Eye and body tracking in the lab, in the wild, and in the clinic

by

Scott Allan Stone

A thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Psychology
University of Alberta

Abstract

In behavioural psychology experiments, we try to ask and answer questions about real life phenomena. For example, we may be interested in learning how long it takes for someone to react to a red light turning green whilst driving. Arguably, this is a hard question to answer due to the difficulty in actually collecting the data. One concession that might be made is to ask a similar question in a much more controlled environment. For example, if we want to know the average response time to look at a salient target, we may design a study that attempts to isolate this response. Perhaps a dot is shown on a computer screen, and we measure the average response to initiate a saccade. This is valid science, but it is hard to say that it represents real life. Real life is dynamic, and as a consequence, much more noisy. We don't sit in rooms with isoluminant lighting at exactly 40cm from the screen and try to move 'as fast and accurately as possible' in our everyday life. In this thesis, I attempt to challenge these assumptions through the use of eye and body tracking techniques. I conducted three studies, each increasingly giving up experimental control to collect more naturalistic data.

In the first study, I demonstrate that tightly-synchronized simultaneous collection of laboratory-grade eye- and body-tracking can be used to generate accurate 3D gaze vectors that have sub-centimetre accuracy in optimal conditions. Generating 3D gaze vectors is a challenging problem because there are no widely used established calibration routines. This study aimed to answer this question. Participants completed four different calibration routines, where the pupil position data was used as an input to a model that predicts gaze position in 3D space. A validation procedure is used

to assess model performance when the experimenters know exactly when and where the participant is supposed to be looking. Finally, participants complete a previously validated ecologically valid task known as the Pasta Task to assess the model's performance during real-world behaviour. This study provides tangible advice for researchers interested in collecting naturalistic simultaneous eye and body tracking data.

In the second study, I use a consumer-grade eye tracker and a simple hand tracker (i.e. a computer mouse) to record the behaviour of participants as they navigate through a menu interface modeled after the popular video game, Mass Effect 3. This work was conducted as part of a Mitacs Accelerate internship with BioWare in Edmonton, Alberta. The purpose of this study was, in technical terms, to find key performance indicators to aid user experience researchers at BioWare. In other words, I collected eye and mouse tracking data and analyzed the data to find when users were having trouble. Typically user experience researchers use qualitative measures such as interviews, surveys, and technical reports to uncover problematic areas of user experience known as friction points. Here, I use eye and hand tracking to attempt to quantify friction as a dynamic (i.e. not static) process. I demonstrate a methodology that allows for the detection of friction points based on gaze and movement signals. I conducted both an in-person (local) cohort as well as a remote cohort collected using webcam based eye tracking. I show that many of the same patterns of friction that occur in the local cohort also occur in the remote cohort, suggesting that the increased noise (due to decreased experimental control) in the remote cohort did not overpower the signal. The study provides evidence that many of the techniques and tools used in the laboratory can be adapted for use in digital environments.

In the third study, I assessed the feasibility of using eye tracking in a clinical setting to augment the neurological assessment of vertigo. Vertigo is a condition that manifests as extreme dizziness due to imbalance in the vestibular system. Diagnosis is challenging because vertigo is not a single disease entity but the cardinal symptom

of different diseases of varying etiology ranging from benign to deadly. The high stakes involved in assessing a potentially life-threatening prognosis (in the case of a brainstem stroke, for example) requires extra care and resources dedicated to these patients such as computed tomography (CT), or magnetic resonance (MR) imaging. A typical neurological battery will assess the vision of the patient to test oculomotor function, which is critical for diagnosis. Here, we created a simple systematic screen-based set of stimuli that approximates the neurological assessment given to a patient with vertigo symptoms. We used a portable eye tracker to collect eye position during the approximately five minute stimulus presentation. Moreover, the data collection was performed by someone without any background in eye tracking, testing the claims of modern eye tracker manufacturers regarding ease-of-use and portability. Data were collected from both normative controls as well as patients. The assessment was designed to be systematic and easy to use, completing in around 5 minutes. We found that control participants worked well for the stimulus, yet patients produced poor quality data. We created analysis pipelines for the data and speculate on the use of such a device to increase the efficacy of a physician performing neurological examinations on dizzy patients.

These studies, taken together, are an investigation of how we can start to use many of the laboratory technologies out in the wild and the resultant successes and limitations.

Preface

All ideas, data, and analyses presented here are my own, developed in collaboration with my supervisors Drs. Craig Chapman and Anthony Singhal. This thesis received research ethical approval from the University of Alberta human research ethics committee (Pro00087329).

Chapter 2 of this thesis has been published as S.A. Stone, Boser, Q.A., Dawson T.R., Vette, A.H., Hebert, J.S., Pilarski, P.M., & Chapman, C.S. Generating accurate 3D gaze vectors using synchronized eye tracking and motion capture. *Behavior Research Methods*, doi:10.3758/s13428-022-01958-6.

Chapter 3 of this thesis has been accepted for publication as part of the Eye Tracking Research Applications conference and will be published in the *Proceedings of the ACM on Human-Computer Interaction* as S.A. Stone, Chapman, C.S. Unconscious frustration: using eye and mouse tracking to dynamically assess user experience on a menu interface. *Proceedings of the ACM on Human-Computer Interaction.*

Chapters 1, 4, and 5 are original works.

To my lovely wife-to-be and partner in life, Becky.

Acknowledgements

I've never been any good at writing things like these, so if you don't see your name here, please know that you are important and appreciated.

First and foremost, I would like to thank my supervisors, Craig Chapman and Anthony Singhal. Your guidance and knowledge has been instrumental and invaluable throughout my PhD. Anthony, thank you for always finding time to chat, whether it was about research or not. Craig, I always appreciated your open-door policy (at least in the "before" times), and your willingness to help me work through whatever problem I was having. Thanks for being a great mentor and great colleague at the same time.

I would also like to thank everyone on my examining committee: Claudia Gonzalez, Alona Fyshe, Bin Zheng, and Chris Dijkerman. Your insightful comments helped ground me to make sure my research was not going completely off the rails, whilst also encouraging me to push the boundaries of what was possible.

