies focus specifically on operational aspects of the United States military. For instance, some research is into relational parameters for aircrew performance with the purpose to extend knowledge on aircrew selection and medical examination in terms of depth perception [11, 14]. Somewhat more dedicated to control tasks are the research projects committed to operational use of (hyper)stereoscopic display setups in aerial refueling [12, 13]. These experiments test various display parameters e.g. quality of vision and introduction of hyper-stereo for impact on task performance. Although these studies imply a significant contribution of stereoscopic displays to control task performance and possibly difference in manual control behavior, their addition to understanding how depth control and stereoscopic vision affect control behavior is limited.



(a) Perspective view on boom (follower) and receptacle (target) system



(b) Visually enhanced boom (green crosshair) and receptacle (red crosshair) system



(c) Boom operator perspective

Figure 4.6: Illustrations of aerial refueling showing (a) A perspective view on a boom and receptacle system in an aerial refueling scenario where the boom operator's point of view is highlighted in red and the line of sight is shown as a black line; (b) An identical perspective view and highlighted follower (green crosshair on boom nozzle) and target (red crosshair at a distance to receptacle); (c) The perspective perceived by the boom operator with aforementioned follower and target visual enhancements

In fact, much of the research focusing on control tasks dealing with depth control consider some measure of performance [5, 6, 48–51]. These studies often provide a great scientific justification for the addition of stereoscopic displays in situations where the operator is heavily reliant on visual depth perception [48, 50]. A recent study by Karasinski and Robinson (2019) concluded no improved performance in general on the other hand. Several literature reviews point out the differences between experiment results on the adoption of stereoscopic displays three-dimensional control tasks [5, 6, 51]. McIntire (2014) bundled the results of over 180 research experiments over 50 years of time. The largest part of these studies (70%) focused on general human factors or human-machine interactions. The remaining experiments were carried out with a medical motivation. What stands out is the stronger support for stereoscopic 3D displays in medical literature (70% of all studies declare advantageous) compared to literature on human-machine interaction (55% of all studies declare advantageous). Additionally, McIntire distinguished the results by analysing possible advantages of stereoscopic 3D displays for six different categories of tasks:

1. Judgments of position and/or distances;
2. Finding/identifying/classifying objects;
3. Spatial manipulations of real or virtual objects;
4. Navigation;
5. Spatial understanding/memory/recall;
6. Learning/training/planning.

Interestingly, out of all categories, the category "Spatial manipulations of real or virtual objects" was found to benefit most (67% of all spatial appreciation studies) from the addition of stereoscopic 3D displays [6]. Considering only human-machine interaction studies, the number even rises slightly higher to 69%. Of course this category best represents the classical control tracking tasks in cybernetics studies and is therefore particularly interesting to the research project at hand. Additionally, McIntire states "*stereoscopic 3D displays seem to be most beneficial for depth-related tasks performed in the near-field (tasks in close spatial proximity to the viewer, such as spatial manipulations of objects). This is perhaps not surprising, given that in real-world viewing situations, the closer an object is to a viewer, the larger its potential binocular disparity cues, leading to a larger relative benefit of binocular cues over monocular cues when perceiving depth.*" This large contribution of binocular disparity was already found in the overview of dominant depth cues over the three ranges distinct ranges in Figure 2.16.

McIntire moreover states that stereoscopic 3D displays are especially useful for "*difficult, complex, or unfamiliar depth-related tasks, or for tasks where monocular cues were degraded or absent*." This statement is of great importance to the remote vision of the boom operator discussed earlier. One must recall the weighted additive model in addition with Figure 2.16 that the most dominant depth cues -binocular disparity, motion parallax and occlusion- may not at all be available or usable in an aerial refueling situation. Firstly motion parallax in terms of head movement is omitted due to fixed camera positions. Additionally, interposition in safe refueling conditions always results in the receiver aircraft behind visually positioned behind the refueling boom or not overlaid at all. In short, of all dominant depth cues, only binocular disparity is available and reliable. This once again supports the enhanced binocular disparities through hyper-stereoscopic vision of the TRVS in the KDC-10 tanker aircraft of the Royal Netherlands Air Force.

All in all, the use of stereoscopic 3D displays is no evident no-brainer for all depth related control tasks. McIntire concludes: "*since stereoscopic 3D displays seem to offer performance benefits for **specific** (depth-related spatial) tasks, these human factors concerns suggest that the technology should be wielded delicately and applied carefully in order to assure both high comfort and high performance for users*." Many studies focus on either qualitative and quantitative assessment of stereoscopic 3D displays in control tracking tasks. The results predominately concern measures of performance and conclusions often are not in agreement with other experiments. Moreover, the impact of depth control and stereoscopic displays on current knowledge of manual control behavior is unknown [18]. Mulder et al. (2018) recognize the increasing adoption of modern interfaces such as stereoscopic 3D displays, but state that an understanding of the impact of stereoscopic displays on human control behavior lacks in current knowledge and models.

5

PROPOSAL FOR RESEARCH

At this stage the reader has been introduced to numerous fields of study. Introductions are given to aspects of human vision and control to subsequently define the concept of depth control with stereoscopic vision. It has become clear how stereoscopic displays in human-machine interactions have been a focus point in many studies whilst simultaneously being far from understood in terms of human adaption and behavior with such interfaces.

This chapter serves to state the design choices to do following novel research into providing a first understanding of human control behavior in depth control tracking tasks with stereoscopic vision:

The objective of this research project is to develop an understanding of human control behaviour in manual depth control tracking tasks with stereoscopic vision *by* comparing pilot describing functions of flat-plane pursuit tracking tasks against pilot describing functions of depth pursuit tracking tasks with both monocular and binocular vision obtained through experimentally gathered data.

Section 5.1 covers all aspects of the design of the research experiment. The control tasks, experiment setup and procedures as well as participant selection and briefing are discussed. Section 5.2 enumerates the hypotheses that will be tested based on results from experimental data.

5.1. DESIGN OF EXPERIMENT

As the objective of this research states, conclusions will be based on data gathered from experiments on manual control tracking tasks. This section describes the various aspects that must be considered and selected.

5.1.1. CONTROL TASK

The participants will be tasked with controlling a single-axis tracking task in pursuit display setup. The control task adheres to the practical application of aerial refueling where both the boom nozzle and receiver receptacle can be clearly identified as follower and target, respectively. A simplification of the task is required to consider solely the third-dimensional movements of follower and target. A schematic representation of the pursuit task is given in Figure 5.1. Although the human controller may explicitly visually perceive all system input signals, i.e., target signal f_t, controlled system output x and control error e as indicated in the figure, a compensatory control strategy is expected by the operator as the controlled element dynamics are chosen as rate-controlled $H_c(s) = K/s$. That is, even though a pursuit display is shown to the operator, only the error signal e and therefore H_{p_e} are considered [46, 47, 52]. The surplus signals f_t and x are greyed out in Figure 5.1.

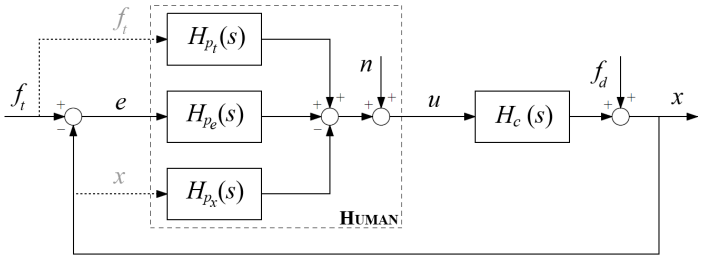


Figure 5.1: An illustration with schematic representation of a pursuit tracking task.

The figure moreover shows target signal $f_t(t)$ and output signal $f_d(t)$. The target signal sets the target position and the output signal adds noise to the follower. Should the human operator make use of more than solely the H_{p_e} response, the dual channel control strategy may still be identified [52–54]. The forcing functions $f_t(t)$ and $f_d(t)$ are constructed as quasi-random multisine signals using:

$$f_{t,d}(t) = \sum_{k=1}^{N_{t,d}} A_{t,d}(k)\sin(\omega_{t,d}(k)\,t + \phi_{t,d}(k)) \tag{5.1}$$

The specifications for both target signal f_t and disturbance signal f_d are given in Table 5.1. Both signals have been constructed threefold with unique phase shifts ϕ_t and ϕ_d. Experiment trials will run with alternating phase sets for f_t and f_d to further minimize signal recognition and therefore ensure random appearing forcing functions.

Table 5.1: Signal forcing function characteristics.

Target signal f_t							Disturbance signal f_d						
k	n_t	ω_t, rad/s	A_t, m	ϕ_t, rad			k	n_d	ω_d, rad/s	A_d, m	ϕ_d, rad		
				I	II	III					I	II	III
1	5	0.383	0.920	3.042	0.217	6.150	1	6	0.460	0.956	1.293	5.882	2.323
2	11	0.844	0.695	4.499	2.851	3.298	2	13	0.997	0.669	4.454	5.624	5.493
3	23	1.764	0.346	5.980	0.911	0.323	3	27	2.071	0.302	2.632	3.489	6.010
4	37	2.838	0.176	1.269	1.129	0.336	4	41	3.145	0.162	4.173	0.654	2.639
5	51	3.912	0.107	2.110	0.877	2.899	5	53	4.065	0.109	1.037	0.820	3.771
6	71	5.446	0.065	0.645	3.996	2.674	6	73	5.599	0.068	5.492	1.370	5.449
7	101	7.747	0.041	0.435	5.621	1.399	7	103	7.900	0.043	4.946	4.811	4.877
8	137	10.508	0.029	0.073	1.516	2.049	8	139	10.661	0.031	2.961	1.520	6.079
9	171	13.116	0.025	3.318	3.352	1.274	9	194	14.880	0.024	3.068	2.955	5.615
10	226	17.334	0.021	3.082	4.610	2.080	10	229	17.564	0.022	2.833	3.515	6.273

5.1.2. EXPERIMENT SETUP

The research encompasses a human-in-the-loop experiment at the Faculty of Aerospace Engineering at Delft University of Technology. The experiment platform is the 3D Aerial Refueling Simulator designed and built by multiSIM Ltd. Simulation & Training for the Royal Netherlands Air Force. The simulator can display stereoscopic three-dimensional environments with the 3D PluraView 28" 4K Monitor. The stereoscopic 3D display consists of two 28-inch monitors with 60 Hz frame rates and resolutions of 3840 by 2160 pixels per monitor. Each monitor displays a single half-image, i.e., the monitor presents the observer with a unique image representing the perspective of a single eye. A semi-transparent mirror is placed between the two monitors to see both half images -one directly through the mirror, one reflected onto the mirror- at approximately 80 cm distance. Passive polarized glasses are to be worn by the observer so only the appropriate half image is perceived by either eye. Figure 5.2 shows both the left- and right-hand control stick (CLS-E Active Force Feedback Joystick by Brunner Elektronik AG), as two are needed for realistic aerial refueling tasks. However, the experiment uses only the right-hand control stick. The control stick's travel reaches $\pm 20.5^\circ$ and the haptic characteristics can be customized using the manufacturer's control software.

Figure 5.2: An illustration of the 3D Aerial Refueling Simulator designed and built by multiSIM Ltd. Simulation & Training for the Royal Netherlands Air Force.

5.1.3. EXPERIMENT CONDITIONS

The experiment conditions vary for two aspects of perception. Firstly, four unique observation settings with respect to movement of the target and follower forms the first set of conditions. These conditions are summarized in Table 5.2.

Table 5.2: Experiment conditions.

Condition	Axis	Description	Interocular distance
FP	Vertical	Isometric (flat-plane) view	–
D-Mono	Depth	Monoscopic view	0 m
D-Stereo	Depth	Natural stereoscopic view	0.06 m
D-Hyper	Depth	Hyperstereoscopic view	0.5 m

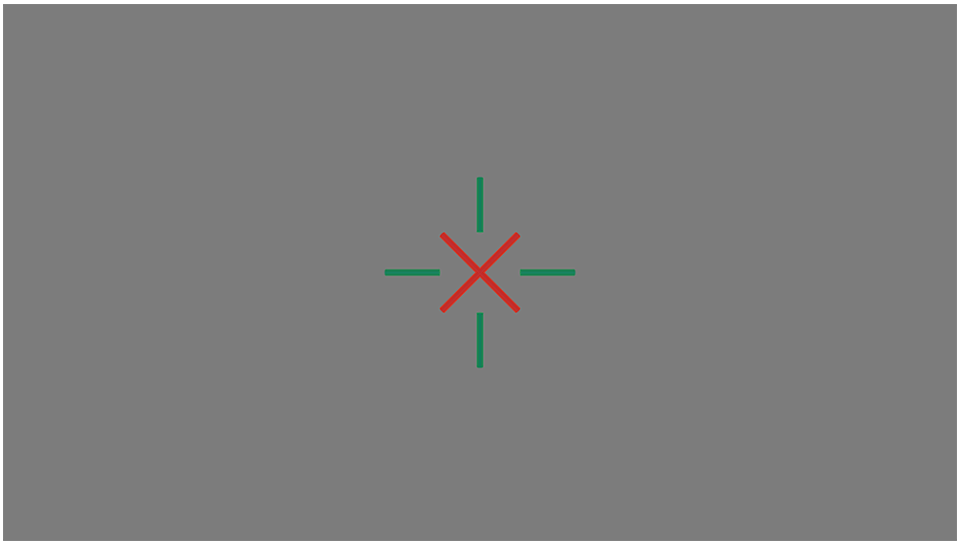
The condition *FP* is chosen as a reference condition to the bulk of flat-plane control task experiments. The subsequent conditions *D-Mono*, *D-Stereo* and *D-Hyper* comprise identical forcing functions and observation distance, but have coinciding lines of sight with the movement axes of $f_t(t)$ and $f_d(t)$. The depth conditions have a toe-in camera configuration rather than parallel configuration. Although the use of toe-in camera configurations for binocular disparity capable stereoscopic vision systems can be susceptible to keystoning effects, the objects of interest in this particular case are centered and so little distortion is expected [55].

The parallax geometry (see Figure 2.12) was selected based on the aerial refueling task. All conditions deal with teleoperation control task of a follower and target at a distance of approximately 18 m from the observer. This value was found in aerial refueling to be the distance between the stereo camera position and the neutral boom nozzle position. The value for distance to target E is therefore fixed at 18 m for all conditions. Interocular distance a is a condition variable based on monocular, binocular and TRVS by TNO eye distances and provided in Table 5.2 for the four conditions.

Due to the slightly different half images in conditions *D-Stereo* and *D-Hyper* stereopsis is induced as a additional depth cue. Most other depth cues are omitted to isolate the effects of stereoscopic vision. Only the depth cues of perspective and motion parallax (self-motion such as head movements excluded) are kept within the artificial environments. Special consideration of the orientation of visual representations (red and green reticles) of the target and follower is necessary to omit the dominant depth cue of occlusion. For any state of target and follower, one may not visually blockade the other.

In addition to the way the environment is observed, the environment itself is altered between two more experiment conditions. The first experiment environment is designed to adhere to previous experiments on manual control behavior in tracking tasks. The environment is shown in Figure 5.3a. The second environment is a simulated aerial refueling scenario. The much more advanced visual environment makes for a more complex analysis of available depth cue and impact on control behavior. Environment 1 (E1) is believed to provide a better basis for theoretical modelling of manual control behavior in depth control tasks. However, Environment 2 (E2) may provide insight in the applicability of found results on more visually advanced, i.e., more practical situations.

The follower and target, i.e., refueling boom and receptacle translate solely along the camera's line of sight to maintain pure depth movements. The task therefore remains exactly identical to the control task in E1. The line of sight in this environment is shown in Figure 4.6. The orientations of the visual representations of the follower and target (green and red reticles) must have slightly different orientations to not only exclude potential occlusion of either reticle, but also reticle with the actual target F-16 fighter jet and follower refuelling boom.



(a) Abstract environment E1 based on classical control tracking task experiments

(b) Aerial refueling environment E2 based on TNO TRVS boom operator display [16]

Figure 5.3: The experiment environment designs are shown by the two screenshots. (a) Two marker objects represent the follower and target elements in an empty abstract environment. (b) The two marker objects are slaved to physical objects to visually enhance follower (refueling boom) and target (receiver receptacle) in an simulated airborne environment.

5.1.4. PARTICIPANTS AND PROCEDURES

Two groups with different levels of experience are selected to investigate applicability of controller model to larger groups of population. Current results on performance measures for stereoscopic displays namely indicate possible difference between novice and expert group [6]. Novice group was selected to be (somewhat) familiar with control tasks, but not with stereoscopic 3D displays and (artificial) depth perception. The expert group on the other hand was selected for expertise in field of depth control task. These are experienced and operational refueling boom operators of the Royal Netherlands Air Force.

A sufficient number of participants for statistical significance is desirable. Additionally, to reduce the effect of chronological order for conditions, each possible sequence of conditions should ideally be followed an identical number of times. For the novice group, who will perform solely the four conditions in E1, sixteen participants were selected as a multiple of four. The expert group comprises all eight available boom operators.

Before the start of experiment all participants are firstly briefed on the experiment they are about to take part in. The experiment setup is introduced and all safety precautions are explained. Before any actions on the setup are taken, all participants will firstly do stereoacuity tests to ensure stereoacuity of 50-100 seconds of arc [31, 56, 57]. This step ensures undesirable biases due to inability to perceive binocular disparity depth cues are omitted. The participants are then shown all conditions and/or environments as training to familiarise participants with the task they are expected to perform.

The experimental data is then gathered during the actual measurement phase of the experiment. The conditions and environments are randomized as can be seen in Table 5.3. Each condition will be completed for five runs where a single run lasts ninety seconds.

Table 5.3: Experiment Latin Square design for participant groups

Participant	Condition						
	1	2	3	4	5	6	7
Novice 01	E1-D-Mono	E1-FP	E1-D-Stereo	E1-D-Hyper	-	-	-
Novice 02	E1-D-Hyper	E1-D-Stereo	E1-FP	E1-D-Mono	-	-	-
Novice 03	E1-D-Stereo	E1-D-Mono	E1-D-Hyper	E1-FP	-	-	-
Novice 04	E1-FP	E1-D-Hyper	E1-D-Mono	E1-D-Stereo	-	-	-
Novice 05	E1-FP	E1-D-Stereo	E1-D-Mono	E1-D-Hyper	-	-	-
Novice 06	E1-D-Mono	E1-D-Hyper	E1-D-Stereo	E1-FP	-	-	-
Novice 07	E1-D-Hyper	E1-D-Mono	E1-FP	E1-D-Stereo	-	-	-
Novice 08	E1-D-Stereo	E1-FP	E1-D-Hyper	E1-D-Mono	-	-	-
Novice 09	E1-D-Mono	E1-D-Hyper	E1-FP	E1-D-Stereo	-	-	-
Novice 10	E1-FP	E1-D-Stereo	E1-D-Hyper	E1-D-Mono	-	-	-
Novice 11	E1-D-Stereo	E1-FP	E1-D-Mono	E1-D-Hyper	-	-	-
Novice 12	E1-D-Hyper	E1-D-Mono	E1-D-Stereo	E1-FP	-	-	-
Novice 13	E1-D-Stereo	E1-D-Mono	E1-FP	E1-D-Hyper	-	-	-
Novice 14	E1-D-Hyper	E1-FP	E1-D-Mono	E1-D-Stereo	-	-	-
Novice 15	E1-FP	E1-D-Hyper	E1-D-Stereo	E1-D-Mono	-	-	-
Novice 16	E1-D-Mono	E1-D-Stereo	E1-D-Hyper	E1-FP	-	-	-
Expert 01	E1-D-Hyper	E1-FP	E1-D-Stereo	E1-D-Mono	E2-D-Hyper	E2-D-Stereo	E2-D-Mono
Expert 02	E1-FP	E1-D-Mono	E1-D-Hyper	E1-D-Stereo	E2-D-Mono	E2-D-Hyper	E2-D-Stereo
Expert 03	E1-D-Mono	E1-D-Stereo	E1-FP	E1-D-Hyper	E2-D-Stereo	E2-D-Mono	E2-D-Hyper
Expert 04	E2-D-Stereo	E2-D-Hyper	E2-D-Mono	E1-D-Stereo	E1-D-Hyper	E1-D-Mono	E1-FP
Expert 05	E2-D-Hyper	E2-D-Mono	E2-D-Stereo	E1-D-Hyper	E1-D-Stereo	E1-D-Mono	E1-FP
Expert 06	E2-D-Mono	E2-D-Stereo	E2-D-Hyper	E1-FP	E1-D-Mono	E1-D-Stereo	E1-D-Hyper
Expert 07	E1-D-Mono	E1-D-Hyper	E1-FP	E1-D-Stereo	E2-D-Hyper	E2-D-Stereo	E2-D-Mono
Expert 08	E2-D-Stereo	E2-D-Mono	E2-D-Hyper	E1-D-Stereo	E1-D-Mono	E1-D-Hyper	E1-FP

5.2. HYPOTHESES

The hypotheses that are to be tested with the experimental data are stated below. These hypotheses cover the four main variables that are analyzed on impact on manual control behavior in depth control tracking tasks with stereoscopic vision.