A huge thank you to everyone in the ACELab of past and present: Nathan Wispinski, Jennifer Bertrand, Jeff Sawalha, Ewen Lavoie, and Alex Ouellette Zuk. Thanks for letting me stop everyone from being able to work for long periods of time so I could rant about whatever was bothering me that day. In particular, a huge thank you to Nathan Wispinski for being the other half of the Booth Boys while being an excellent colleague and collaborator, but also a great friend in the process. Jenn, thank you for your deep insights and long chats / vent sessions about research and life.

Obviously, I would never be able to do anything like this if it wasn't for the support

I received from my close friends and family. Mom and Dad: thank you for always encouraging me and being there when I needed your support. Kayla and Chase: I appreciate that you preemptively moved to Calgary just for me! Thank you for always being there when I needed to chat. All of my video game boys: bless all of you. Gus and Horatio: please stay off the counter.

Becky, thank you for all that you do to bring brightness to my life. Thank you for always being there to listen to me, help me work through my problems (even if I don't seem like I want to), and always encouraging me to finish what I started. Who would have thought that MATLAB camp could lead to this? You are the love of my life and everything you do is deeply appreciated.

Finally, a big thank you to you, whoever you are, for reading this thesis. I promise I didn't write it just for myself.

Table of Contents

1 Introduction **1**

1.1 Motivation 2

1.1.1 The role of experimental control 2

1.1.2 Why record gaze and movement data? 5

1.1.3 What counts as useful? 7

1.1.4 This thesis in context 8

1.2 Vision and eye tracking 10

1.2.1 From vision to vision for action 10

1.2.2 History of eye tracking 12

1.2.3 Contemporary eye tracking 13

1.3 Movement and motion capture 15

1.3.1 History 15

1.3.2 Contemporary motion capture 17

1.4 Synchronization of eye and motion capture streams 20

1.5 Previous attempts of “in the wild” eye tracking and movement data collection 21

1.6 Aims and objectives 26

2 Generating accurate 3D gaze vectors using synchronized eye tracking and motion capture **36**

2.1 Introduction 38

2.2 Methods 41

2.2.1 Equipment 41

2.2.2 Participants 42

2.2.3 Procedure 42

2.2.4 Data Processing 49

2.3 Results 55

2.3.1 Factors for rmANOVA 55

2.3.2 Calibration Trials 56

2.3.3 Validation Trials 56

2.3.4 Task Trials 58

2.4 Discussion 61

2.5 Conclusion 65

3 Unconscious frustration: dynamically assessing user experience using eye and mouse tracking 70

3.1 Introduction 72

3.1.1 Background: user experience 72

3.1.2 Background: tracking the eyes and hands 72

3.1.3 Related works: eye and hand tracking in UXR 73

3.1.4 Assessing dynamic behaviours 75

3.2 Methods 76

3.2.1 Participants 76

3.2.2 Equipment 77

3.2.3 Task 78

3.2.4 Procedure 81

3.3 Dependent Variables 82

3.4 Results 84

3.4.1 Statistical test designs 84

3.4.2 Statistics: Time 84

3.4.3 Statistics: Gaze 87
3.4.4 Statistics: Coordination 91
3.5 General Discussion 92
3.5.1 Main findings 92
3.5.2 Local versus Remote cohorts 93
3.5.3 Moving beyond the laboratory 95
3.5.4 Privacy 96
3.5.5 Conclusion 97

4 Eye tracking in the clinic: an investigation of an easy-to-use automated clinical eye movement assessment 102

4.1 Introduction 104
4.2 Materials and methods 109
4.2.1 Task & system 110
4.2.2 Equipment 110
4.2.3 Eye tracking procedure 111
4.2.4 Motivation 114
4.2.5 Data collection 114
4.2.6 Software 115
4.3 Environment 115
4.4 Participants 117
4.4.1 Controls 117
4.4.2 Patients 117
4.4.3 Data 118
4.5 Analysis 118
4.5.1 Stare 118
4.5.2 Pursuit 122
4.5.3 Vestibulo-ocular reflex (VOR) 123

4.5.4 Jump . . . 125
4.5.5 Brightness . . . 127
4.6 Results . . . 128
4.6.1 Performance . . . 129
4.6.2 Control data . . . 129
4.6.3 Patient data . . . 132
4.7 Discussion . . . 135
4.8 Conclusion . . . 143
4.9 Acknowledgements . . . 143

5 General Discussion 149

5.1 What is the real world? . . . 151
5.1.1 Ecological validity . . . 152
5.1.2 Each study in context . . . 153
5.2 Study 1: eye and body tracking in the lab . . . 153
5.2.1 Summary . . . 153
5.2.2 Spectrum . . . 156
5.2.3 Limitations . . . 156
5.2.4 Utility . . . 157
5.3 Study 2: eye and body tracking in the wild . . . 158
5.3.1 Summary . . . 158
5.3.2 Spectrum . . . 159
5.3.3 Limitations . . . 160
5.3.4 Utility . . . 161
5.4 Study 3: eye tracking in the clinic . . . 162
5.4.1 Summary . . . 162
5.4.2 Spectrum . . . 163
5.4.3 Limitations . . . 164

5.4.4 Utility 165
5.5 Retrospection and future directions 165

Bibliography **172**

Appendix A: ETRA 3D gaze vector publication **188**

Appendix B: Supplemental mouse movement statistics **193**

List of Tables

4.1 A description of all admitted patients in the study. General information about sex and age were collected, as well as whether nystagmus was present. The HINTS column marks whether or not the battery was used on the patient. Stroke presence and the appropriate diagnosis is listed in the last 2 columns. 117

4.2 Table showing the means of all the measures collected for the control group, with between-subject means and standard deviations on the bottom two rows. Where applicable, data is show for each eye (P_0, P_1), and for dispersion measures the major (width) and minor (height) axes (x & y) are noted. All units, unless specified are in standardized (0–1) space. Stare dispersion shows the amount of dispersion calculated by the ellipse for each eye. The Pursuit dispersion at the hold locations is shown in the same form as the Stare data. The average Pursuit hold velocities (i.e. gaze should be stable) is shown. The average Pursuit movement velocity (i.e. when the eyes are smoothly tracking the object) are shown. The VOR cross-fitting root-mean-squared error is shown. The Jump accuracy (as a percentage) and average reaction times are shown. 133