H1: Controlling the position of an object in depth given identical visual characteristics results in lower accuracy control compared to controlling its position in a flat plane.

The first hypothesis may be tested by comparison of results attained in the *FP* and *D-Mono* conditions. Both conditions have identical visual characteristics and differ by viewing angle on the axis of movement.

H2: Amplifying stereopsis in depth control tasks has a positive effect on control accuracy.

Between conditions *D-Mono*, *D-Stereo* and *D-Hyper* an increasing interocular distance is implemented, i.e., increasing the depth cue of stereopsis with subsequent experiment conditions. Result comparison between these three conditions may therefore accept or reject the second hypothesis.

H3: Visually realistic (background) environments positively stimulate control accuracy.

The practical applicability of results found in abstract environment E1 is tested by result comparison with real world environment E2. The third hypothesis is tested accordingly.

H4: Frequent usage of stereoscopic 3D displays during depth control tasks enhances control accuracy.

Holding results of the novice group and expert group against each other provides a basis for acceptance or rejection of the fourth hypothesis

The acceptance or rejection of the above stated hypotheses follows after conduction of the experiments and analysis of the results. These are published in a paper following this document.

6

Analysis of Preliminary Results

The experiment described in the previous chapter is carried out prior to actual experiment phase in the research project verify desired functioning of the setup and conditions. This chapter comprises some preliminary results to analyze whether the envisioned experiments yields the aimed results. Additionally, an analysis of the attained magnitude of error as a function of the distance to the observer is given. This is done to investigate the impact of image scaling due to distance on a measure of task performance.

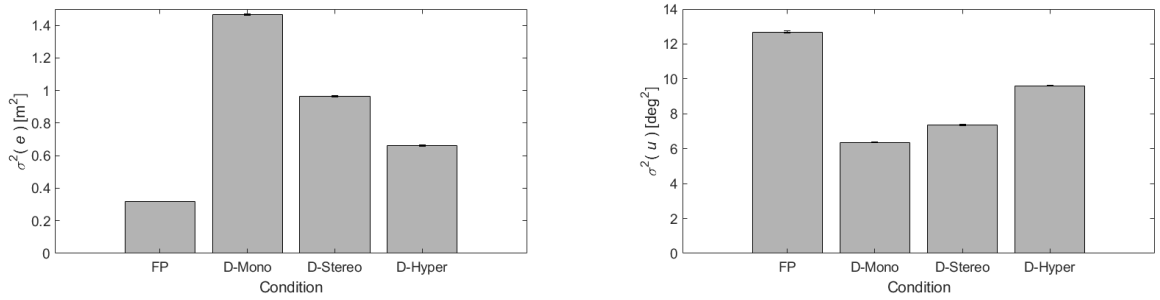
The preliminary results discussed in this chapter are obtained from a single participant ($n = 1$) not taking part in the actual experiment. The participant was deemed experienced with control tracking tasks, but not with depth perception and stereoscopic 3D displays. The participant scored highest possible stereoacuity results in the Titmus Fly test (40 seconds of arc) and TNO test (15 seconds of arc).

Section 6.1 provides a general indication of difference between the four experiment conditions by showing the variances of error and control effort. A second analysis of error is subsequently done to identify the relation of magnitude of error with observation distance. The results are given in Section 6.2. Lastly, in Section 6.3, an attempt is done to model human control behavior by constructing the operator describing functions for the four conditions.

6.1. Tracking Performance and Control Effort

The task performance during the preliminary experiment rehearsal is provided by means of the variance in error. Additionally, the variance of control effort is given. These two forms of result are believed to give a first insight in how the conditions differ to the participant. Both plots are given in Figure 6.1.

Consider firstly the variance of error plot in Figure 6.1a. The *FP*-condition has a much lower variance compared to all other depth conditions. Especially *D-Mono* is a factor of almost five times higher than **H1**. The difference between these two conditions is merely the orientation of the line of sight with respect to the axis of movement of the target and follower. Subsequent depth conditions then show an increasing drop in error variance as the interocular distance is increased, but error variance remains higher than the *FP* even if hyper-stereoscopic vision is introduced in *D-Hyper*. A glance at the variance in control effort shows somewhat similar indications. The control effort drops from *FP* to *D-Mono* and gradually rises again when stereopsis is increased for the depth conditions. Based on these preliminary results, a strong indication is found that **H1** and **H3** may be true.



(a) Variance analysis of control error *e*

(b) Variance analysis of controller input *u*

Figure 6.1: Graphical illustrations of two measurements from preliminary experiment trials. The graphs show results of one single participant over five different runs. (a) The variance of error signal *e* as a measure of task performance. (b) The variance of input signal *u* as a measure of control effort.

6.2. ANALYSIS OF DEPTH CONTROL ERROR SENSITIVITY

Earlier in this document the statements by Wickens indicated a potential ambiguity due to lower accuracy of absolute distance [7]. As the distance to a target and/or follower increases, it was found that the perceived error scaled by perspective. In essence, this may lead to non-constant control behavior for different parts of the movement envelope of system states. This section serves to analyse how magnitude of error holds with different distances to target

The preliminary test experiment tasked participant with control of an object at 18.0 m from the operator. The object moved along an axis with distances of ± 2.0 m. The horizontal axis in Figure 6.2 shows the position of $x_t(t)$. A value of 0.0 m indicates 18.0 m from the operator where positive values means the target is closer to the observer and negative values further away. The vertical axis shows the mean of the error signal of five trials.

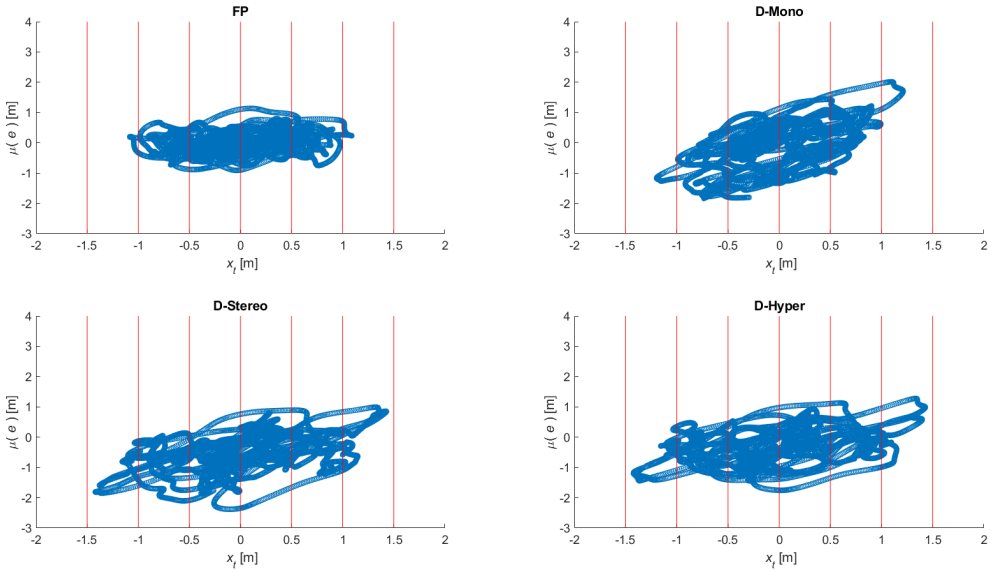


Figure 6.2: Results of error analysis over the range of movement (positive values for position indicate in front of convergence point and negative values behind). The left column shows every time step's mean error over the movement range. The right column shows box plots for the specific sections (section is 0.5 m) of the movement range.

To better visualize the magnitude of error, the range of motion has been divided into eight sections in Figure 6.3. The box plots show the results for the magnitude of error attained in each particular section. The behavior of error magnitude with observation distance can so be graphically represented.

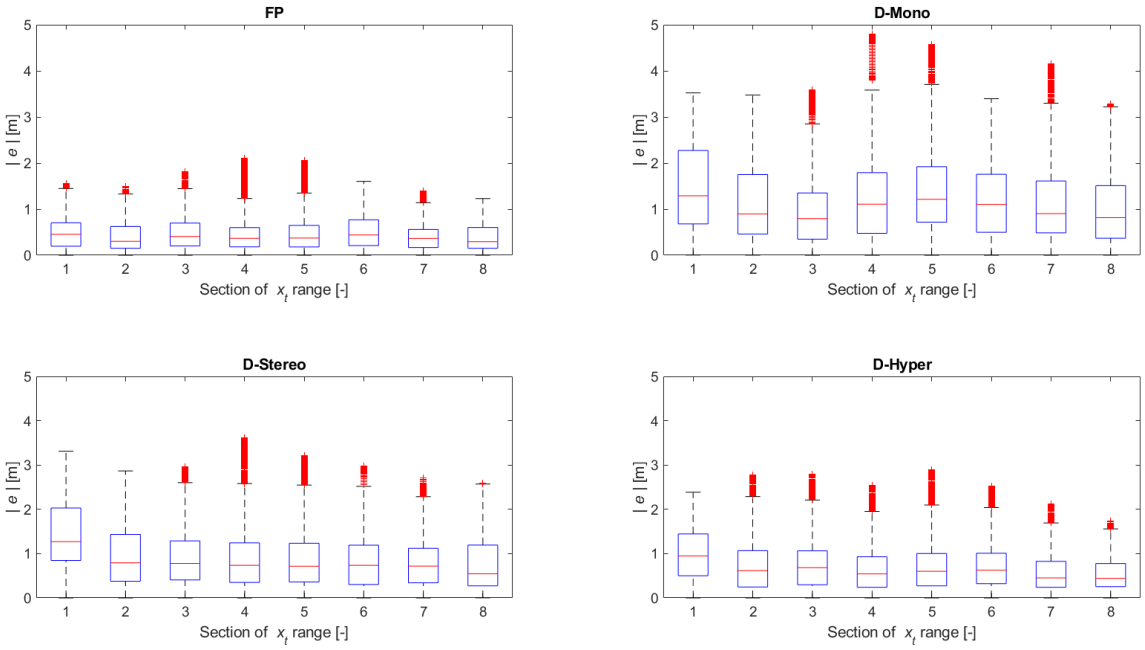


Figure 6.3: Results of error analysis over the range of movement (positive values for position indicate in front of convergence point and negative values behind). The left column shows every time step's mean error over the movement range. The right column shows box plots for the specific sections (section is 0.5 m) of the movement range.

Although no claim can be made that the error is constant over the full range of the system's motion, it can hardly be said that the error increases with increasing distance. If error would have significantly scaled with the distance to the observer, larger magnitudes of error should be found for negative values of $x_t(t)$, i.e., in lower sections 1-4. Except for the extremes near ± 2.0 m from the neutral point (sections 1 and 8 where there are only a limited number of data points at those values for $x_t(t)$), this is not clearly deduced from Figure 6.3. For now, the effect of perspective scaling on system signals $f_t(t)$, $x_t(t)$ and $e(t)$ is not deemed a void to the experiment and subsequent required identification techniques. A similar analysis should be done to the full set of experimental data after the experiments have been conducted to determine similar or different sensitivities to error perception.

6.3. MODEL OF HUMAN CONTROLLER

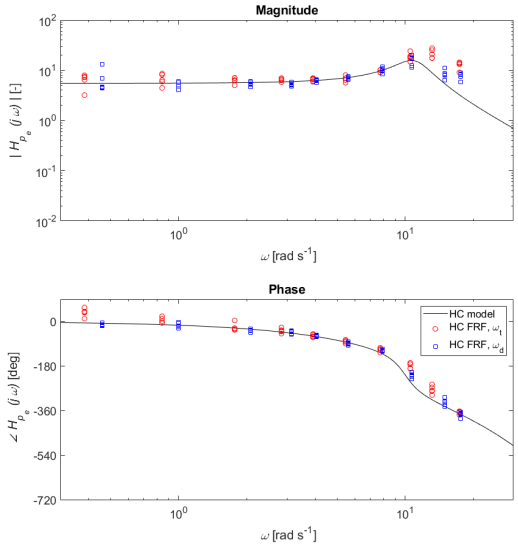
Previous methods of data analysis -like many depth control studies before- compare various measures of performance to conclude on the impact of control variables on the task. One may actually identify and model the participant's control behavior using the experiment data. The frequency response function of H_{p_e} can be calculated by Fourier transformation of error signal e and control signal u. The human operator's error response dynamics H_{p_e} can be computed:

$$H_{p_e}(j\omega_{t,d}) = \frac{U(j\omega_{t,d})}{E(j\omega_{t,d})} \tag{6.1}$$

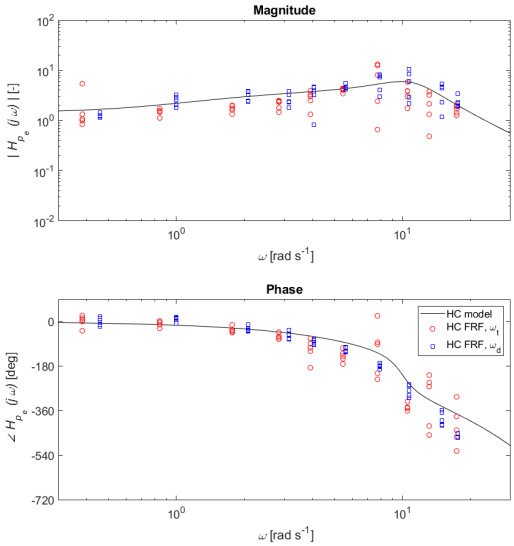
A model of H_{p_e} is initially estimated using the human operator model [1, 2, 18]:

$$H_{p_e}(j\omega) = K_p e^{j\omega\tau_p} \frac{\omega_{nm}^2}{(j\omega)^2 + 2\zeta_{nm}\omega_{nm} j\omega + \omega_{nm}^2} \tag{6.2}$$

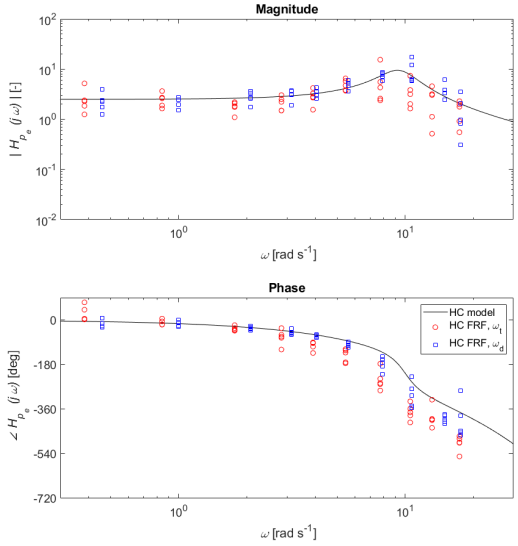
where K_p serves as the human operator's gain and τ_p the time delay. Moreover, the neuromuscular system (NMS) is included in the model by the natural frequency (ω_{nm}) and damping ratio (ζ_{nm}) of that system.



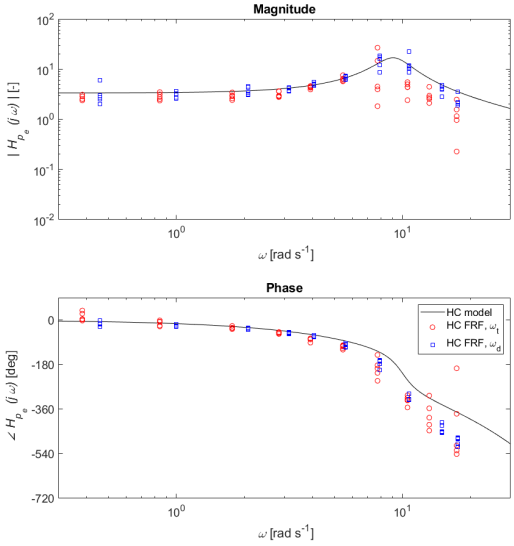
(a) H_{p_e} FRF and computed model for FP condition



(b) H_{p_e} FRF and computed model for D-Mono condition



(c) H_{p_e} FRF and computed model for D-Stereo condition



(d) H_{p_e} FRF and computed model for D-Hyper condition

Figure 6.4: The four graphs represent the frequency response function (FRF) estimations of the human operator's error response dynamics H_{p_e} for the four different experiment conditions.

Observations of the graphs of Figure 6.4 firstly show no apparent difference between the frequency responses of the target and disturbance signals at ω_t and ω_d, respectively. The preliminary results in Figure 6.4 therefore suggest that the human operator made no use of any additional signals in addition to H_{p_e}. They give a strong indication the human operator is showing compensatory control behavior.

What is striking and quite divergent from other magnitude frequency responses is the modelled frequency response function for the *D-Mono* condition. A possible explanation could be the large errors found between the target and follower in Figure 6.1a, but the odd shape in the magnitude response function is an interesting find. All in all, it is believed that the operator's control behavior is identifiable and can most likely be modelled using earlier described methods. Analysis of actual experimental data should be done to tell how well the different experiment conditions can be modelled and with what certainty.

7

CONCLUSION

The report in this document concludes the literature study on the human visual system, manual control behavior in control tracking tasks and the definition of depth control tasks. Moreover, the preliminary phase of this research is closed off to continue with the actual experiments and subsequent data analysis. An endeavour was made to introduce the fundamental bases for various fields of study and motivate the need for research into human control behavior in depth control tracking tasks. The research has the main research question:

What effect does binocular vision in manual depth control tracking tasks have on human control behavior?

The answering of the main research question will have to done after the experiments have been conducted as will subquestions 6-9. This document is concluded with answering subquestions 1-5 based on Chapter 2, 3 and 4.

1. **What are the key variables affecting depth perception in the human visual system?**
   Depth perception, as a high-order neural process in the human visual system, is a form of visual perception where the observer is informed about third-dimensional distances or shapes. Depth perception is a summation of depth cues obtained through either physiological, pictoral or motion sources. Numerous visual sources of information may assist the observer in three-dimensional vision, although binocular disparity, motion parallax and occlusion are believed to be the most dominant depth cues.

2. **What are the key variables affecting manual control behavior in flat-plane control tracking tasks?**
   The variables affecting human control behavior in a human-machine system are identified by McRuer and Jex: task variables, environmental variables, operator-centered variables and procedural variables. Of all variables, they task variables known as forcing functions, displays, manipulators and controlled elements are believed to be the key variables affecting human control behavior.

3. **What are the characteristics of manual control behavior in flat-plane control tracking tasks?**
   The characteristics of manual control behavior are highly dependent on the task variables defined by McRuer and Jex. Different models describe different control behavior types like the crossover model for compensatory control behavior. Since the focus of this literature study focused mainly on the compensatory control behavior, the main characteristics of manual control behavior in scope of this research is the crossover model with the addition of a second order system to accommodate for the human neuromuscular system.

4. **What is currently known about manual control behavior in depth control tracking tasks?**
   It was found that our understanding of the implication of controlling an object along the third-dimension rather than in a flat plane on human control behavior is rather limited. Many studies on the effect of controlling depth compared to flat-plane or perspective view control tasks consider some measure of performance. These are indicative of a human operator finding the control of depth more difficult, but fail to explain the adaption and control behavior of the human operator.

5. **What is currently known about the effects of binocular vision in depth control tracking tasks?**
   Binocular vision has been adopted in control tracking tasks quite extensively over the last years. Various industries have a growing interest in a stereoscopic three-dimensional display as a modern interface within their system. As a result, many research experiments have been conducted to establish the potential advantageous implementation of these displays. Results on whether the implementation is useful are not quite complementary, however. Numerous literate studies report varying results found among various applications. The usefulness of binocular vision in depth control tracking tasks is said to be subject to parameters like availability of other depth cues and nature of the control task.

The remaining unanswered subquestions as well as the main research questions shall be answered after the experiment proposed in this document is conducted. Ultimately results through experimental data may help in doing a first attempt to understand human adaptation to and behavior in manual depth control tracking tasks with stereoscopic vision.

BIBLIOGRAPHY

[1] D. T. McRuer, D. Graham, E. S. Krendel, and W. J. Reisener, *Human Pilot Dynamics in Compensatory Systems: Theory, Models, and Experiments with Controlled Element and Forcing Function Variations*, Tech. Rep. AFFDL-TR-65-15 (Air Force Flight Dynamics Laboratory, Wright-Patterson Air Force Base (OH), 1965).

[2] D. T. McRuer and H. R. Jex, *A Review of Quasi-Linear Pilot Models,* IEEE Transactions on Human Factors in Electronics **HFE-8**, 231 (1967).

[3] D. T. McRuer and D. H. Weir, *Theory of Manual Vehicular Control,* IEEE Transactions on Man-Machine Systems **10**, 257 (1969).

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

[5] J. P. McIntire, P. R. Havig, and E. E. Geiselman, *What is 3D Good for? A Review of Human Performance on Stereoscopic 3D Displays,* in *Head- and Helmet-Mounted Displays XVII; and Display Technologies and Applications for Defense, Security, and Avionics VI*, Vol. 8383, edited by P. L. Marasco, P. R. H. II, D. D. Desjardins, and K. R. Sarma, International Society for Optics and Photonics (SPIE, 2012) pp. 280 – 292.

[6] J. P. McIntire, P. R. Havig, and E. E. Geiselman, *Stereoscopic 3D Displays and Human Performance: A Comprehensive Review,* Displays **35**, 18 (2014).

[7] C. D. Wickens, *Three-dimensional Stereoscopic Display Implementation: Guidelines Derived from Human Visual Capabilities,* in *Stereoscopic Displays and Applications*, Vol. 1256, edited by J. O. Merritt and S. S. Fisher, International Society for Optics and Photonics (SPIE, 1990) pp. 2 – 11.

[8] T. B. Sheridan, *Teleoperation, Telerobotics and Telepresence: A Progress Report,* Control Engineering Practice **3**, 205 (1995).

[9] P. Batsomboon and S. Tosunoglu, *A Review of Teleoperation and Telesensation System,* in *1996 Florida Conference on Recent Advanced in Robotics, Florida Atlantic University, Florida* (1996).

[10] Modrzejewski, F., *European Air Refueling Training 2015,* Available at `http://foto.poork.pl/en/fotoreportaze/european-air-refueling-training-2015/` (accessed: 2020-05-15) (2015).

[11] M. D. Winterbottom, J. P. Gaska, S. L. Wright, S. Hadley, C. Lloyd, H. J. Gao, F. Tey, and J. P. McIntire, *Operational Based Vision Assessment Research: Depth Perception,* (2014).

[12] M. D. Winterbottom, *Individual Differences in the Use of Remote Vision Stereoscopic Displays,* (2015).

[13] M. D. Winterbottom, C. Lloyd, J. P. Gaska, S. Wright, and S. Hadley, *Stereoscopic Remote Vision System Aerial Refueling Visual Performance,* Electronic Imaging **2016**, 1 (2016).

[14] M. D. Winterbottom, C. Lloyd, J. P. Gaska, L. Williams, E. Shoda, and S. Hadley, *Investigating Aircrew Depth Perception Standards Using a Stereoscopic Simulation Environment,* Electronic Imaging **2017**, 29 (2017).

[15] P. Bijl, S. De Vries, and J. Kalthof, *Upgrading the KDC-10 RARO Refueling Vision System: WP3b*, Confidential Internal Report TNO-DV3 2005-A 161 (TNO, 2005).

[16] J. C. P. Bol, L. van Breda, and M. Ringers, *Tanker Remote Visual System (TRVS)*, Tech. Rep. DenV S070047 (TNO, 2007).

[17] A. Schoor (Royal Netherlands Air Force), *Air to Air Refueling van F-16 Door de KDC-10 van RNLAF,* Available at `https://nimh-beeldbank.defensie.nl/foto-s/` (accessed: 2020-05-15) (2007).

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

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

[20] A. J. Grunwald and S. J. Merhav, *Vehicular Control by Visual Field Cues,* IEEE Transactions on Systems, Man, and Cybernetics **6**, 835 (1976).

[21] M. Mulder and J. A. Mulder, *Cybernetic Analysis of Perspective Flight-Path Display Dimensions,* Journal of Guidance, Control, and Dynamics **28**, 398 (2005).

[22] D. Purves, G. J. Augustine, D. Fitzpatrick, W. C. Hall, A. LaMantia, J. O. McNamara, and S. M. Williams, *Neuroscience*, 3rd ed. (Sinauer Associates, Inc., 23 Plumtree Road, Sunderland, Massachusetts 01375, 2004) ISBN: 0-87893-725-0.

[23] The Discovery Eye Foundation, *The Optic Nerve and its Visual Link to the Brain,* Available at `https://discoveryeye.org/optic-nerve-visual-link-brain/` (accessed: 2020-03-24) (2015).

[24] J. J. Gibson, *The Perception of the Visual World* (Houghton Mifflin Company, Boston, 1950).

[25] I. P. Howard and B. J. Rogers, *Binocular Vision and Stereopsis* (Oxford University Press, Inc., 198 Madison Avenue, New York, New York 10016, 1995) ISBN: 0-19-508476-4.

[26] L. A. Remington and D. Goodwin, *Clinical Anatomy of the Visual System*, 3rd ed. (Elsevier Butterworth-Heinemann, 3251 Riverport Lane, St. Louis, Missouri 63043, 2012) ISBN: 978-1-4377-1926-0.

[27] J. R. Leigh and D. S. Zee, *The Neurology of Eye Movements*, 3rd ed. (Oxford University Press, Inc., 198 Madison Avenue, New York, New York 10016, 1999) ISBN: 0-19-512972-5.

[28] M. Fahle, *Wozu Zwei Augen?* Naturwissenschaflen **74**, 383 (1987).

[29] B. Sweet and M. Kaiser, *Depth Perception, Cueing, and Control,* AIAA Modeling and Simulation Technologies Conference (2011), 10.2514/6.2011-6424.

[30] L. Kaufman, *Sight and Mind: An Introduction to Visual Perception* (Oxford University Press, Inc., 1974) ISBN: 0-19-501763-3.

[31] J. Lee and A. McIntyre, *Clinical Tests for Binocular Vision,* Eye **10**, 282285 (1996).

[32] J. P. Wann, S. Rushton, and M. Mon-Williams, *Natural Problems for Stereoscopic Depth Perception in Virtual Environments,* Vision Research **35**, 2731 (1995).

[33] D. M. Hoffman, A. R. Girshick, K. Akeley, and M. S. Banks, *Vergence Accommodation Conflicts Hinder Visual Performance and Cause Visual Fatigue,* Journal of Vision **8**, 33 (2008).

[34] L. Lipton, *StereoGraphics Developers Handbook* (StereoGraphics Corp., 198 Madison Avenue, New York, New York 10016, 1997) ISBN: 0-19-508476-4.

[35] C. D. Wickens, S. Todd, and K. Seidler, *Three-Dimensional Displays: Perception, Implementation, and Applications*, Tech. Rep. CSERIAC-SOAR-89-001 (University of Dayton Research Institute, 1989).

[36] S. Nagata, *How to Reinforce Perception of Depth in Single Two-Dimensional Pictures,* in *Pictorial Communication in Virtual and Real Environments* (Taylor & Francis, Inc., USA, 1991) p. 527545.

[37] J. Cutting and P. Vishton, *Perceiving Layout and Knowing Distances: The Interaction, Relative Potency, and Contextual Use of Different Information about Depth,* Percept. Space Mot **5**, 69 (1995).

[38] A. Tustin, *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**, 190 (1947).

[39] J. I. Elkind, *Characteristics of Simple Manual Control Systems*, Ph.D. thesis, Massachusetts Institute of Technology (1956).

[40] T. B. Sheridan, *Humans and Automation: System Design and Research Issues* (John Wiley & Sons, Inc., USA, 2002).

[41] D. M. Pool, *Objective Evaluation of Flight Simulator Motion Cueing Fidelity Through a Cybernetic Approach*, Ph.D. thesis, Delft University of Technology, Faculty of Aerospace Engineering (2012).

[42] H. J. Damveld, G. C. Beerens, M. M. van Paassen, and M. Mulder, *Design of Forcing Functions for the Identification of Human Control Behavior,* Journal of Guidance, Control, and Dynamics **33**, 1064 (2010).

[43] E. S. Krendel and D. T. McRuer, *A Servomechanics Approach to Skill Development,* Journal of the Franklin Institute **269**, 24 (1960).

[44] P. M. T. Zaal, D. M. Pool, J. de Bruin, M. Mulder, and M. M. van Paassen, *Use of Pitch and Heave Motion Cues in a Pitch Control Task,* Journal of Guidance, Control, and Dynamics **32**, 366 (2009).

[45] F. M. Nieuwenhuizen, P. M. T. Zaal, M. Mulder, M. M. van Paassen, and J. A. Mulder, *Modeling Human Multichannel Perception and Control Using Linear Time-Invariant Models,* Journal of Guidance, Control, and Dynamics **31**, 999 (2008).

[46] R. J. Wasicko, D. T. McRuer, and R. E. Magdaleno, *Human Pilot Dynamic Response in Single-loop Systems with Compensatory and Pursuit Displays*, Tech. Rep. AFFDL-TR-66-137 (Air Force Flight Dynamics Laboratory, 1966).

[47] R. A. Hess, *Pursuit Tracking and Higher Levels of Skill Development in the Human Pilot,* IEEE Transactions on Systems, Man, and Cybernetics **SMC-11**, 262 (1981).

[48] W. S. Kim, S. R. Ellis, M. E. Tyler, B. Hannaford, and L. W. Stark, *Quantitative Evaluation of Perspective and Stereoscopic Displays in Three-Axis Manual Tracking Tasks,* IEEE Transactions on Systems, Man, and Cybernetics **17**, 61 (1987).

[49] J. A. Karasinski and S. K. Robinson, *Evaluating Augmented Reality in a Three-Axis Manual Tracking Task,* in *Proceedings of the AIAA Modeling and Simulation Technologies Conference, San Diego (SA)*, AIAA-2019-1227 (2019).

[50] W. S. Kim, F. Tendick, and L. W. Stark, *Visual Enhancements in Pick-and-Place Tasks: Human Operators Controlling a Simulated Cylindrical Manipulator,* IEEE Journal on Robotics and Automation **3**, 418 (1987).

[51] S. Dixon, E. Fitzhugh, and D. Aleva, *Human Factors Guidelines for Applications of 3D Perspectives: A Literature Review,* in *Display Technologies and Applications for Defense, Security, and Avionics III*, Vol. 7327, edited by J. T. Thomas and D. D. Desjardins, International Society for Optics and Photonics (SPIE, 2009) pp. 172 – 182.

[52] M. C. Vos, D. M. Pool, H. J. Damveld, M. M. van Paassen, and M. Mulder, *Identification of Multimodal Control Behavior in Pursuit Tracking Tasks,* in *Proceedings of the 2014 IEEE International Conference on Systems, Man, and Cybernetics, San Diego (CA)* (2014) pp. 69–74.

[53] M. M. van Paassen and M. Mulder, *Identification of Human Operator Control Behaviour in Multiple-Loop Tracking Tasks,* in *Proceedings of the Seventh IFAC/IFIP/IFORS/IEA Symposium on Analysis, Design and Evaluation of Man-Machine Systems, Kyoto Japan* (Pergamon, 1998) pp. 515–520.

[54] A. van Lunteren, *Identification of Human Operator Describing Function Models with One or Two Inputs in Closed Loop Systems*, Ph.D. thesis, Delft University of Technology, Faculty of Mechanical Engineering (1979).

[55] Z. Gao, A. Hwang, G. Zhai, and E. Peli, *Correcting Geometric Distortions in Stereoscopic 3D Imaging,* PLoS One **13**, 33 (2008).

[56] S. Lee and N. Koo, *Change of Stereoacuity with Aging in Normal Eyes,* Korean journal of ophthalmology : KJO **19**, 136 (2005).

[57] M. E. F. Piano, L. P. Tidbury, and A. R. OConnor, *Normative Values for Near and Distance Clinical Tests of Stereoacuity,* Strabismus **24**, 169 (2016).

III

BOOK OF APPENDICES

A

Task Performance and Control Effort

The task performance and control effort analysis of Part I is presented in this appendix with the variance in either e or u calculated per participant. The plots combine task performance and control effort for *PG1*, *PG2* in *E1* and *PG2* in *E2*.

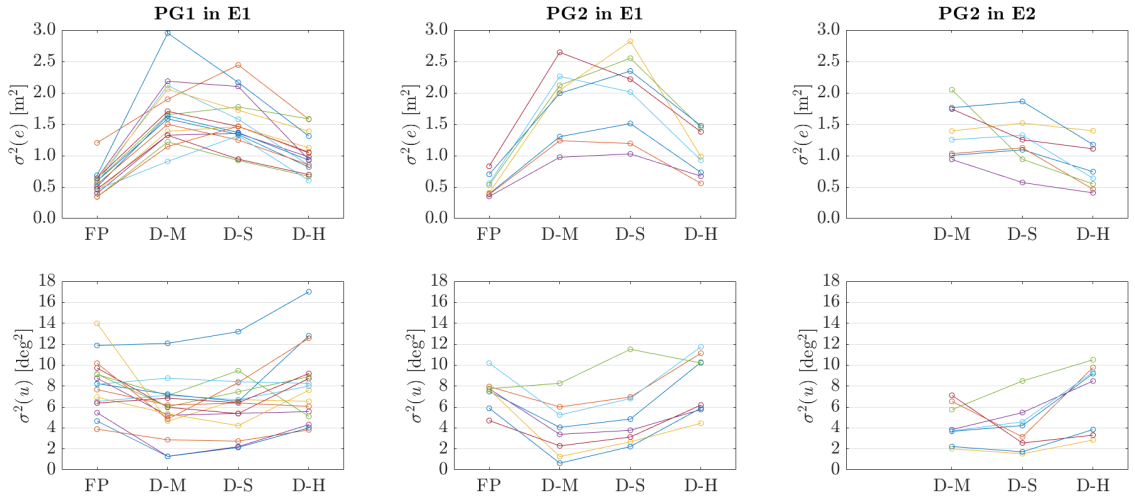


Figure A.1: Task performance and control effort in variance of control error e and control input u for *PG1* in *E1*, *PG2* in *E1* and *PG2* in *E2*.

B

Control Correlation

This section of the book of appendices provides plots with the control correlation coefficients for each individual participant and display condition combination. In addition to the control correlation plot in Part I, this appendix provides both the mean control correlation coefficients per frequency of either the target or disturbance signal (top plot) as well as the correlation coefficients per individual measurement run (bottom plot).