4.3 Table showing the means of all the measures collected for the patient group, with between-subject means and standard deviations on the bottom two rows. Where applicable, data is show for each eye (P_0, P_1), and for dispersion measures the major and minor axes (x & y) are noted. All units, unless specified are in standardized (0–1) space. Stare dispersion shows the amount of dispersion calculated by the ellipse for each eye. The Pursuit dispersion at the hold locations is shown in the same form as the Stare data. The average Pursuit hold velocities (i.e. gaze should be stable) is shown. The average Pursuit movement velocity (i.e. when the eyes are smoothly tracking the object) are shown. The VOR cross-fitting root-mean-squared error is shown. The Jump accuracy (as a percentage) and average reaction times are shown. It is worth noting that the overall quality of the data appears to be quite low. 136

List of Figures

1.1 Example of DeepLabCut applied to an astronaut exercising in space. 19

1.2 High-level diagram of the functionality of Lab Streaming Layer 22

2.1 The four Calibration routines used in the study: A) Experimenter Sweep, B) Self Sweep, C) Experimenter Paint, and D) Stationary Target 44

2.2 Example gazes corresponding with each of the calibration routines. A) Experimenter Sweep, B) Self Sweep, C) Experimenter Paint and D) Stationary Target 46

2.3 Example still shot of a participant midway through completing the Pasta Task. 47

2.4 Locations, objects and markers critical for all three of the tasks. Neutral, Side Cart, Home, Green Shelf, and Blue Shelf locations are shown. The Pasta Box is shown. The Head Rigid body is shown on the participant. 48

2.5 Visual demonstration of the differences between the Cartesian and Spherical coordinate systems used. 52

2.6 Plots showing the average performance of the models in linear distance (LD) and visual angle (VA) errors for the Calibration trials. 57

2.7 Plots showing the average performance of the models in linear distance (LD) and visual angle (VA) errors for the Validation trials. 59

2.8 Plots showing the average performance at each location, split by the pupil input data in linear distance (LD) and visual angle (VA) errors for the Task trials. 60

3.1 A waterfall representation of the order of frames that the participant will encounter in a typical trial. 80
3.2 The eye distance traveled over the course of an entire trial split by the phase. Data for both the Local and Remote cohorts are shown, split by the type of eye tracker used. 86
3.3 The average gaze distance traveled navigating to the goal frame. Data for both the Local and Remote cohorts are shown, split by the type of eye tracker used. 89
3.4 The average proportion of $\text{Tlead}_{\text{NaN}}$ values for each type of goal frame. Data for both the Local and Remote cohorts are shown, split by the type of eye tracker used. 91

4.1 Each of the stimuli presented the patient, as described in section 4.2.3. Each stimulus represents what would be presented on the screen. . . . 116
4.2 Here, four plots of eye dispersion data are shown. The X and Y axis of each plot are scaled to the control data to allow for easy reference. 121
4.3 A visual example of data from a single eye collected during the VOR task, with the XY data points plotted in blue, the mean of the XY data plotted in as a green dot, and the fitted cross drawn in red. . . . 124
4.4 An example plot of the eye velocities during a jump in the Jump task. The x axis is time (seconds) and the y axis is velocity as a standardized unit (z-score). 126
4.5 Overall results, showing the average Stare dispersion, Pursuit dispersion, Pursuit velocity, VOR error, and Jump performance. 137

5.1 A continuum of objects falling along a continuum of realness. 168

Chapter 1

Introduction

1.1 Motivation

One of the reasons we do science is to learn something about ourselves, and the world around us. Science is a vigorous process; we have to be incredibly careful with our assumptions, data collection procedures, and analysis techniques. We do not want to get anything wrong, or at least get it wrong because we are being careless. We achieve this through control. Control of our experimental setups, control of the dependent variables we collect, and control over the statistical analyses that we perform. We do this because, again, we want to be right. But at some point, being right does not matter if it does not reflect reality anymore. Sometimes, the way that we study behaviour does not reflect to how people *actually* behave.

The world is a dynamic and noisy place. To study a phenomenon, we try to distill it into its individual parts to make it easier to study. However, this reductionist approach can lead us to oversimplify the complex phenomena we set out to study and ignore the richness and diversity of real-world situations. Understanding the roles and limitations of control, and how they affect our study of psychology, is important for contextualizing the findings that emerge from our research.

1.1.1 The role of experimental control

Experimental control is the ability to manipulate and measure variables in a way that allows for a cause and effect relationship to be established. In scientific terms, we want to manipulate only the independent variable and measure the dependent variable and test if there are any differences following the manipulation. For example, imagine you are a chemist who is interested in investigating the effect of two different temperatures on a chemical reaction. Here, it would be important to ensure that the amount and concentration of each chemical, the amount of stirring, and the length of reaction is identical between the two samples. In other words, the only thing that is different between the two samples is the temperature applied. Following measurement, any differences noted could be more easily attributed to the change of temperature and not anything else. But, not all fields can afford the same level of control over the environment that the data is collected in. In this example, there is not a lot of wiggle room for variability. The two chemical samples are assumed to be identical, we can measure the exact temperature of the reaction, and we can ensure we are completing the task in the same amount of time. While in this example, it is relatively simple to achieve experimental control, psychology presents some unique challenges due to the complexity and variability of human behaviour and the difficulty in isolating and controlling for all relevant variables.

Achieving experimental control in a human psychology experiment is not as straightforward. In this case, there are several considerations that need to be made that do not apply to the chemistry example above. First, humans are variable. Not everyone is the same, nor do they necessarily have the same personality, preferences, or abilities. Of course, these are factors that can alter the outcome of an experiment. Second, there are ethical considerations that need to be made for humans; some manipulations may not be appropriate or feasible. As researchers, we must respect and care for our subjects. Third, many of the phenomena are incredibly complex and difficult to isolate from other behaviours. As an example, one such behaviour, social attention, can be challenging to study because it is hard to isolate the essence of a social interaction.

Probing this idea of what defines a social interaction, a study by Laidlaw *et al.* [1] investigated participants' eye movements while sitting in a waiting room. In this study a confederate posing as another research participant was either physically present or shown on a computer screen. The authors found that participants looked at the other person less when they were physically present, suggesting that social attention (via eye movements) is heavily modulated by the 'mere opportunity for social interaction'. This study demonstrates how easily behaviour can be modified by something as simple as another person being physically present. Additionally, we know that humans tend to behave differently when they know they're being watched. For example, many people are familiar with *white coat syndrome*, a well-known phenomenon in the medical community which describes the rise in blood pressure, anxiety and perspiration that some patients experience during routine checkups [2, 3]. Even in a laboratory setting, participants have been found to be more likely to 'choke' when being watched by an experimenter [4]. It is perhaps useful to look at these responses not as noise in the data, but rather as part of the data itself; this is how people are *truly* feeling in the moment.

So, how do we know what signals matter and what do not? For example, in a psychophysics experiment, it may not be clear what should be used to measure reaction time—brain signals, eye movements, body movements or some other signal. This is not to say that it is not possible to achieve some level of control in psychology experiments, but rather it requires careful consideration of the inherent variability of human beings. One of the ways researchers do this is by being clever in the way they design experiments to isolate the behaviours they wish to study. However, it is important to note that sometimes these experiments may not have a real-world analog, yet we still try to apply the results to real life. So how do we create analogous tasks?

Imagine you are a psychology researcher who is interested in understanding what people pay attention to when choosing a box of cereal off the shelf at the supermarket. To assess attention, you elect to record where people look and how they move, which is treated as a proxy for attention. One of the problems here is that it is very difficult to record these kinds of signals in a supermarket. One reason might be the increased noise that is inherent to this type of environment. Second, the ergonomics of recording this kind of data are challenging to say the least. To alleviate these concerns, a concession you might make is to take the person out of the supermarket entirely, and put them in a setting that that controls for the noise and distractors. Since you only care about where someone is looking and what they ultimately choose, you might reduce the task to a digitized version that shows three different rows of cereal brands on a computer screen. Then, to make sure you are getting consistent data, you make sure the person is seated in a chair at a fixed distance from the screen, and adjust the lighting to ensure optimal data collection. The study as described does a good job of isolating the phenomena we are interested in studying, but it does not reflect the natural action anymore. If a researcher wants to study the natural behaviour, it logically follows that the task should be as natural as possible, even if it means including more noise in the data. To achieve this, the researcher should choose signals that are robust to noisy environments, such as eye and body movements. Eye and body movement signals have value because they are the modalities we use to collect information and interact with our environments.

1.1.2 Why record gaze and movement data?

Humans are visual creatures. One of the ways we learn about our world is by looking at it. The way that we move our eyes reflects the cognitive processes occurring in the brain [5]. It naturally follows then, that our recording of the eyes should grant insight into the inner cognitive workings of the mind. Previous work has demonstrated the intrinsic link between visual behaviour and cognition in a variety of tasks including visual search [6, 7], reading [6, 8], visual working memory [9, 10], and during naturalistic movements [11–14]. It tends to be difficult to obtain high quality data from naturalistic tasks due to the challenges with control explained previously. For example, Land & Lee [11] investigated where participants looked when they were steering a car around a bend. The presumption for this study was that no equivalent laboratory analog task existed, and the best option was to record the eye movements of actual drivers whilst driving. Here, the authors found that drivers relied on a tangent point in the bend approximately 1 to 2s prior to beginning the turn. It is likely that the pressures (and distractors and noise) the driver faced impacted their behaviour, which cannot be modeled in a laboratory task. For instance, the eye movements (and intense manual labour to analyze them) are robust enough to survive extraneous noise that the environment provided, which suggests that eye movements can (and should) be recorded in natural contexts to get the best representation of the behaviour a researcher in interested in. Of course, eye movements are just one part of a set of complex behaviours such as guiding the hands (or steering wheel in this case) to complete a larger goal. Clearly, how we look and how we move is important to understanding cognition.

Our limbs are the primary way we interact with our environment. For instance, envision waking up to an unfamiliar hotel room to an alarm during a business trip, which you fumble to deactivate. To an outside observer, your hand movement's speed, accuracy, and trajectory may indicate that you're not acquainted with the environment. Just like when we watch someone attempt a new sport, we can perceive their lack of experience through their actions. The way we move, similar to eye movements, is also an indicator of the dynamic cognitive processes in brain. For example, studies investigating reaching behaviour have shown that direct reaches are associated with higher confidence than indirect reaches in both hand [15–19] and mouse movements [20–24]. Recording movement then, is important to understanding the goals of an individual. Movements of the hand in particular are closely linked with the movements of the eyes. This is because we use visual input to direct our actions—known as eye-hand coordination. Evidently, studying both of these modalities simultaneously will give the most information about the intentions of someone's actions, which could be useful to understand the context of the behaviour.

Studying visuomotor activity provides a unique and dynamic understanding of human perception. By tracking where someone looks and how they interact with their surroundings, we can gain insight into their attentional focus and decision-making processes. While static measures such as attention heatmaps or movement trajectories may provide some useful data, they fail to capture the complexity of our dynamic interactions with the world around us. Furthermore, the relationship between vision and action is not static, and visuomotor recordings allow us to track this relationship over the course of an entire task dynamically. In the above example studying cereal brand preference on a shelf, a static measure such as fixation time would fail to capture important dynamics. For example, if someone alternates between two different choices, a measure such as total fixation time on each object does not include how many times the person oscillated between the choices. In this example, the oscillating behaviour is an important indicator of indecision.

Clearly, multimodal dynamic visuomotor data has a lot of value for understanding human behaviour. Despite its richness, it does tend to be more challenging to interpret. While this may seem like a downside, it gives researchers the opportunity to decide what they consider useful in their data.