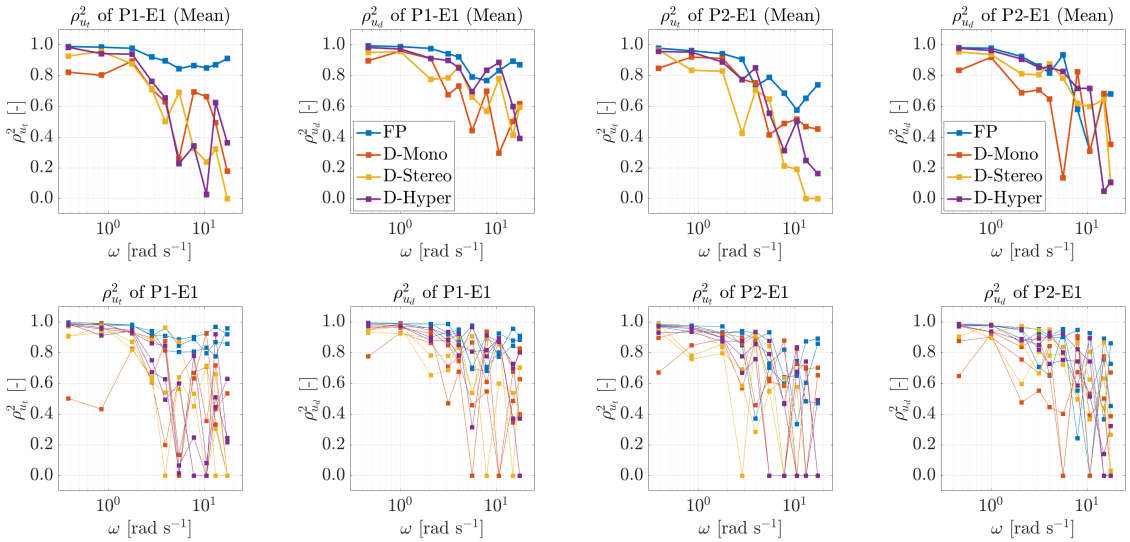


Figure B.1: Correlation *P1*

Figure B.2: Correlation *P2*

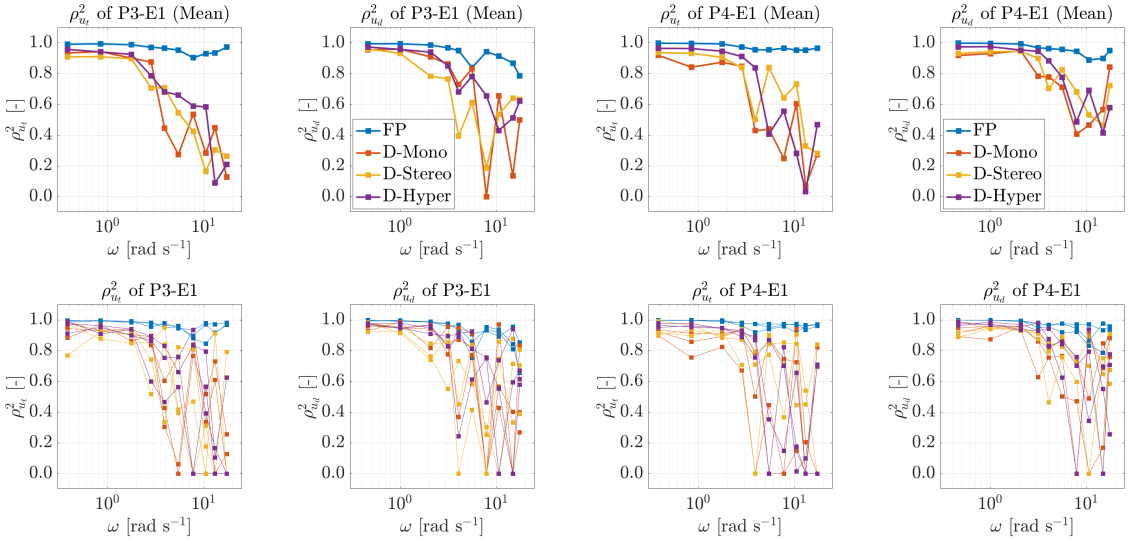


Figure B.3: Correlation *P3*

Figure B.4: Correlation *P4*

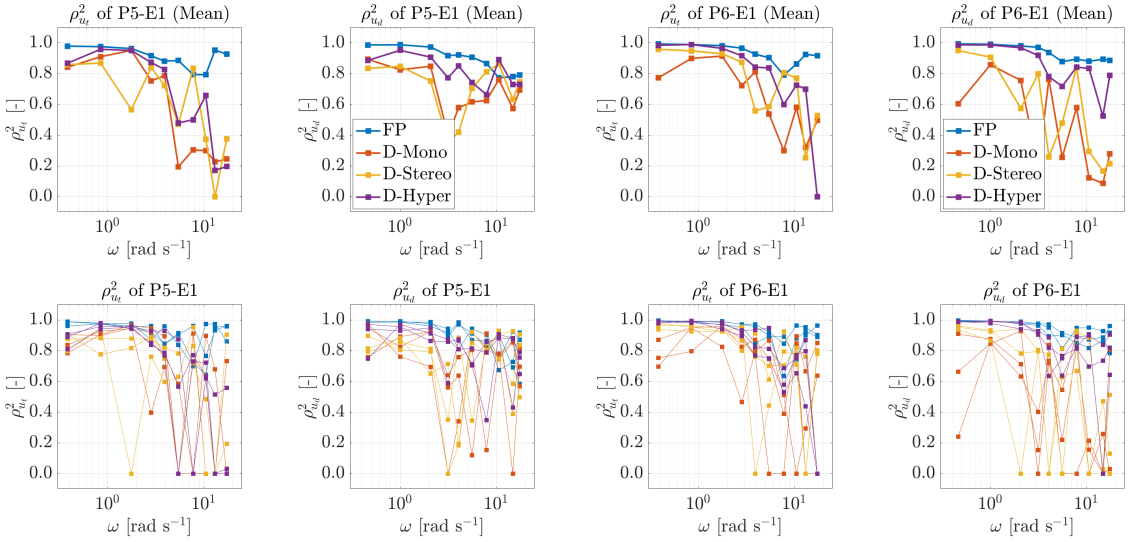


Figure B.5: Correlation *P5*

Figure B.6: Correlation *P6*

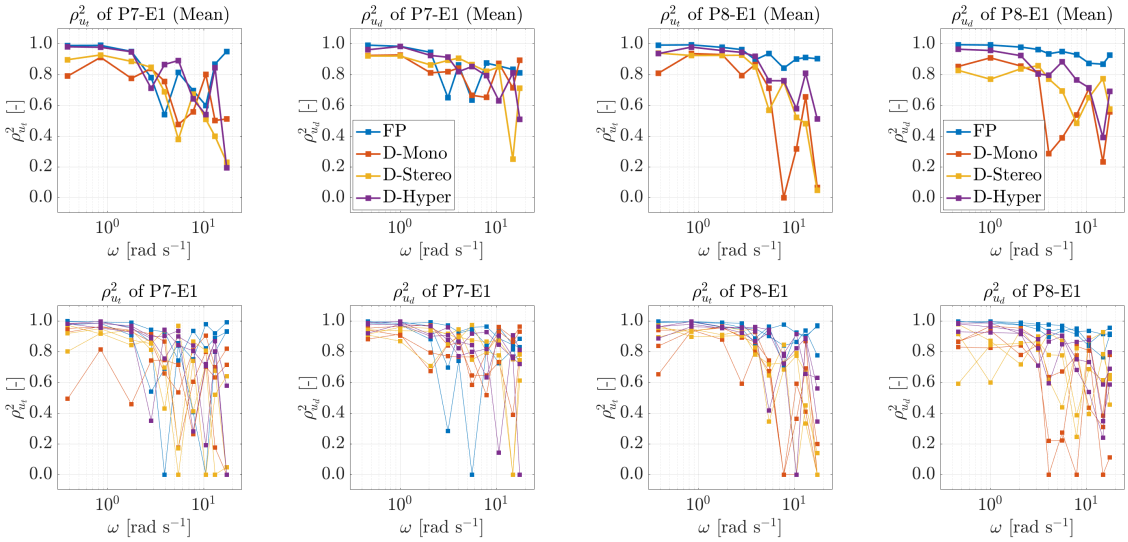


Figure B.7: Correlation *P7*

Figure B.8: Correlation *P8*

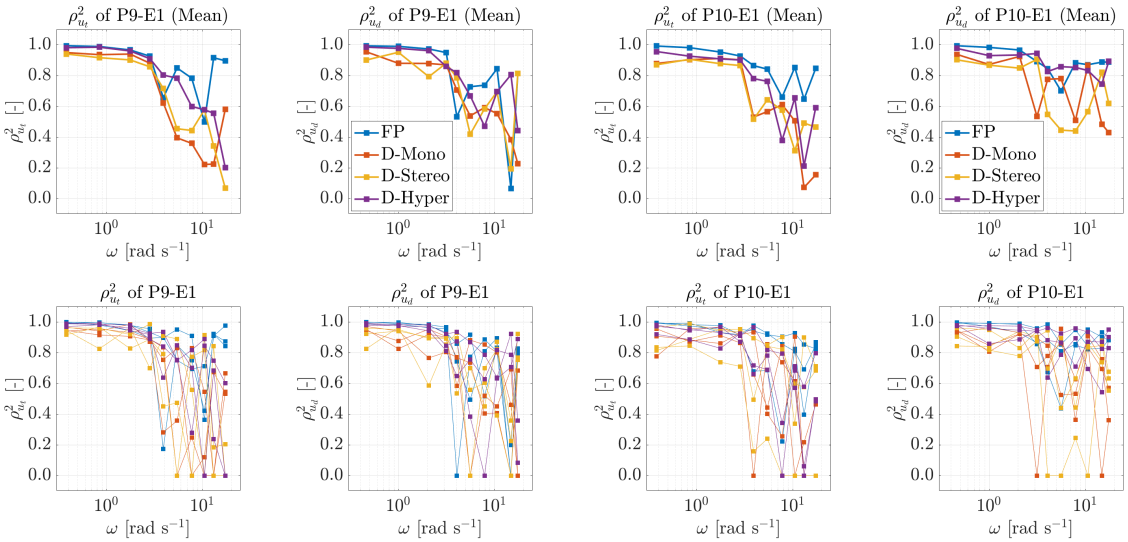


Figure B.9: Correlation *P9*

Figure B.10: Correlation *P10*

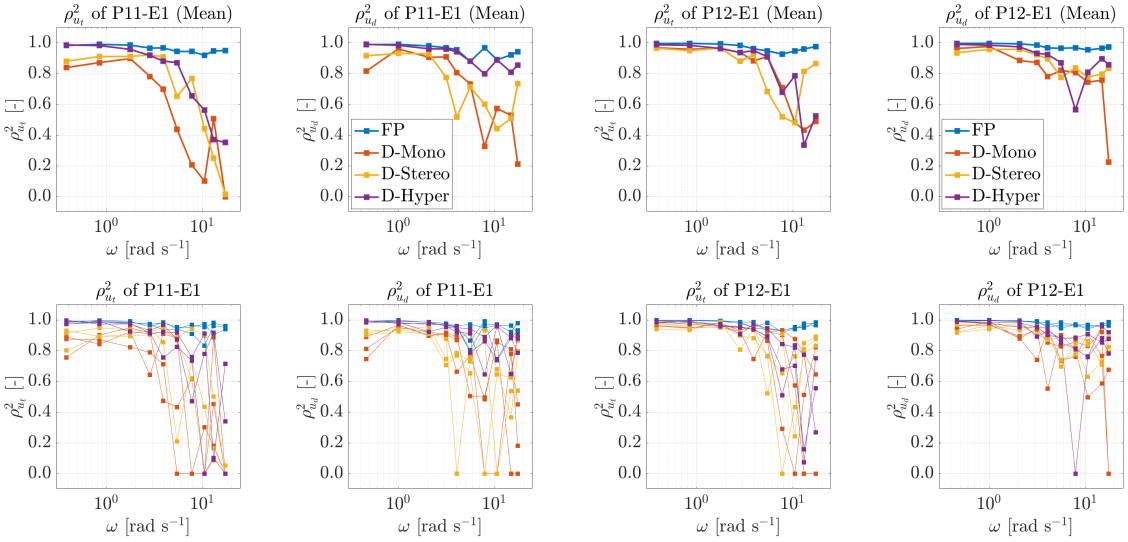


Figure B.11: Correlation *P11*

Figure B.12: Correlation *P12*

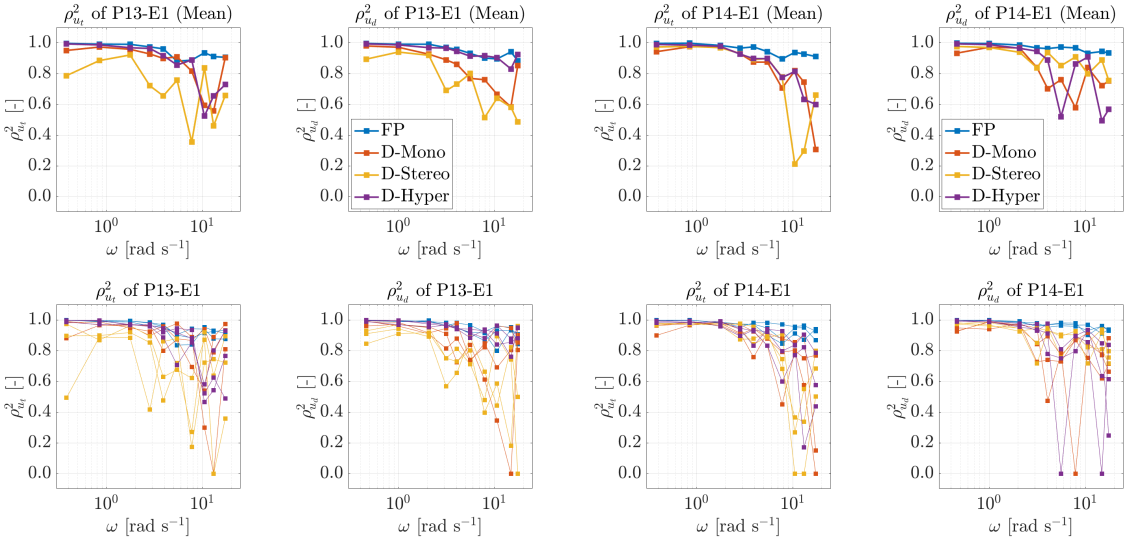


Figure B.13: Correlation *P13*

Figure B.14: Correlation *P14*

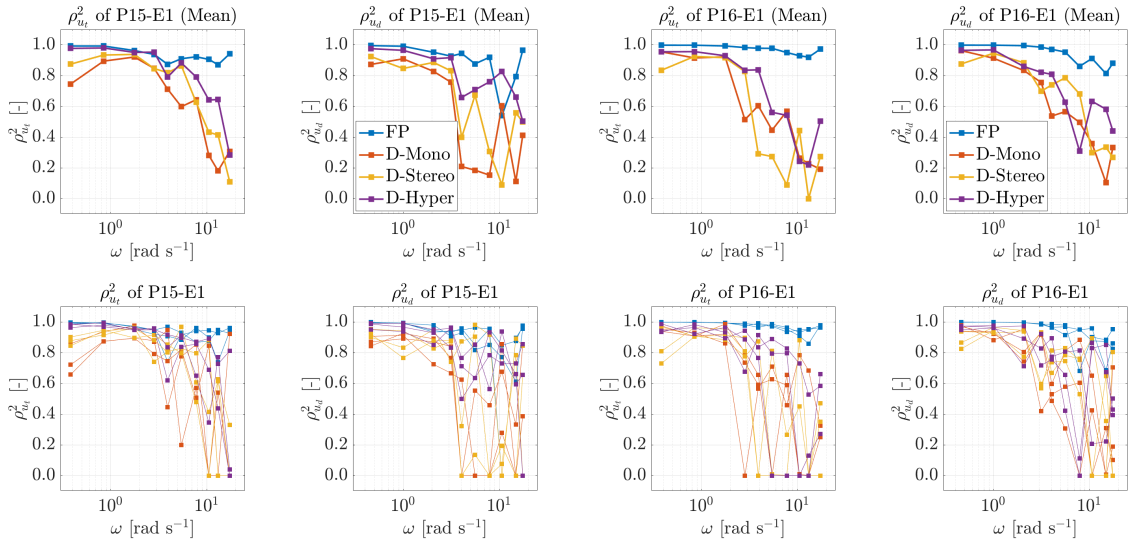


Figure B.15: Correlation *P15*

Figure B.16: Correlation *P16*

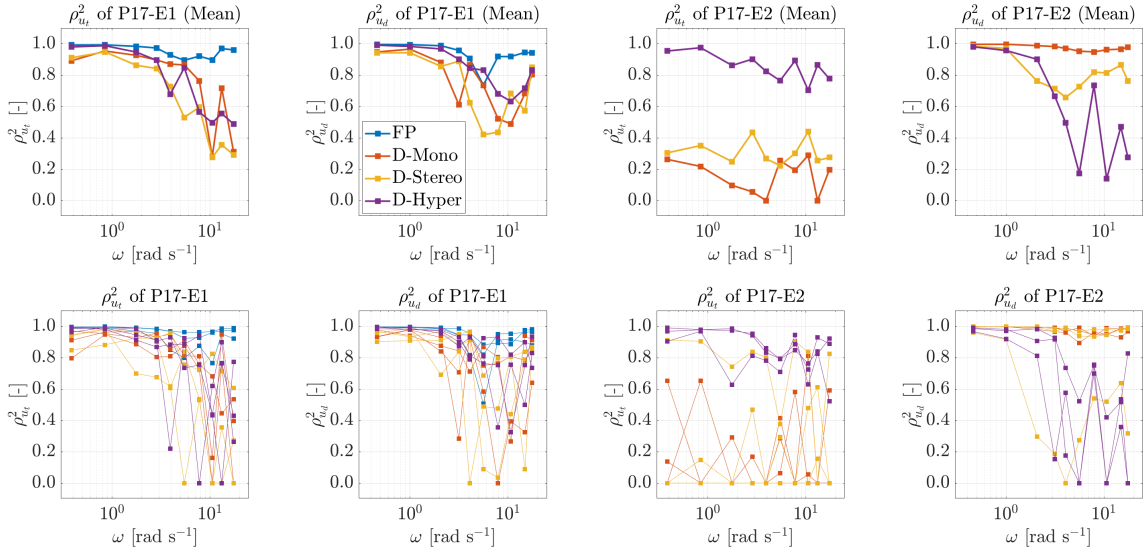