1.1.3 What counts as useful?

People do not naturally sit in isoluminant rooms at exactly 40 cm from their computer screen and consistently move "as fast and accurately as possible". Of course, specific questions require specific setups, but if we want to understand how people *naturally* behave, we should move towards more natural data collection techniques and experimental designs. And, as the preceding section details, focusing data collection on gaze and movement behaviours is a prime target. Arguably, experimental designs that effectively capture the dynamics and complexities of the environment and encourage real behaviours will have more utility than analogous lab-based tasks.

Utility is often defined by the person who is receiving the data. To a grocery store manager, understanding how people navigate and interact with the shelves in their store is much more useful than the laboratory equivalent described above. Of course, a professor running a laboratory that studies attentional mechanisms may have an entirely different set of criteria they consider important. In a similar vein, if a website designer wants to optimize their website, it would make sense to have users navigate their actual website rather than an equivalent recreation. Knowing how people focus their attention on the website is useful in understanding attentional mechanisms, but it is unlikely that the website designer will be able to apply this knowledge directly. Useful data should make predictions while representing the underlying phenomena being studied. Sitting in a dark room staring at dots on a screen has little ecological value because it does not reflect what humans actually do.

Here, I will attempt to persuade you that giving up control in order to better understand natural human behaviour is a justifiable trade-off. Visuomotor recordings are the best way to measure natural human behaviours because they are objective measures of how someone actually operates in the world. While more noise may be present in the data, I will show that it does not prevent useful metric extraction whilst allowing for data collection in more real-world settings.

1.1.4 This thesis in context

In the following thesis, I will present three distinct studies that have varying levels of control applied to each. The link between the studies is intended to be a gradated loss of control the experimenter has over the participants, the environment the data is collected in, and the dependent variables themselves while encouraging natural behaviours to emerge. If we want to study these behaviours, we have to measure them in the context that we are interested with the tools that we have. Additionally, it is also important to target the end-user of the generated data for each study. Below I will give a high-level description of the studies in this thesis, with a more fulsome description later on.

In the first study, I took one step away from a traditional controlled data collection paradigm and targeted scientists who are interested in collecting data from real-world tasks in a laboratory. Gaze behaviours were assessed during calibration sequences to generate models to predict where someone is looking in 3D space. The models were validated and assessed on a naturalistic task known as the Pasta Box task [14, 25, 26], where participants are given free movement of their bodies to complete the task.

The second study moved out of the lab and into the world of user experience (UX). UX is a field that necessitates real-world behaviours because researchers want to assess how users actually use their products. I assessed the gaze patterns and mouse movements of participants tasked with navigating a simple video-game inspired menu system. Natural movements were critical because I aimed to detect friction points: unhelpful slowdowns experienced while using the menu system. A simple instructional prompt manipulation was used to encourage exploration of the menu system without giving the users any specific instructions. The output of this study was intended for UX designers and researchers to allow for the improvement of their products.

The third study moved into the most difficult environment thus far: the clinic. This was the most ambitious study, which ultimately hit the limit of the current technology. Eye tracking data was collected from patients experiencing dizziness at the University of Alberta hospital. A clinically inspired set of stimuli were presented to the patients, designed to be akin to a clinical neurological exam. The study assessed the feasibility of collecting data in a completely uncontrolled environment such as a hospital stroke clinic. The target user here was the attending physician, who could use the output of such a device to improve clinical diagnostic efficiency. Here, although we tried to collect reach trajectory data, ultimately the patients were unable to complete this portion of the task.

I use both eye tracking and movement analysis to assess real-world questions about the behaviours that humans exhibit within their environments, ranging from a dedicated laboratory motion-capture room, to measuring user experience whilst navigating video game inspired menus, to eye movements from patients at the hospital. I hope to convince you, the reader, that as control decreases (and noise increases), there are still useful metrics buried in the data that can be captured through eye and body movement recordings.

1.2 Vision and eye tracking

1.2.1 From vision to vision for action

Understanding the relationship between vision and action is crucial in studying natural behaviours from a psychological perspective. Contemporary psychology has swung the pendulum between control and naturalistic behaviour towards control over the last few decades. This is perhaps in part because for a long time we thought of perception, and specifically visual perception, as a passive process. Because earlier psychology research neglected the embodied aspect of human experience, we thought it was appropriate to put people in dark rooms hitting buttons to respond to targets. Of course, on the surface, we know that these types of behaviours are not reflective of how people actually behave. Rather, we believe they isolate the neural and behavioural mechanisms that we are interested in studying. Over time, it was recognized that perception plays a critical role in facilitating action. In the following section I will discuss the gradual shift from pure vision research towards vision for action.

Prior to the 1980's, vision research was almost exclusively perceptual in nature [27]. In fact, in 1948 Lashley [28] once famously concluded that visual processing did not extend beyond the striate cortex. There have been debates about whether the primary function of vision is to construct perceptual experiences through bottom-up edge-detection techniques [29], or whether it involves top-down processes moderated by the frontal cortex [27], which enable the selection of predefined stimuli present on the retina [30]. Much of the work done at this point was investigating the functional segregation of the brain, opting to study the mechanistic features of visual processing, and not studying the behavioural outputs of vision.

The visual system's non-monolithic nature was some of the first evidence presented by Schneider [31], who used golden hamsters to demonstrate that lesions in the striate cortex led to pattern discrimination deficits, while tectal ablations eliminated visually guided head orienting responses toward food. Schneider called this evidence for two visual systems at play. Work in the early 1970's demonstrated that two visual systems also existed in parallel in the frog brain [32]. Ingle sought to dissociate the function of the visual thalamus from that of the optic tectum by ablating the tectum. Accidentally, he found that allowing fibers from the affected eye to regenerate over the course of 6-8 months to the ipsilateral tectal hemisphere resulted in mirror-symmetrical frog snapping behaviors towards prey, but barrier avoidance was unaffected. This suggested the presence of two distinct visual systems, similar to the work by Schneider.

Mishkin & Ungerleider [33] provided some of the first evidence that vision was split into two distinct cortical pathways in primates. Mishkin & Ungerleider demonstrated lesions that disconnected parieto-preoccipital areas from striate cortex resulted in the monkeys having trouble either 1) identifying an object or 2) being able to localize it in space. They determined that if the lesion was found to be dorsal to the parietal cortex it resulted in localization deficits, but ventral lesions lead to object identification deficits. These deficits were evidence of what Mishkin *et al.* called the "where" and "what" visual pathways. A fundamental issue with this interpretation is that is does not question what the output of such a system would be used for—how does the brain use this information for interaction?