Figure B.17: Correlation *P17*

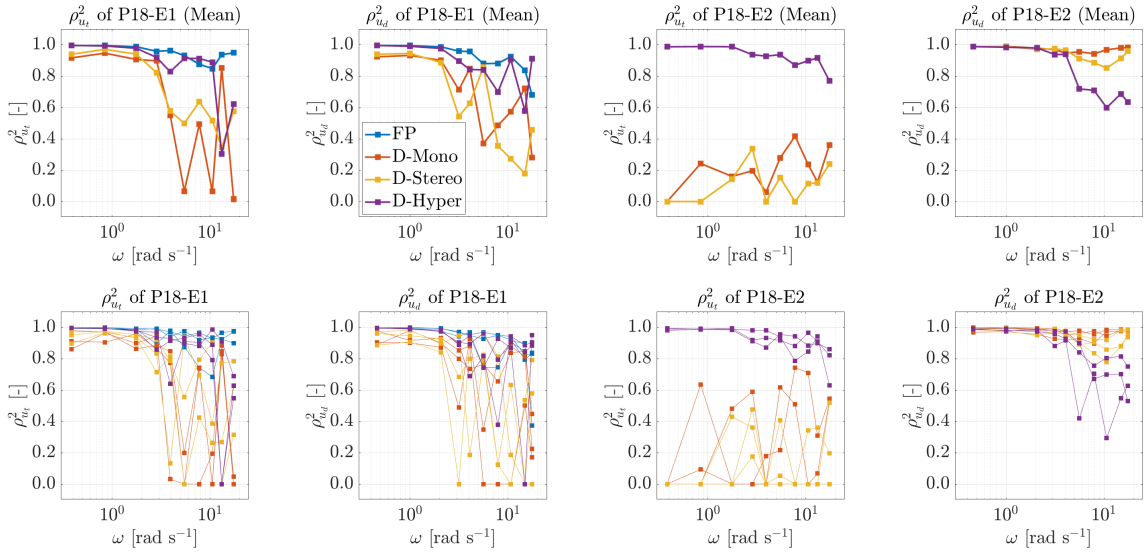


Figure B.18: Correlation *P18*

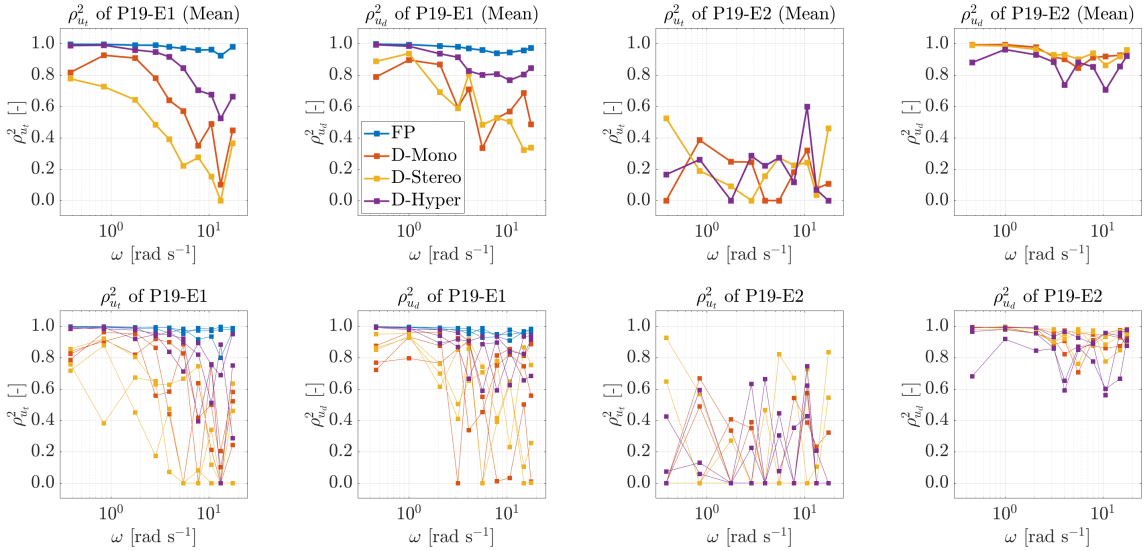


Figure B.19: Correlation *P19*

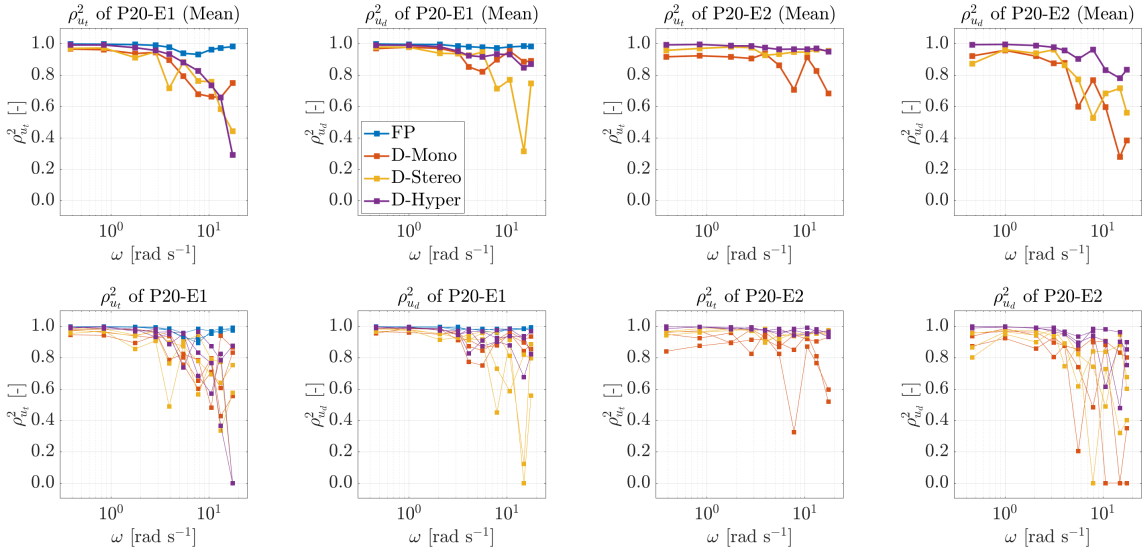


Figure B.20: Correlation *P20*

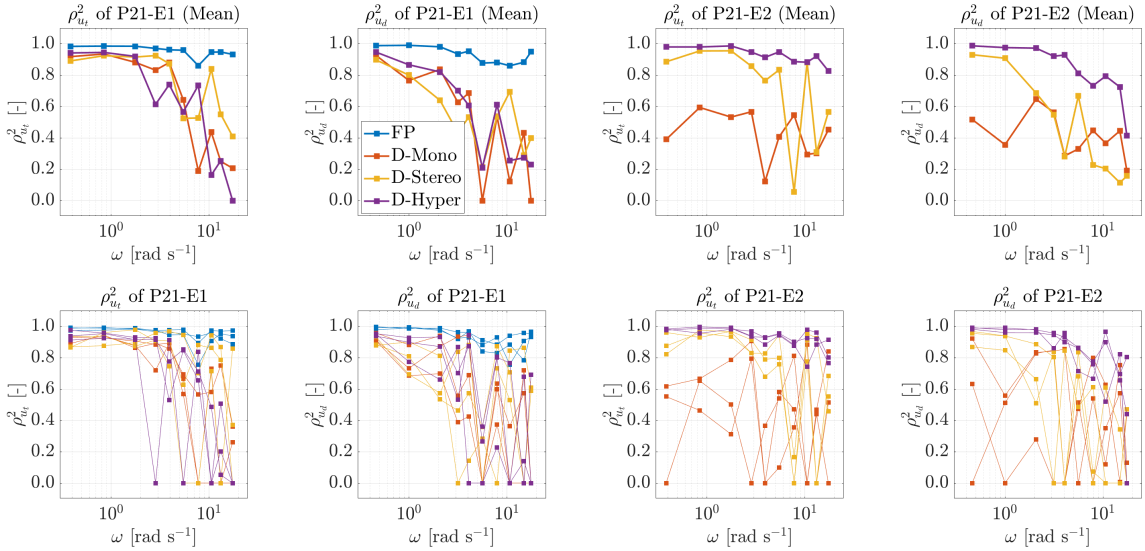


Figure B.21: Correlation *P21*

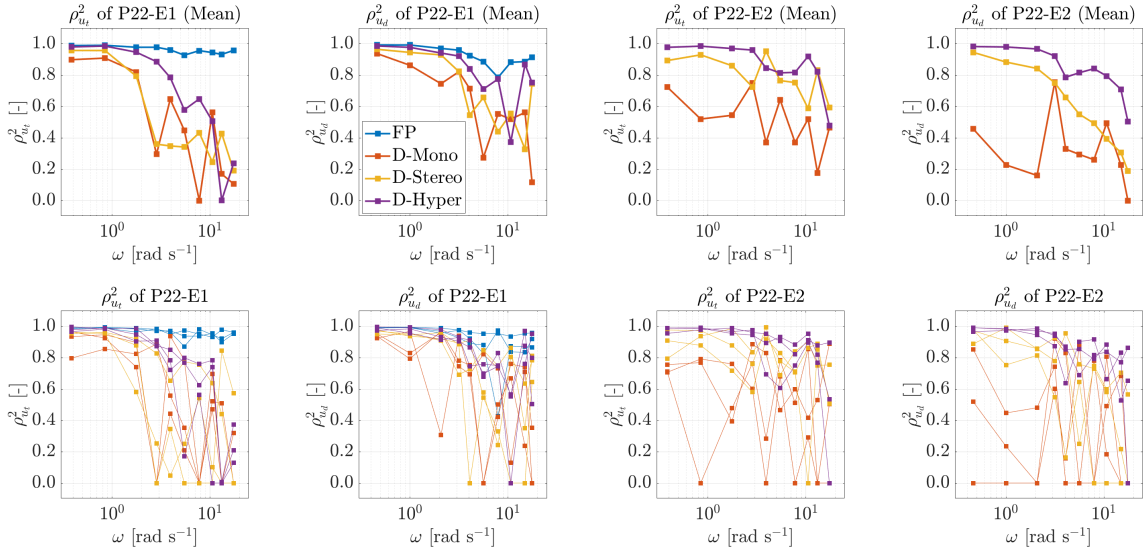


Figure B.22: Correlation *P22*

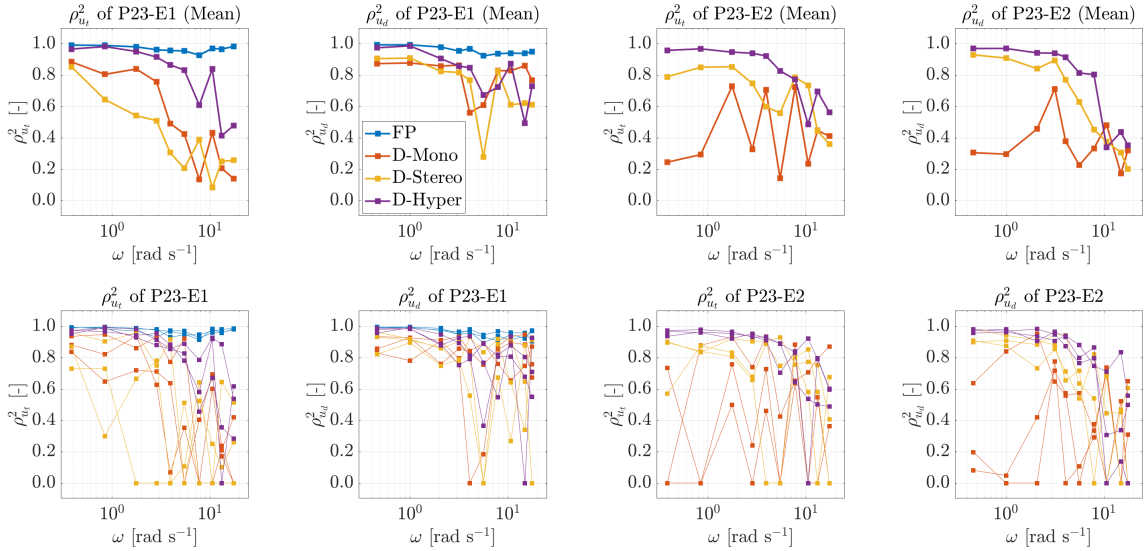


Figure B.23: Correlation *P23*

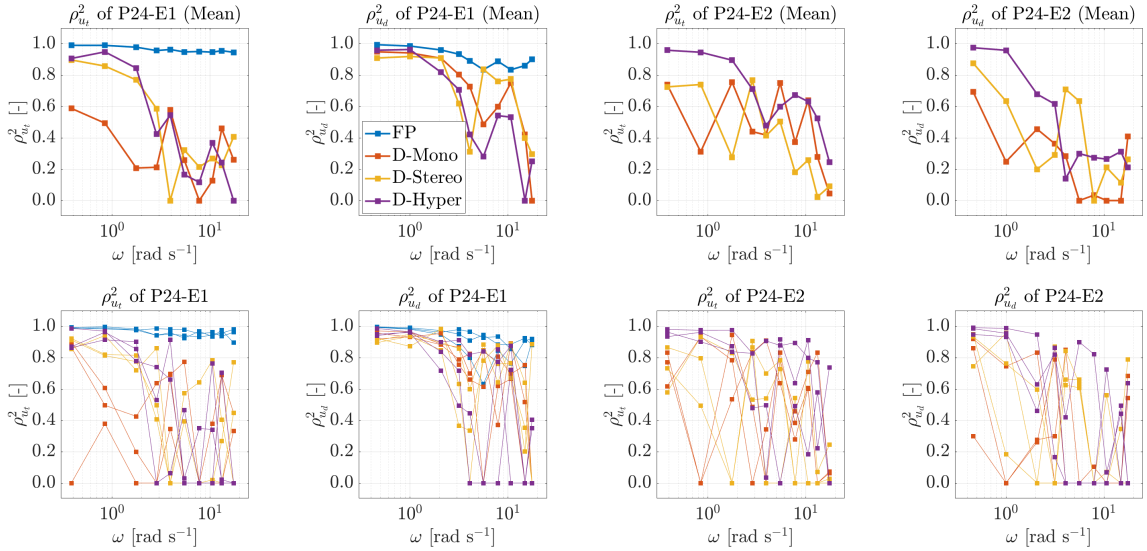


Figure B.24: Correlation *P24*

C

Error Consistency

The error consistency analysis plots (explained in the Methods section in Part I) are presented in this appendix for each individual participant and display condition. Each plot contains the (e_{RMS}) per section of each measurement run, i.e., three values for (e_{RMS}) per section.