Answering this question was the seminal work of Goodale and Milner [35]. This work work demonstrated vision is critical for initiating and guiding movements through space, suggesting that cortical visual processing was not necessarily "what" and "where", but rather "what" and "how". Going further, Goodale [27] argues that vision for perception is literally only half of the story; vision is also used extensively for action [35–38], and in fact is controlled through a different set of mechanisms than those used for perception (for an excellent review, see Gallivan & Goodale [39]). These separate pathways have now become known as the dorsal and ventral visual streams. Colloquially, the dorsal and ventral visual streams are now known as the

”action” and ”perception” streams, respectively [27]. In short, the control of action is necessarily an online process, requiring quick moment-to-moment adjustments for efficient interaction. In contrast, perception typically does not update as quickly, as objects do not tend to change frequently over time. As a result, the perception stream updates much slower. The two streams work in parallel and are not thought to be hermetically sealed from one another [27, 40]. Both streams are intimately linked and work together in complementary roles, controlling our behaviour and allowing for adaptation. For example, as I am sitting here, I would like to take a drink from my coffee cup. My ventral (perception) stream processes the image on my retina and tells me that there is a coffee cup. To actually reach for the cup, however, my dorsal (action) stream processes my end-effector’s (i.e. my hand’s) path to the cup by converting the visual coordinates into motor commands. The dorsal stream processes information at a much higher rate than the ventral stream, making it much more suitable for temporally sensitive adjustments around any potential obstacles [41]. Just as I am about to grasp the cup, a cat jumps on the desk and nearly knocks it over. The dorsal stream is able to quickly recalculate my reach trajectory so I can save the cup before it hits the ground. Of course, in the real world, we perform these kinds of movements all the time. When reaching into our pantry to get a can of soup, we have to avoid all of the other items that are in the way to navigate our hand to the can. The visual system’s ability to convert between visual and motor coordinates is nothing short of incredible, and tracking the movement of the eyes is one of the primary ways that we can gain insight into how this conversion takes place.

1.2.2 History of eye tracking

Eye tracking is a technique that records the pupils, which is typically used to calculate gaze position in space. One of the earliest known attempts to non-invasively track the eye’s position was by Dodge & Cline [42], who were able to successfully quantify horizontal eye angle velocities using a photographic method that recorded eye movements onto a piece of sensitive film. This proved to be a useful endeavor, as it was the first demonstration of what normal movements of the eyes look like. Crucially, this method did not provide real-time feedback to researchers who were interested in using the eye movements themselves as a feedback signal into their experimental design. It was not until 1939 that both horizontal and vertical movements of the eyes were recorded using electrooculographic electrodes placed around the eyes [43, 44]. This method, although more invasive, was the first time that eye movements were recorded in real-time (and in fact were, to attempt to quantify the presence of nystagmus in the eyes [45]). Later on, Yarbus [46] used a complex mirror system to record movements of the eyes and was the first to observe that eye movements were intrinsically linked to the cognitive goals of the participant. In this regard, Yarbus was a pioneer in the eye tracking community and the first to make this association. When a subject was told to look at painting with varying instructions, he noted that the gaze patterns differed depending on what the participant was trying to achieve. Early on, fixations were mostly on the faces of the subjects in the painting, whereas later on fixations tended to be on extraneous objects (e.g. chairs, tables) [47]. Over time, it was observed that the subjects continued this pattern, suggesting there was structure behind these eye movements. This observation, which may seem obvious by today's standards, was critical for the understanding of the role of eye movements for perception. In recent times, rapid technological improvements have meant that tracking the eyes is easier and more affordable than ever before.

1.2.3 Contemporary eye tracking

In modern times, we use specialized devices known as eye trackers. Generally speaking, eye trackers use cameras pointed at the eyes to track the pupil. There are a multitude of methods used to track the pupil, but many eye trackers tend to use a combination of video-based and infrared-based approaches. In modern psychology, there are two main types of eye trackers typically used: desktop and head mounted.

Desktop systems tend to have have much higher spatial and temporal resolution at the cost of being stationary. Additionally, these systems usually require fixing the head position, but this is not always the case. The more portable head-mounted eye tracker allows for free movement of the head, but usually still requires a dedicated computer during recording.

Desktop eye tracking

Desktop eye tracking refers to the method of recording the eye movements of a participant sitting in front of a computer screen. A typical desktop eye tracker, such as the EyeLink 1000 (SR Research, Ottawa, Canada), offers high spatial and temporal resolution, but does not permit a large amount of free head movements. There is, however, a burgeoning field of monitor-mounted eye trackers, such as the Tobii Eye Tracker 4C (Tobii Research AB, Sweden) and the Gazepoint GP3 [48]. These systems are intended for use on a dedicated desktop machine in order to track the gaze position on a monitor. An advantage of these systems is that they do not require the end-user to wear any equipment and can non-invasively record the user's gaze position at relatively high frequencies (typically 90-150hz).

Finally, webcam-based eye tracking algorithms have become popular in the last few years. A consumer grade webcam (e.g. laptop or monitor-mounted webcam) is used to record video of the participant's eyes, from which a pupil position can be ascertained. These techniques typically require a more intensive calibration procedure, but can yield relatively high quality eye tracking data [49]. Webcam eye tracking is especially suitable for remote eye tracking tasks because almost every computer in modern times will have an attached webcam. These options work well for tasks that are relatively stationary, but an eye tracker that allows for free head movements may be desirable.

Head mounted eye tracking

Mounting the eye tracker on the head of the participant allows the researcher to account for head movements. Earlier versions of these eye trackers were typically still quite cumbersome and expensive (e.g. EyeLink II, SR Research Ltd.), but modern head-mounted eye trackers tend to be lightweight, portable, and much more affordable. Portable systems, such as the Pupil Labs Core [50] use a video-based deep-learning approach to detect the pupil, as well as to track other pupillometric data. Because these systems are portable, they are more suitable for deployment outside of traditional eye tracking laboratories, such as in a public grocery store [51]. These devices facilitate data collection in the wild. Recording naturalistic behaviour requires minimal restrictions for the participant, and these devices give freedom of movement to the person being recorded, making them suitable for real-world tasks.