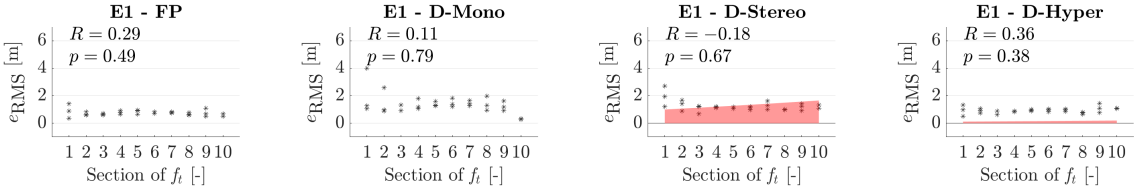


Figure C.1: Error consistency of *P1*

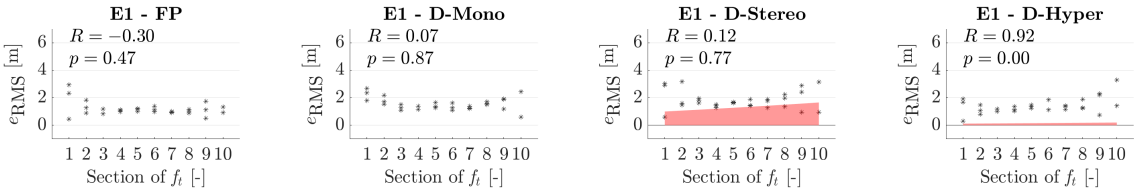


Figure C.2: Error consistency of *P2*

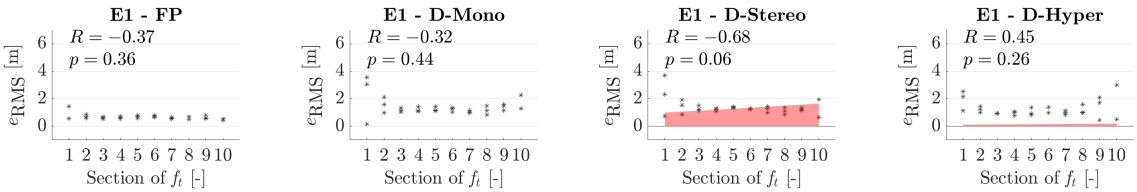


Figure C.3: Error consistency of *P3*

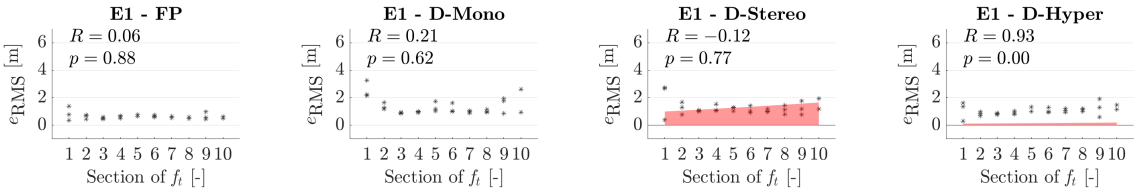


Figure C.4: Error consistency of *P4*

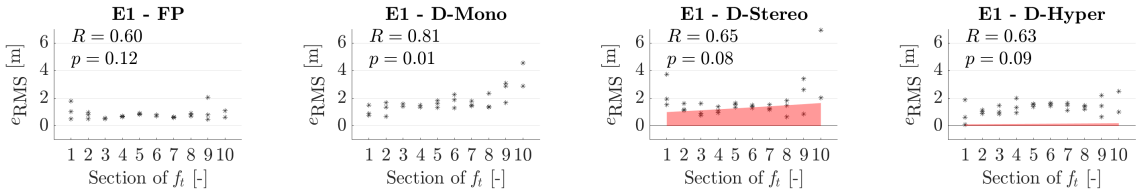


Figure C.5: Error consistency of *P5*

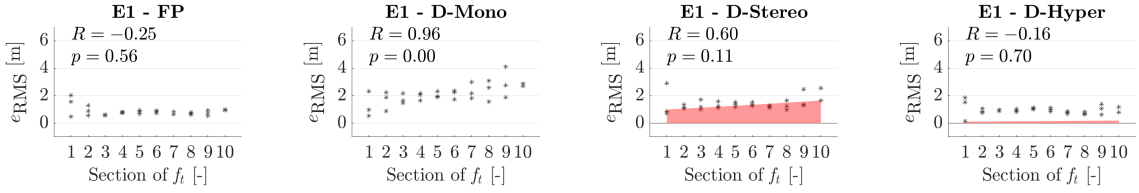


Figure C.6: Error consistency of *P6*

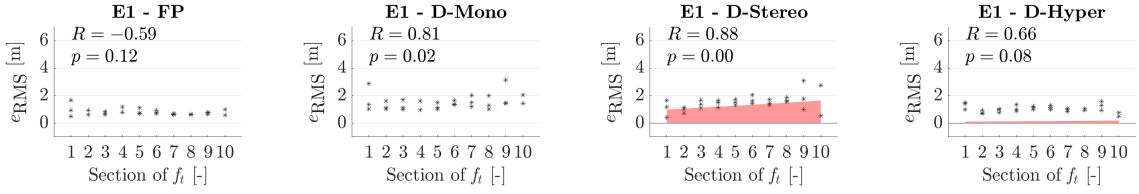


Figure C.7: Error consistency of *P7*

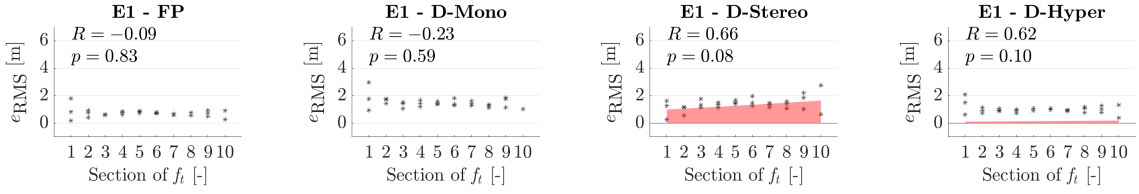


Figure C.8: Error consistency of *P8*

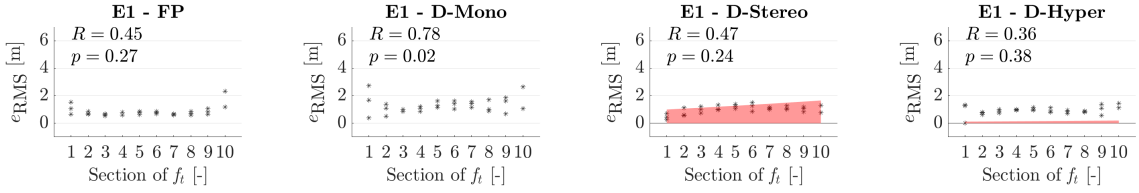


Figure C.9: Error consistency of *P9*

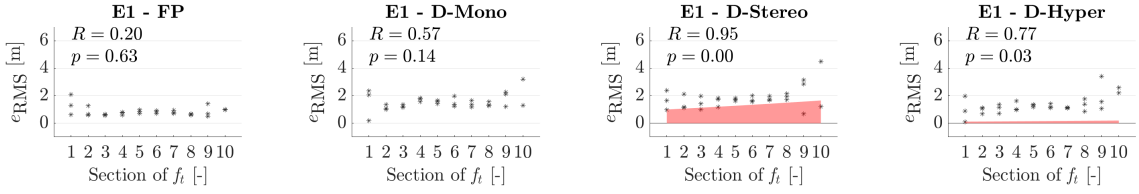


Figure C.10: Error consistency of *P10*

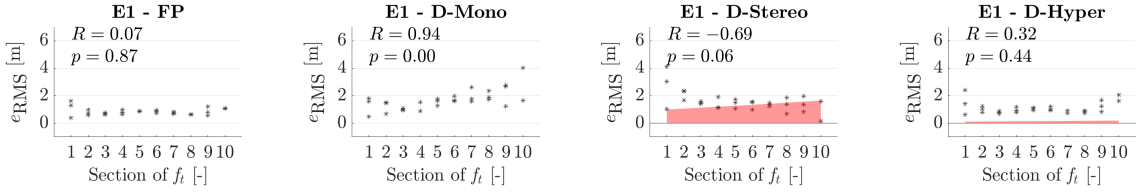


Figure C.11: Error consistency of *P11*

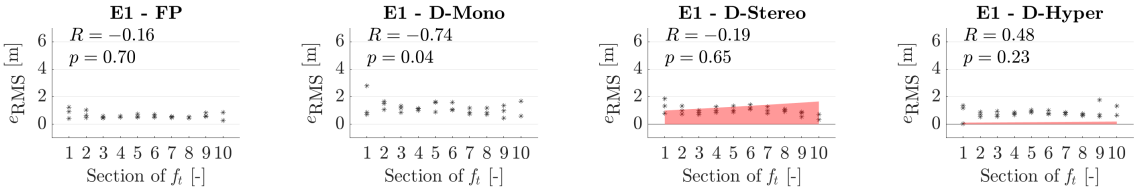


Figure C.12: Error consistency of *P12*

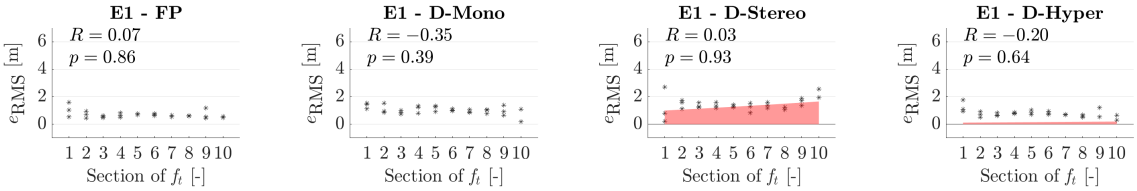


Figure C.13: Error consistency of *P13*

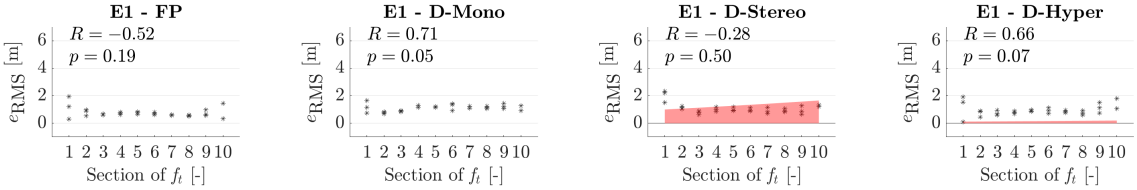


Figure C.14: Error consistency of *P14*

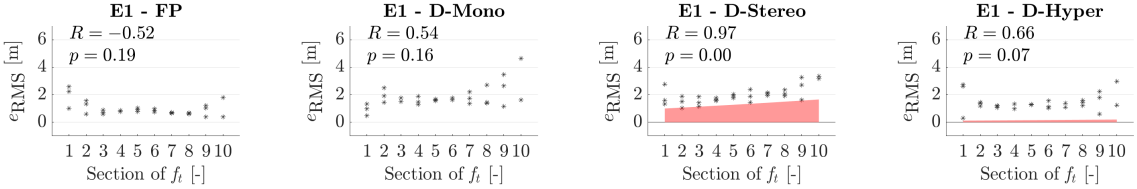


Figure C.15: Error consistency of *P15*

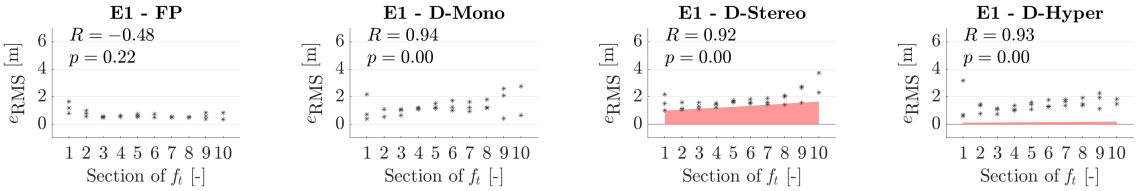


Figure C.16: Error consistency of *P16*

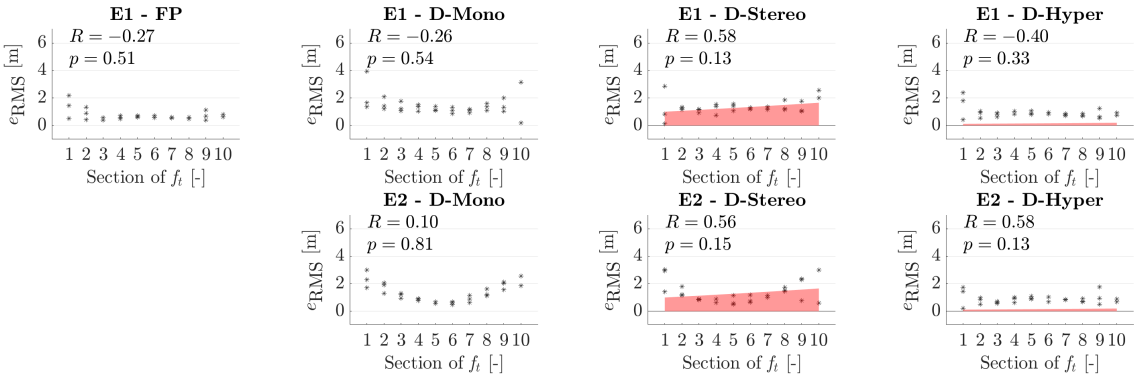


Figure C.17: Error consistency of *P17*

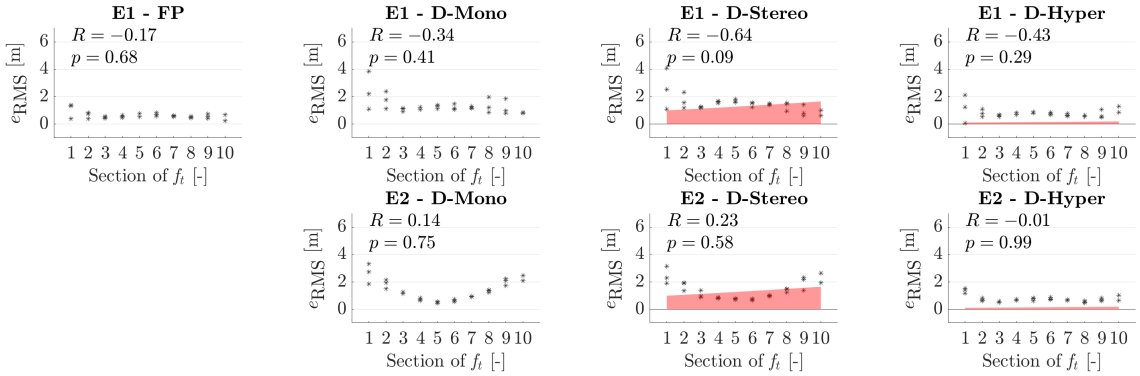


Figure C.18: Error consistency of *P18*

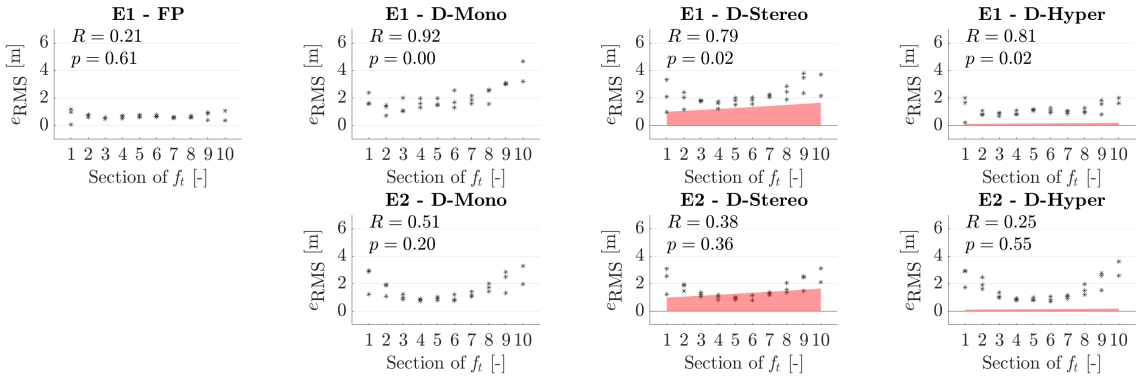


Figure C.19: Error consistency of *P19*

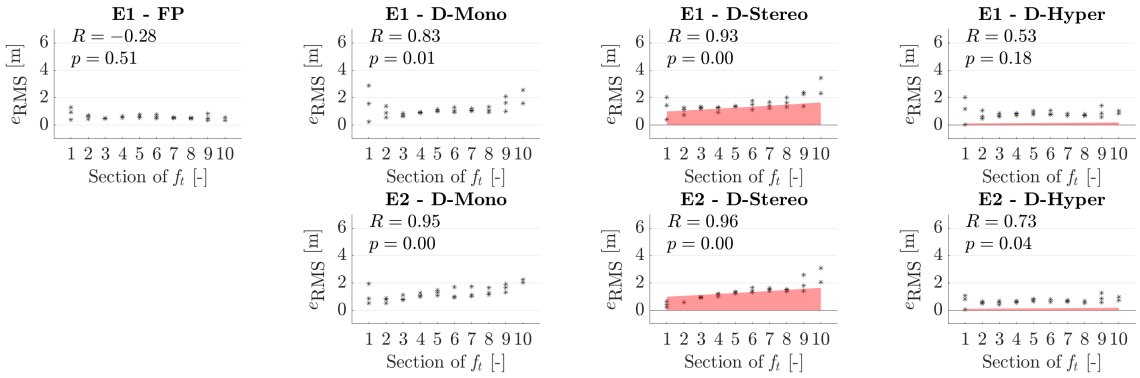


Figure C.20: Error consistency of *P20*

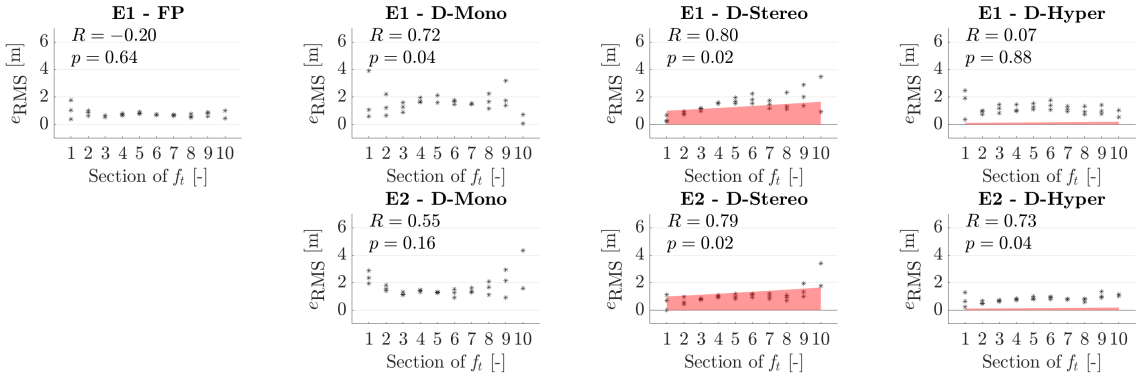


Figure C.21: Error consistency of *P21*

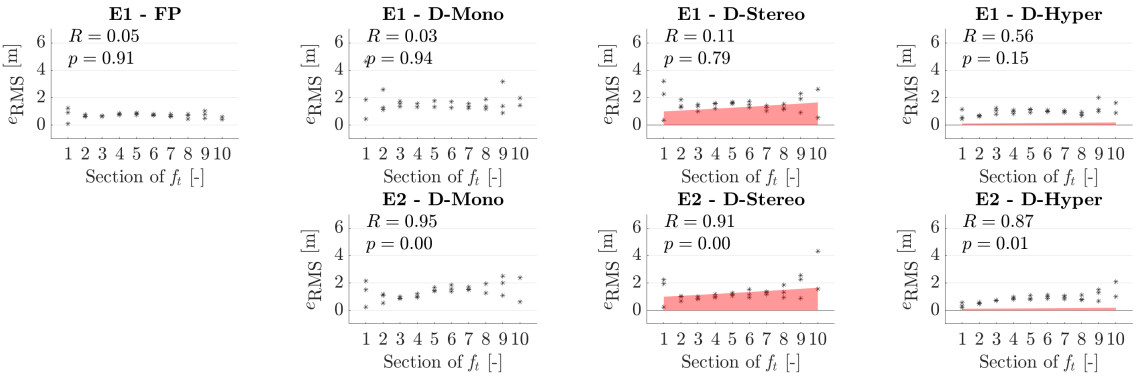


Figure C.22: Error consistency of *P22*

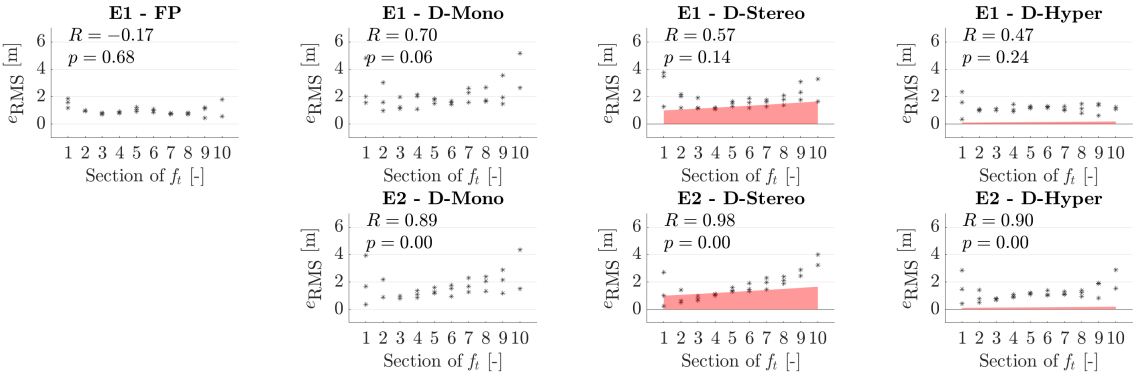


Figure C.23: Error consistency of *P23*

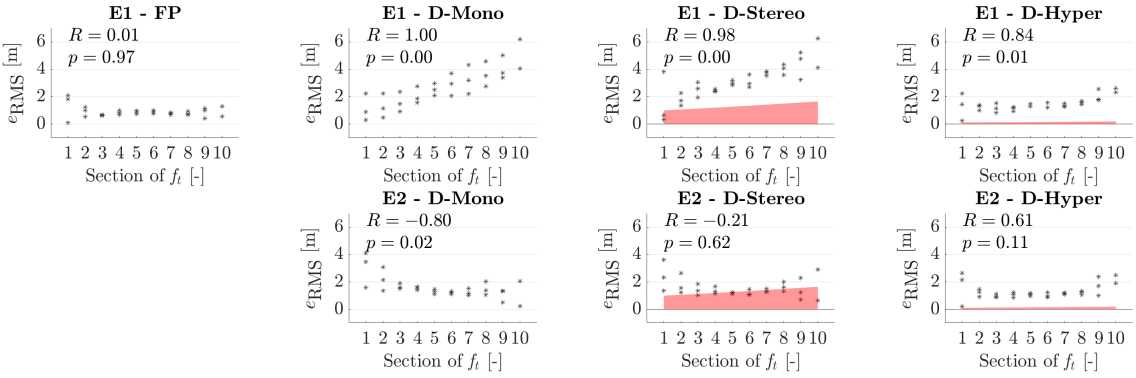


Figure C.24: Error consistency of *P24*

D

HUMAN CONTROLLER MODELING

This appendix presents the human pilot describing functions for each individual participant and display condition as a supplement to the example describing function in Part I. Each plot indicates the quality-of-fit by R^2 for the estimated human controller's model parameters to the mean of the identified FRF of three measurements.