1.3 Movement and motion capture

1.3.1 History

One of the earliest examples of capturing an animal in motion was the famous recording of *The Horse in Motion* by Muybridge [52]. These were a series of chronometric photographs that captured the movement of a horse as it galloped by. This was one of the first earnest attempts to better understand the biomechanics of how a horse moves. This study was no easy feat; multiple cameras were needed to photograph the running horse, which used an electrical shutter system custom engineered for the study. A long wall of white planks were erected to allow for enough brightness to develop the individual frames. Muybridge found that there were times when all four of the horse's feet were off the ground (and tucked underneath, facing each other). This was some of the first evidence that capturing snapshots in time of moving subjects could have great scientific value, such as being able to quantify the movement of animals. This technique would come to be known as motion capture.

Although commercial systems that recorded movement became available in the late 1970's, some researchers still opted to use more primitive methods to effectively motion track animals' movements [53]. As an example, Ellard *et al.* [53] recorded Mongolian gerbils jumping varied distances to investigate the strategies used to estimate distance. To quantify their movement the researchers projected each frame onto a digitizing tablet, which was connected to a computer. On each frame, the head and leg positions were mapped manually. Clearly, this approach was valuable but incredibly cumbersome for anything more than a few seconds worth of data. Over time, affordable devices known as motion capture systems became available that were suitable for use in human biomechanical and movement studies.

Motion capture (sometimes called mo-cap) is a technique used to record human movement in two or three dimensions. In the context of human behaviour, it is an especially useful tool to help quantify various aspects of movement. For example, reach-to-grasp studies can record the exact timing, trajectory, and final destination of the hands [15, 17, 19, 54], from which metrics such as velocity, acceleration, and completion time can be derived. Studying reaching is a non-invasive way of studying the function of the brain. It can be argued that one of the primary outputs of the brain is movement, and using motion tracking to study these outputs gives us deeper insights into the inner workings of cognition and attention [22, 55–57]. Freeman *et al.* [22] gives a compelling argument that hand movements in particular contain "motor traces of the mind", suggesting motion capture can be used as an effective non-invasive tool to study brain systems. As stated earlier, there is extensive evidence that movement is regularly updated via cognitive processing [16, 27, 35, 36, 58, 59], and recording these movements shows us the other side of the coin: how the control signals generated in the brain manifest as movements. Below, I will discuss three prominent and commonly used types of contemporary mo-cap: passive/active systems, markerless motion capture, and mouse tracking.

1.3.2 Contemporary motion capture

Passive and active systems

Today, there are a multitude of options for collecting mo-cap data. One of the most common systems used is an infra-red emitting diode (IRED) camera tracking system (e.g. OptiTrack, or Optotrak devices). As their name suggests, these systems work through the use of infra-red cameras that reflect light off of either passive (i.e. with IRED reflective tape) or active (i.e. powered IREDs) markers to track movement in 3D space. In general, both work approximately the same way; the camera system uses an algorithm to triangulate the location of the marker in 3D space using the known positions of the tracking cameras. Typically, these systems require an extensive calibration procedure prior to use each time. The result is real-time tracking of body movements with the trade-off of requiring a large amount of space for the tracking camera system to be installed. These systems tend to have very high temporal (e.g. >hundreds of samples per second) and spatial (e.g. below 0.1mm with proper calibration) resolution during data collection. Because there are multiple cameras (usually around 8 or more in the case of a full-body motion capture system[1]), data collection methodologies must be restricted to collection in a room dedicated to mo-cap. To put it bluntly, these systems are usually unsuitable for data collection in the real world, as they are much too cumbersome and cannot be properly set up in every space.

Markerless motion capture

Today, it is possible to collect mo-cap data using consumer-grade camera systems, such as a mobile phone. The algorithms powering these techniques are known as markerless mo-cap. More recently, markerless mo-cap systems have been steadily improving as the hardware required to power the algorithms becomes cheaper. This is achieved using deep neural networks to track desired targets on individual video

[1] It is worth noting that depending on the experimental setup, only one or two cameras may be necessary depending on the vantage point of the cameras.

frames. DeepLabCut [60] was originally developed to track animals that could not feasibly have mo-cap markers attached, such as laboratory mice or wild animals. The same principals can be applied to tracking humans (see Figure 1.1 for an example). The resulting data can be exported into a familiar time-series format and can be analyzed in many of the same ways that traditional mo-cap data is processed. One drawback is that DeepLabCut currently does not natively support real-time processing and output at the time of writing[2], making it unsuitable for studies that require real-time feedback for the participant. One such system that specifically allows for real-time feedback is MediaPipe [61]. MediaPipe is a framework that is capable of using video frames as an input, and through a pre-trained model, can detect and superimpose pose estimates, hand position, and even iris detection[3]. These types of systems will likely overtake the passive and active systems in popularity due to their easy ability to deploy almost anywhere using almost any kind of recording medium.

Mouse tracking

In some cases, we may only care about the movement of the hands, such as during a computerized task. Thankfully, the primary way that most users interact with their computers is with a computer mouse, whose movements can be recorded using either custom or already available software (e.g. MouseTracker: [21]). An advantage of recording mouse movements is that the recording does not impact the user's experience at all, as opposed to dedicated motion capture setups which require tracking markers to be physically attached to the participant. Trajectory, velocity, and interaction (e.g. clicks) can be collected very easily, and all contain a wealth of information. Mouse trajectories, like reaches, have been demonstrated to be susceptible to information presented on a screen [20–24]. For example, a study by Rheem *et al.* [24] demonstrated that mouse trajectories were susceptible to cognitive loading, scaled to

[2]There is, however, a tool under active development known as DeepLabCut-Live that aims to allow for this in the near future (https://github.com/DeepLabCut/DeepLabCut-live).