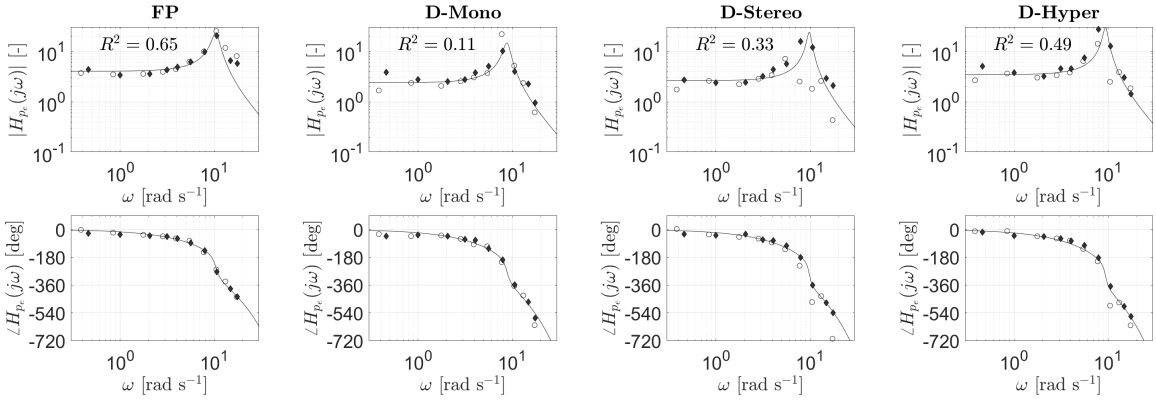


Figure D.1: HC FRFs and estimated model of *P1*

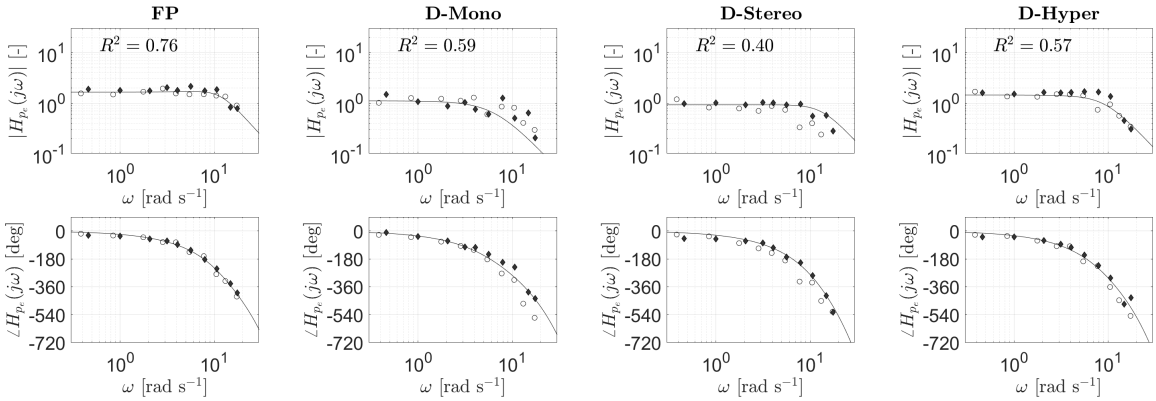


Figure D.2: HC FRFs and estimated model of *P2*

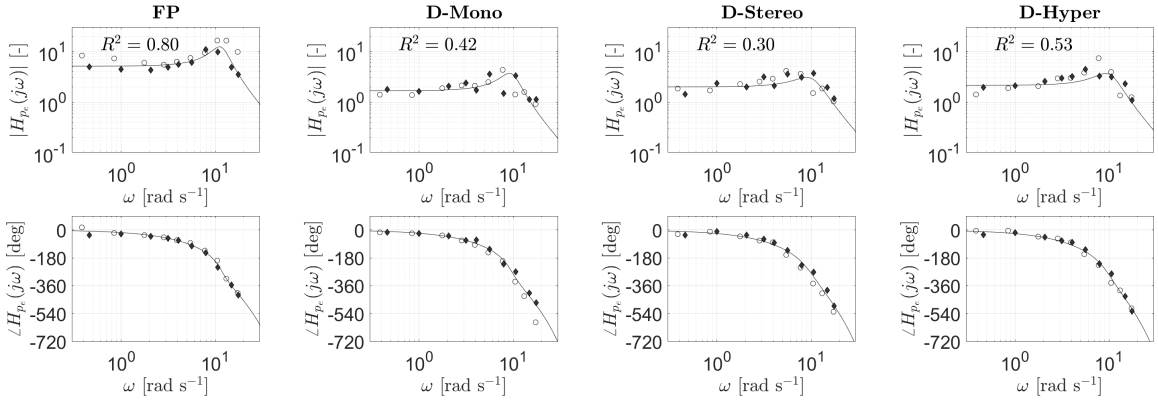


Figure D.3: HC FRFs and estimated model of *P3*

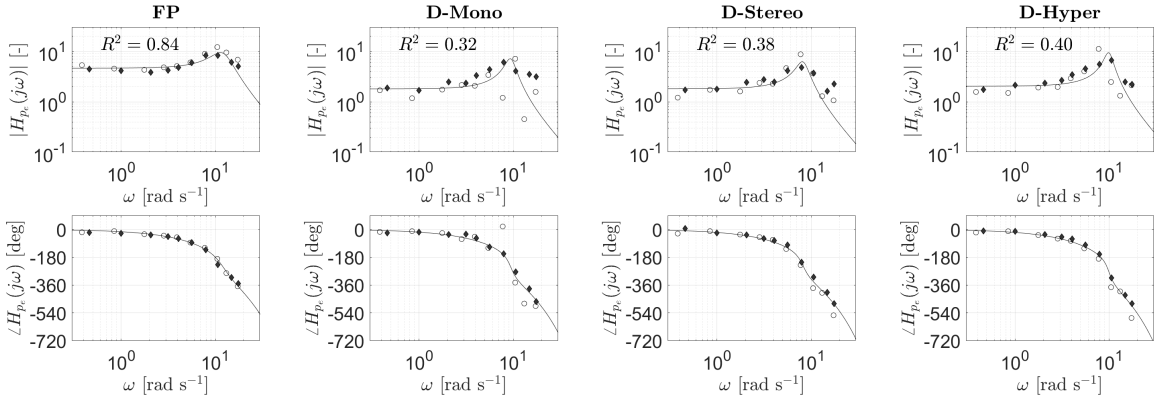


Figure D.4: HC FRFs and estimated model of *P4*

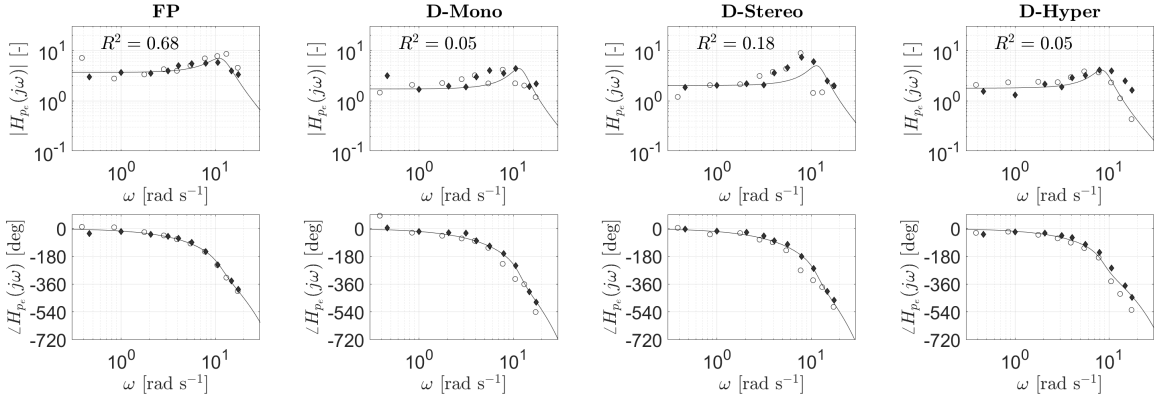


Figure D.5: HC FRFs and estimated model of *P5*

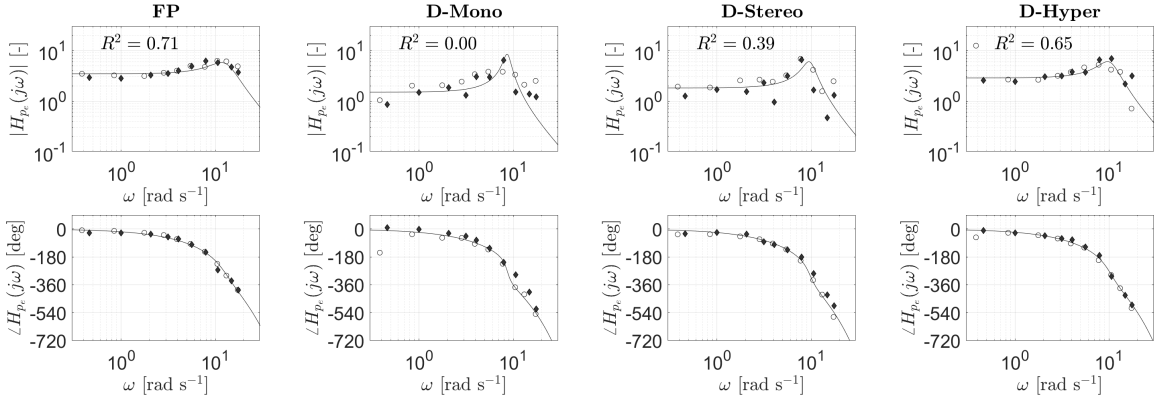


Figure D.6: HC FRFs and estimated model of *P6*

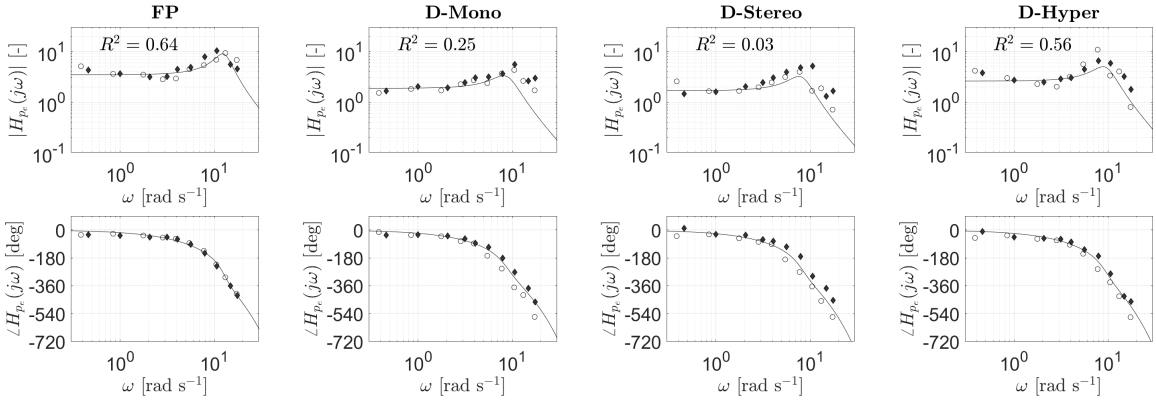


Figure D.7: HC FRFs and estimated model of *P7*

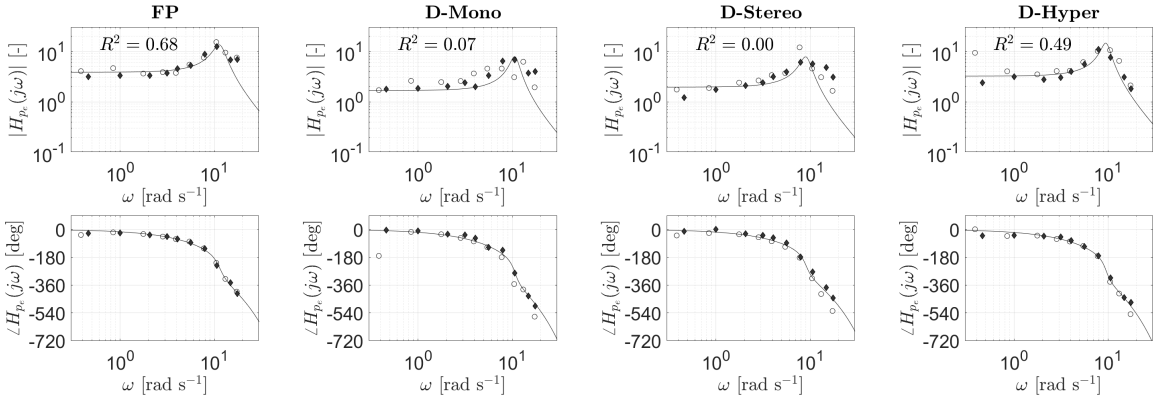


Figure D.8: HC FRFs and estimated model of *P8*

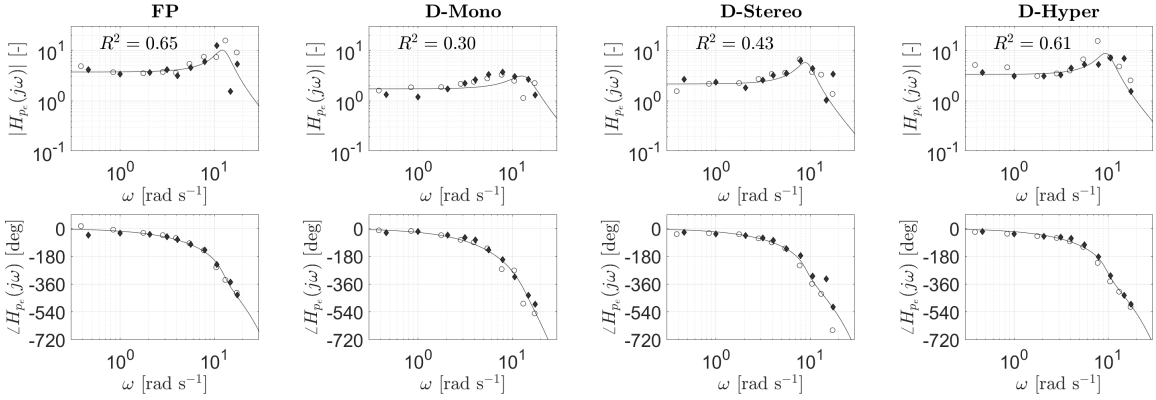


Figure D.9: HC FRFs and estimated model of *P9*

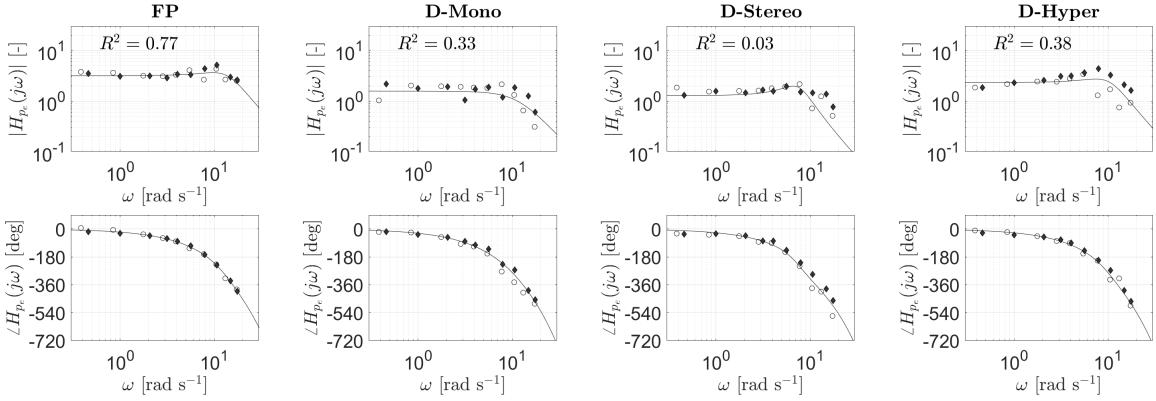


Figure D.10: HC FRFs and estimated model of *P10*

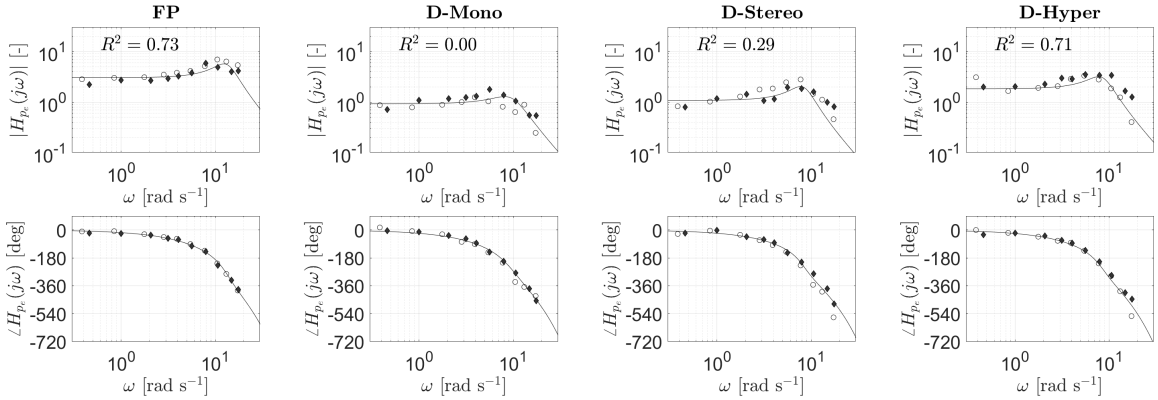


Figure D.11: HC FRFs and estimated model of *P11*

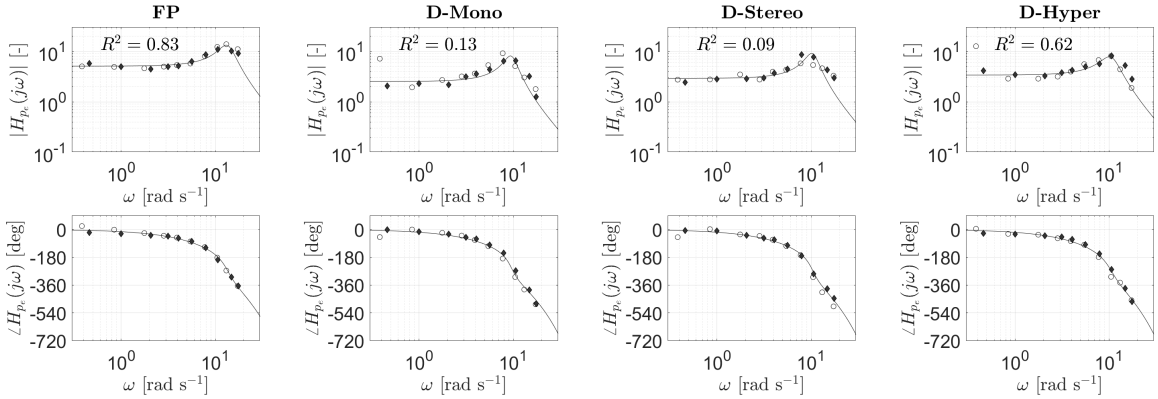


Figure D.12: HC FRFs and estimated model of *P12*

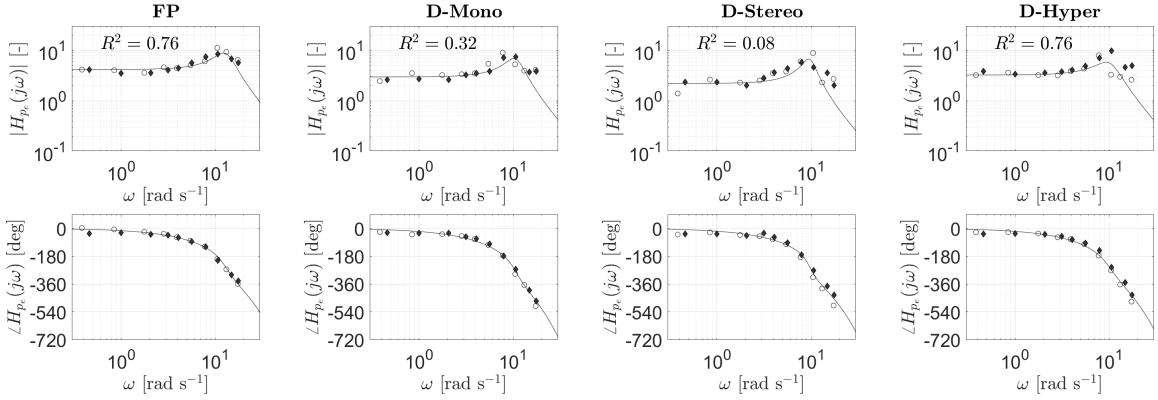


Figure D.13: HC FRFs and estimated model of *P13*

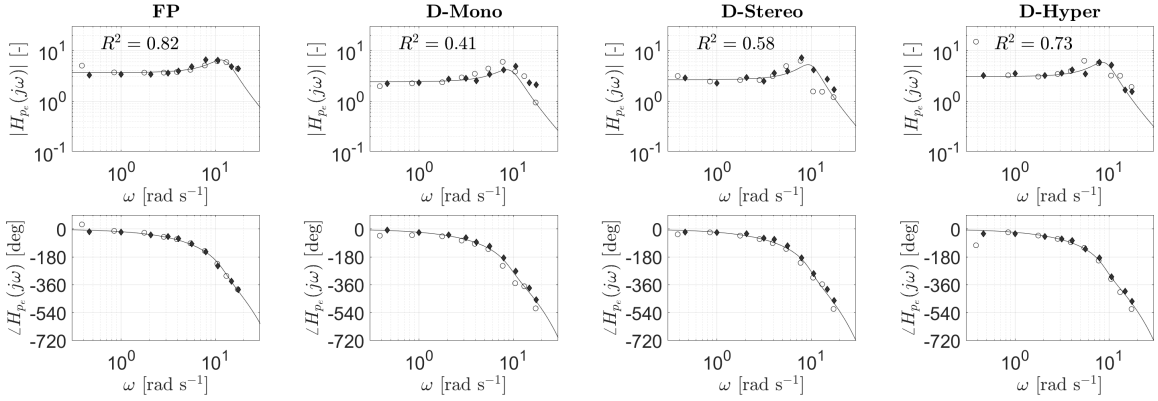


Figure D.14: HC FRFs and estimated model of *P14*

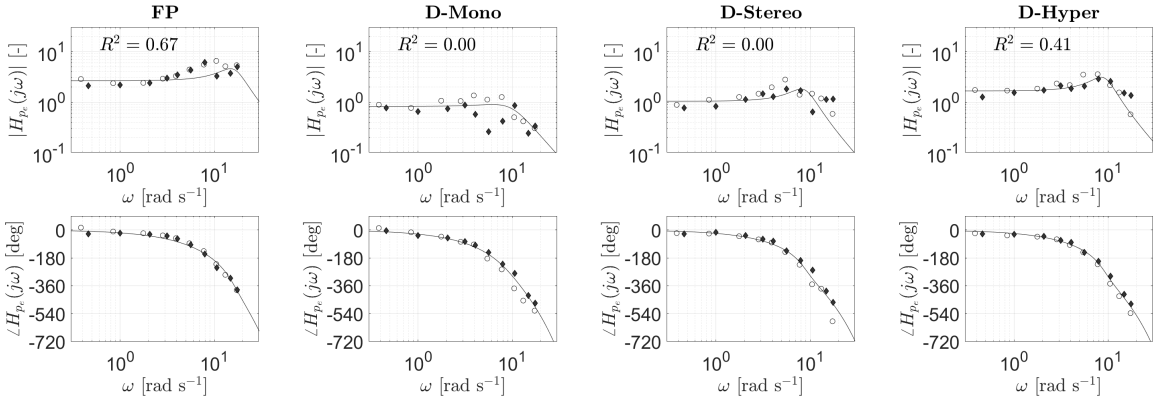


Figure D.15: HC FRFs and estimated model of *P15*

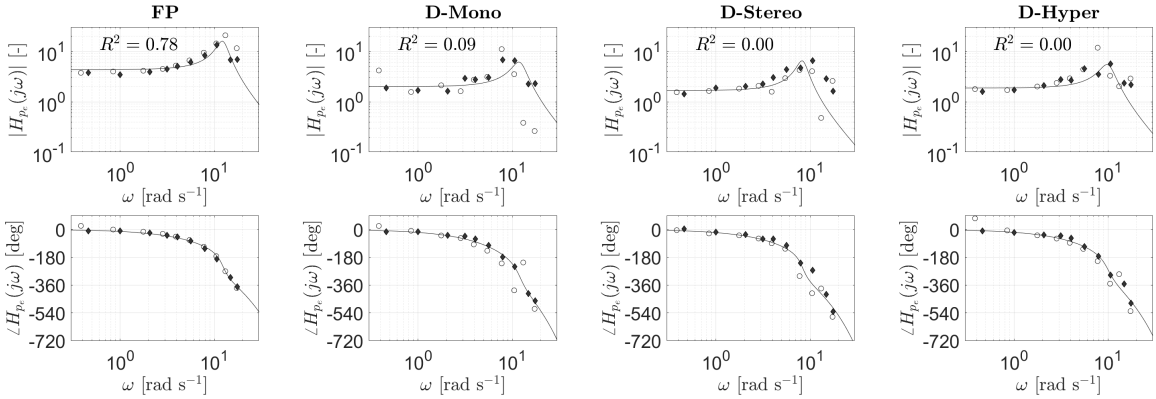


Figure D.16: HC FRFs and estimated model of *P16*

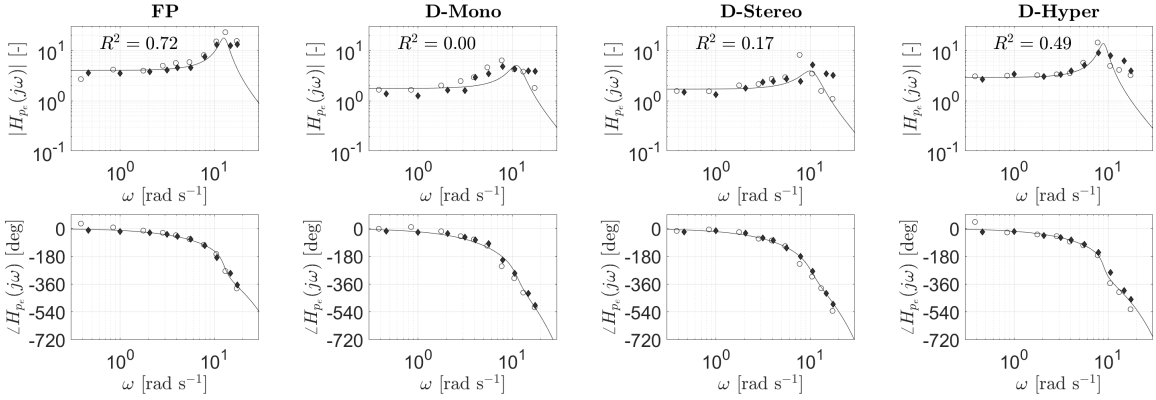


Figure D.17: HC FRFs and estimated model of *P17*

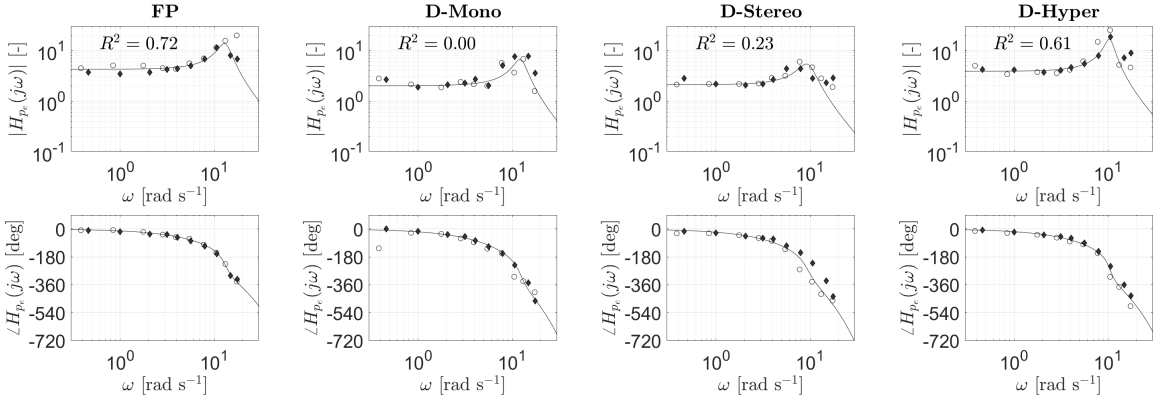


Figure D.18: HC FRFs and estimated model of *P18*

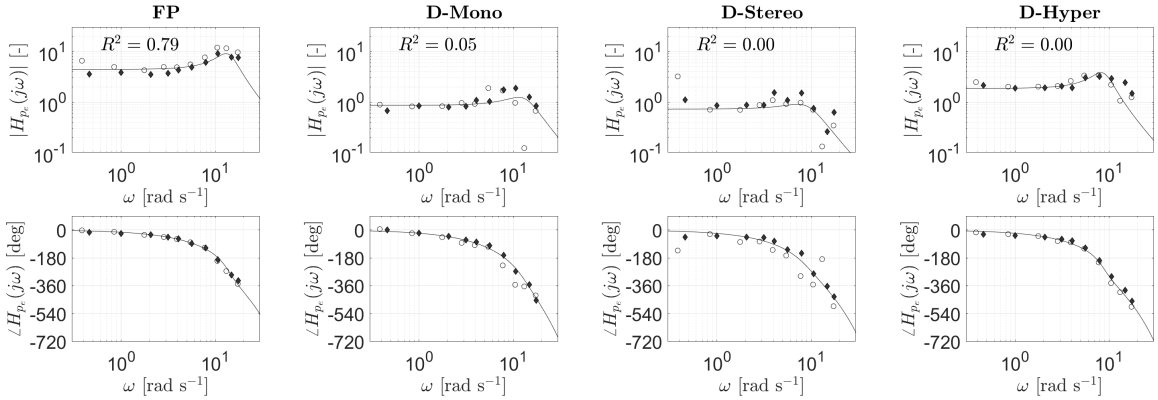


Figure D.19: HC FRFs and estimated model of *P19*

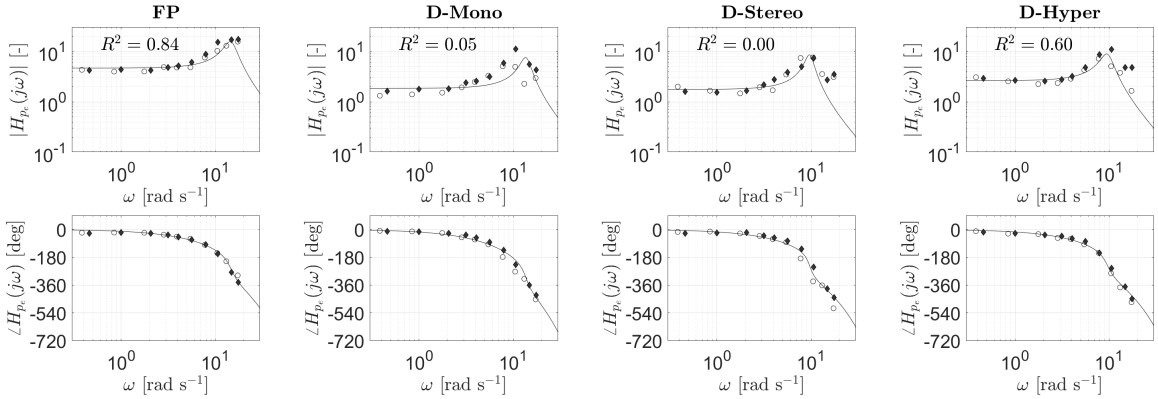


Figure D.20: HC FRFs and estimated model of *P20*

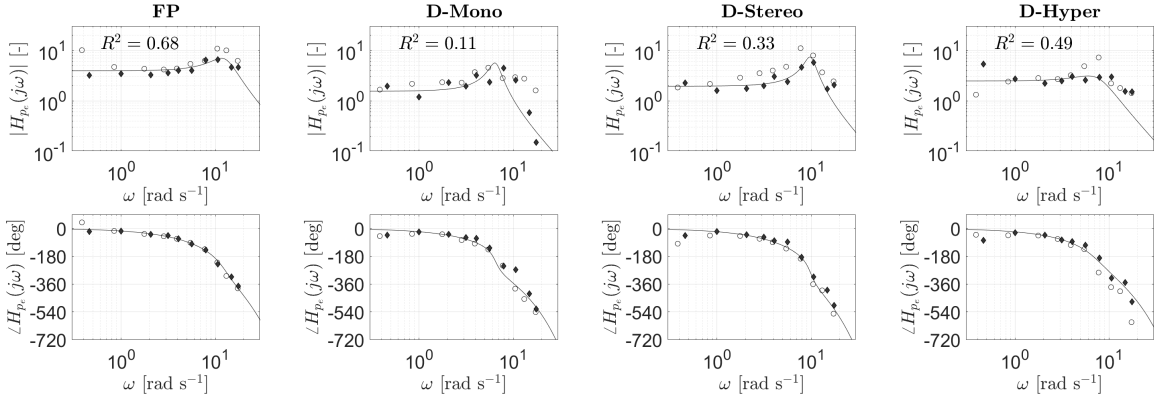


Figure D.21: HC FRFs and estimated model of *P21*

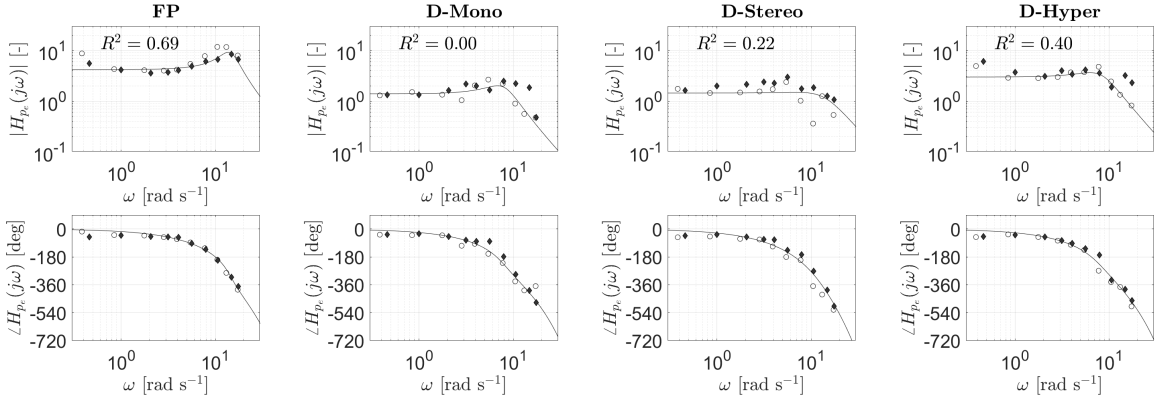


Figure D.22: HC FRFs and estimated model of *P22*

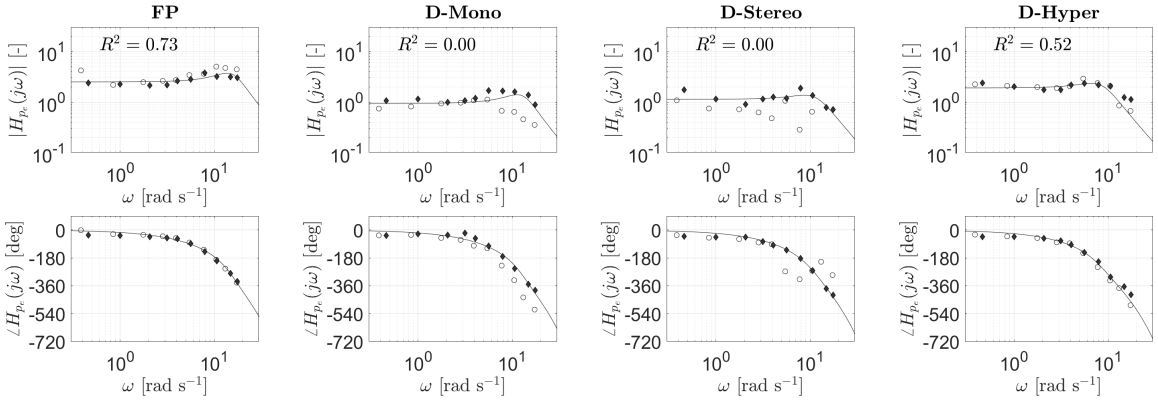


Figure D.23: HC FRFs and estimated model of *P23*

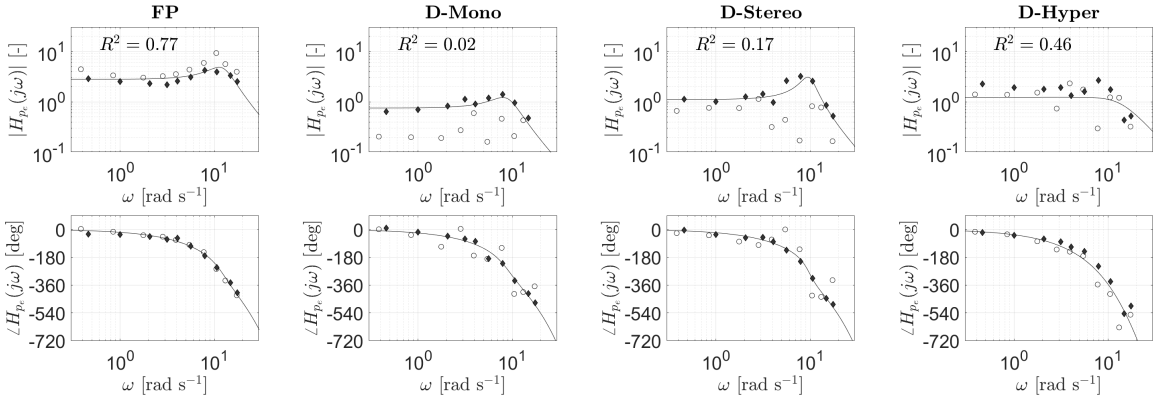


Figure D.24: HC FRFs and estimated model of *P24*

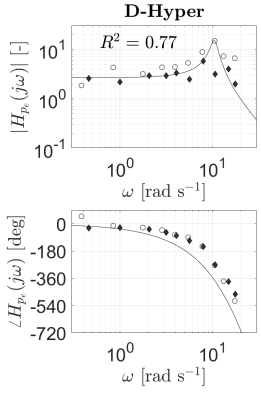


Figure D.25: *P17* in E2

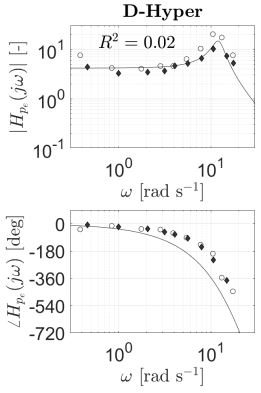


Figure D.26: *P18* in E2

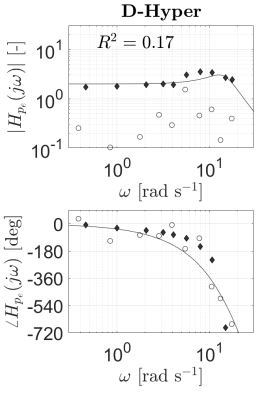


Figure D.27: *P19* in E2

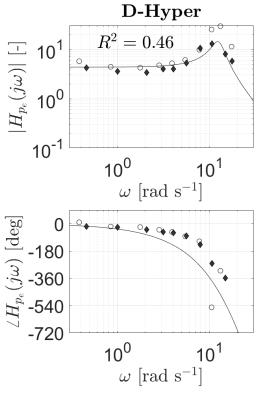


Figure D.28: *P20* in E2

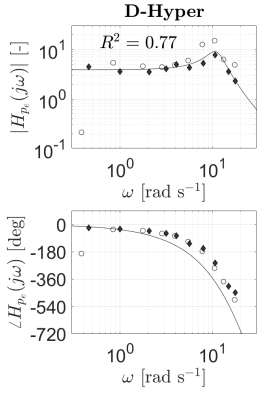


Figure D.29: *P21* in E2

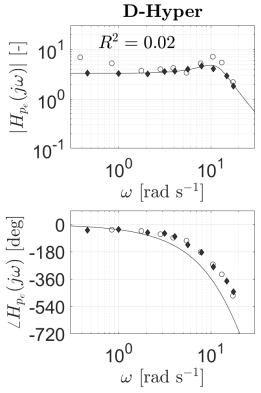


Figure D.30: *P22* in E2

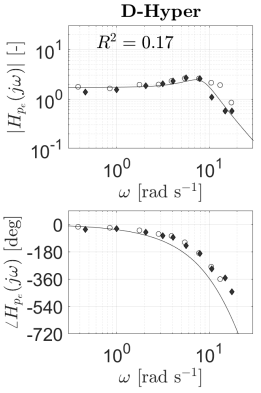


Figure D.31: *P23* in E2

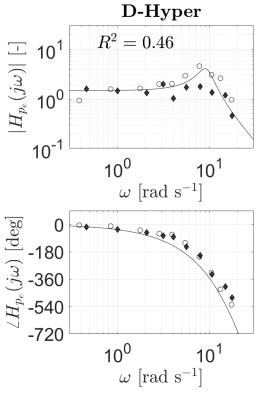


Figure D.32: *P24* in E2

E

STATISTICAL ANALYSES

The Shapiro-Wilk normality test was done to assure justified results from repeated-measures two-way ANOVA tests carried out in Part I. The results are presented hereafter in Table E.1. Display conditions not included in the realistic environment (*FP*) or measures not calculated for certain display conditions (K_p, τ_p, ω_{nm} and ζ_{nm}) appear as a '–' in the table.

Table E.1: Shapiro-Wilk normality test results

		Participant Group 1, Abstract Environment			Participant Group 2, Abstract Environment			Participant Group 2, Realistic Environment		
		Statistic	df	Sig.	Statistic	df	Sig.	Statistic	df	Sig.
σ_e^2	FP	0.951	16	0.51	0.900	8	0.29	–	–	–
	D-Mono	0.933	16	0.27	0.867	8	0.14	0.843	8	0.08
	D-Stereo	0.953	16	0.54	0.924	8	0.46	0.890	8	0.24
	D-Hyper	0.963	16	0.72	0.952	8	0.73	0.965	8	0.86
σ_u^2	FP	0.908	16	0.11	0.986	8	0.99	–	–	–
	D-Mono	0.920	16	0.17	0.951	8	0.72	0.850	8	0.09
	D-Stereo	0.906	16	0.10	0.640	8	0.00	0.940	8	0.61
	D-Hyper	0.967	16	0.79	0.885	8	0.21	0.890	8	0.23
K_p	FP	0.964	16	0.74	0.840	8	0.08	–	–	–
	D-Mono	0.962	16	0.69	0.914	8	0.38	–	–	–
	D-Stereo	0.960	16	0.67	0.959	8	0.80	–	–	–
	D-Hyper	0.901	16	0.08	0.978	8	0.96	0.910	8	0.35
τ_p	FP	0.922	16	0.18	0.968	8	0.88	–	–	–
	D-Mono	0.884	16	0.05	0.964	8	0.84	–	–	–
	D-Stereo	0.973	16	0.88	0.983	8	0.98	–	–	–
	D-Hyper	0.963	16	0.72	0.683	8	0.00	0.935	8	0.56
ω_{nm}	FP	0.905	16	0.10	0.918	8	0.41	–	–	–
	D-Mono	0.957	16	0.61	0.886	8	0.21	–	–	–
	D-Stereo	0.901	16	0.08	0.774	8	0.01	–	–	–
	D-Hyper	0.961	16	0.67	0.816	8	0.04	0.984	8	0.98
ζ_{nm}	FP	0.864	16	0.02	0.949	8	0.70	–	–	–
	D-Mono	0.807	16	0.00	0.800	8	0.03	–	–	–
	D-Stereo	0.809	16	0.00	0.873	8	0.16	–	–	–
	D-Hyper	0.806	16	0.00	0.897	8	0.27	0.882	8	0.20

F

Experiment Procedures

This appendix includes documents Briefing and Consent Form used to inform the participant prior to the experiment and register a written consent. All filled-in consent forms are archived by the researcher.

The experiment was approved by the secretary of the TU Delft Human Research Ethics Committee on August 29, 2019 as application number 862. The experiments were conducted between October 14 and October 18 2019 for the student participants at the Faculty of Aerospace Engineering in Delft and between October 28 and November 14, 2019 for the boom operator participants at the Royal Netherlands Airforce base Eindhoven.

Experiment Briefing

Manual control behaviour in stereoscopic vision-enhanced depth control tasks

Thank you for your contribution to this research. You will be participating in a manual control tracking experiment in a Tanker Remote Vision System (TRVS) Air-to-Air Refuelling (AAR) simulator at TU Delft (see Figure 2). The research focuses on the effects of controlling an element in the depth axis and the influence of having stereoscopic vision. This briefing will introduce you to the experiment and outline what is expected of you as a participant.

Experiment Goal

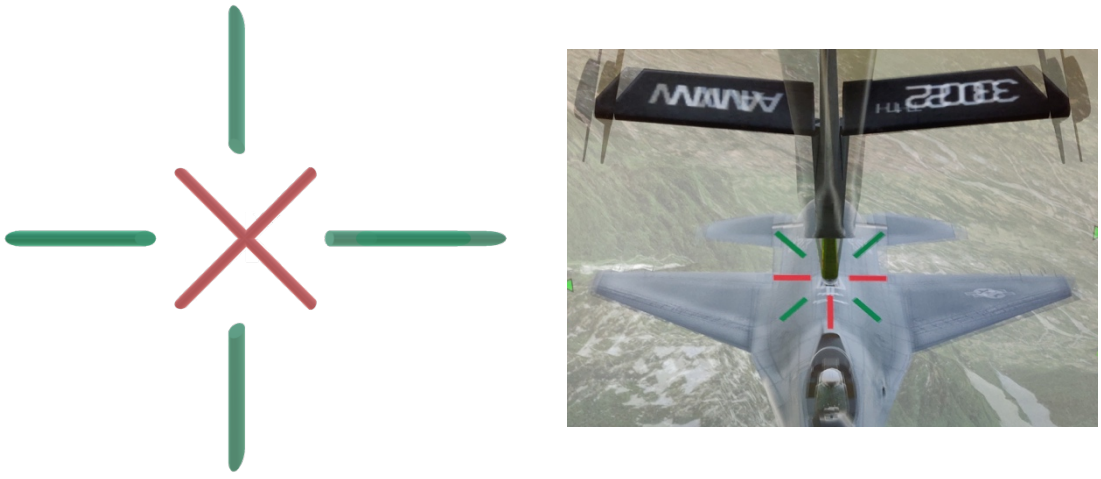
The goal of this experiment is to investigate the effects of stereoscopic vision on human control behaviour in manual depth control tasks. The need for perception of additional/different visual cues in a three-dimensional environment may entail considerable differences to classical control theory on human adaption and control behaviour. An understanding of such difference(s) may aid future implementation of stereoscopic 3D displays as a modern interface for control tasks heavily reliant on depth perception.

Experiment Task

The tasks you will carry out are single-axis depth control tracking tasks with pursuit display. Your goal is to actively minimize the position error in *two* target-following, disturbance rejection tasks:

- An abstract, non-realistic environment where the controlled element (green crosshairs) should be kept on the target element (red cross), minimising the error in position(see Figure 1(a));
- A realistic, simplified aerial refuelling tasks where the controlled element (KDC-10 refuelling boom) should be kept on the target element (F-16 aerial refuelling receptacle), minimising the error in position (see Figure 1(b)). Both elements (KDC-10's boom and F-16's receptacle) will be additionally visually represented by similar shapes as found in the abstract scenario.

The controlled elements in both tasks may be spatially manipulated further away from or closer to the observer by providing forward or aft inputs to a force feedback side-stick on the right-hand side of the seat, respectively (see Figure 2).



(a) **(b)**

Figure 1: A visual representation of the two control tasks. (a) An artificial control task where the green crosshair (controlled element) should be kept over the red cross (target element). (b) A realistic control task where the nozzle of the KDC-10 refuelling boom (controlled element) should be kept on the F-16 refuelling receptacle (target element).

Experiment Procedures

The experiment will go through the two tasks described in the section above for which various display settings are presented. All task/display combinations will be shown before actual measurements are done as a training for the different experiment conditions. After training phase, numerous trials per task/display condition will be completed. Upon completion of either a single trial or a full experiment condition, the researcher will clearly indicate the end of the trial/condition and announce the start of a new trial/condition.

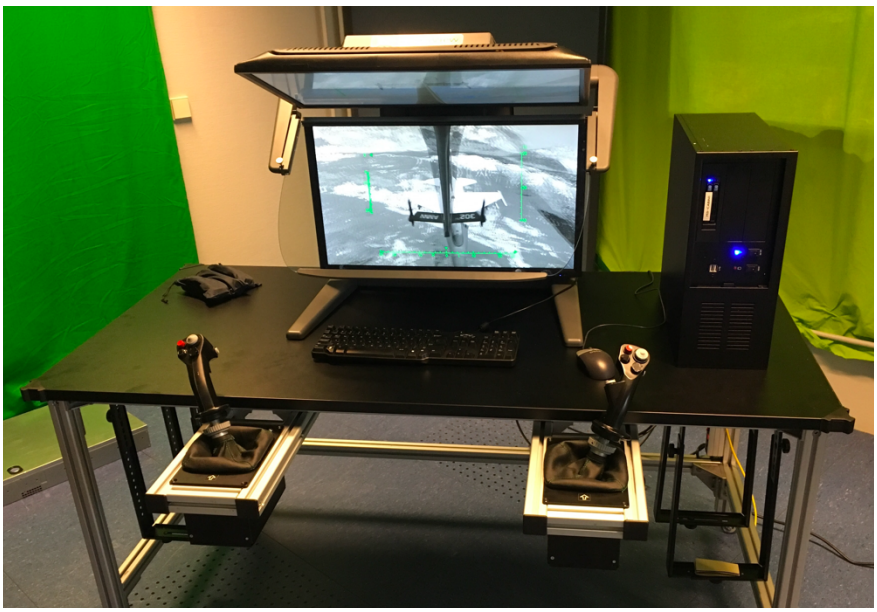


Figure 2: The TRVS AAR simulator loaded with the second task on aerial refuelling. During the experiments, the left side-stick will be removed to avoid confusion of input possibilities

Each trial lasts 90 seconds. Short breaks can be taken between trials to alleviate any discomfort that might occur due to controlling the side-stick, staring at a stereoscopic 3D display or after sitting in a fixed position for a prolonged period of time. Longer breaks will be taken after every 2-3 conditions completed. The complete experiment will last approximately 2.5 hours.

For each tracking run, the subsequent procedure will be followed:

1. The researcher will familiarise the participant with all tasks and display settings;
2. The researcher will clearly announce the end of the training phase and the start of actual measurements;
3. The researcher configures the simulator to the right condition settings before every trial;
4. The researcher checks whether the participant is ready to proceed and initiates the trial;
5. The participant performs the tracking task for the duration of 90 seconds;
6. The participant will be notified of the completion of a trial and/or condition.

Contact information researcher:	Contact information research supervisor
Maarten Kemna	dr. ir. Daan Pool
m.h.h.kemna@student.tudelft.nl	d.m.pool@tudelft.nl
+31 6 42976216	+31 15 2789611

Thank you for participating!

Experiment Consent Form

Manual control behaviour in stereoscopic vision-enhanced depth control tasks

I hereby confirm, by ticking each box, that:

1. I volunteer to participate in the experiment conducted by the researcher (**Maarten Kemna**) 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 and confirm 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 steady, non-motion simulator setup with stereoscopic 3D display. ☐
4. I confirm that the researcher has provided me with detailed safety and operational instructions for the hardware (simulator setup, force feedback sidestick, stereoscopic 3D display) used in the experiment. ☐
5. 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. ☐
6. 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 or, if necessary, the TU Delft HREC (hrec@tudelft.nl). ☐
7. 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
Maarten Kemna	dr. ir. Daan Pool
m.h.h.kemna@student.tudelft.nl	d.m.pool@tudelft.nl
+31 6 42976216	+31 15 2789611