[3]MediaPipe is also capable of doing object detection, tracking, hair segmentation, and many others. A full and current description of capabilities can be found at https://google.github.io/mediapipe/

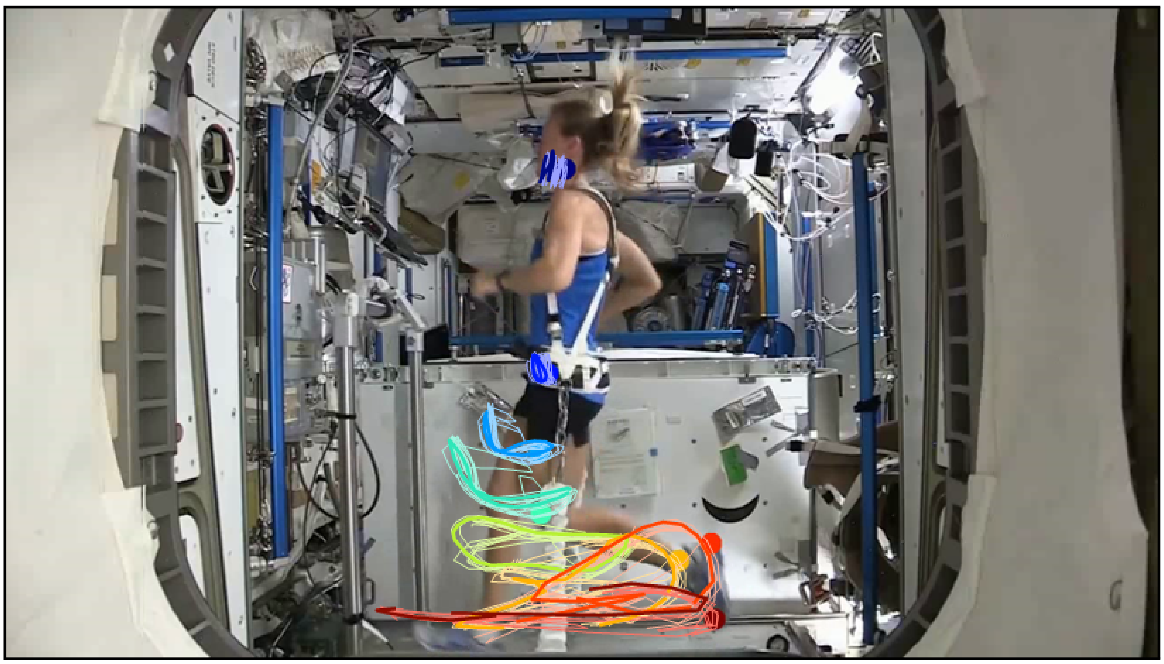


Figure 1.1: An example of a still frame extracted from a video trained in DeepLabCut of an astronaut exercising on the International Space Station. Each dot is a location chosen to track for the duration of the video, with the lines superimposed showing their trajectory over time. It is worth noting that DeepLabCut allows the end-user to track an arbitrary number of locations, joints, or positions in a video.

the perceived task difficulty. Participants performed a dual task, where response time and movement trajectory deviations were larger when presented with a harder task to complete. This is in line with studies demonstrating that movement trajectories are known to deviate more when the actor is less confident in their response [17, 39, 62], meaning mouse movements are analogous mechanistically and functionally to real-world reaches. Given that much of our time as humans is spent using computers (and devices such as smart phones, tablets, and laptops), this is an arguably easier medium to collect real-world data on.

1.4 Synchronization of eye and motion capture streams

Individually collecting eye and motion tracking data is relatively simple. However, many experimental designs benefit from the simultaneous recording of gaze and movement signals. Unfortunately, this is not as straightforward to achieve. To study the temporal dynamics of the coupling of gaze and movement, a methodology is required to ensure the signals can be effectively compared.

For example, in a simple reach-to-grasp task, gaze predicts fixation at the point where the index finger tends to grasp [63, 64]. To effectively calculate the offset between when someone looks at the target (via eye tracking) and when they actually interact (via motion capture), the signals need to be synchronized. Simply recording the streams at the same time is not an appropriate method, as it typically results in a large amount of manual work to synchronize the streams post-hoc [14, 65], as timestamp drift or data collection errors can occur. To effectively extract these kinds of measures, we must use a more sophisticated approach for synchronization.

As previously stated, one of the many challenges of concurrently recording gaze and movement is the time synchronization of the streams. Eye-hand coordination is a tightly coupled dynamic process which necessitates synchronization to accurately study. In their raw formats, gaze and motion capture data are in a time-series format. That is, each sample (e.g. x, y coordinates) is marked with a timestamp. The timestamps associated with each sample however, are not always from the same source. For example, gaze and movement data recorded simultaneously on the same machine may have different timestamps due to the hardware used. When we assess expected relationships, such as gaze appearing on a target prior to reaching for it, it is critical that the timing of these events is known. Luckily, this tight coupling can be achieved through software.

Lab Streaming Layer (LSL) is a software library that aims to solve the multi-modal data streaming problem [66]. At a high level, LSL accepts time series data streams and gives each sample a timestamp from a common source (see Figure 1.2). The timestamp source is the host operating system's internal high resolution clock, which boasts high resolution and accuracy with minimal drift. LSL also applies a drift correction algorithm to ensure that long periods of recording do not have incremental drift. Additionally, LSL has minimal hardware requirements, making it suitable to use on low-powered hardware (such as single board computers or mobile phones) to effectively create portable data collection systems outside of the laboratory.

In this thesis, I use the LSL framework extensively to synchronize gaze, movement, and stimuli control signals. My approach is easy to implement, and results in high-quality synchronization of an arbitrary number of modalities. In the following studies, the data collection streams are programmed using the available language bindings for C# and Python.

1.5 Previous attempts of "in the wild" eye tracking and movement data collection

Depending on the specific task it is deployed for, eye tracking can be synonymous with gaze tracking; collecting information about where the user is looking. When studying natural behaviour, gaze can be used as a tool to gain insight into the inner cognitive workings of the brain. Hayhoe & Ballard [13] posit that one of the main roles of gaze is to guide where and when to fixate on targets so they can be interacted with. It

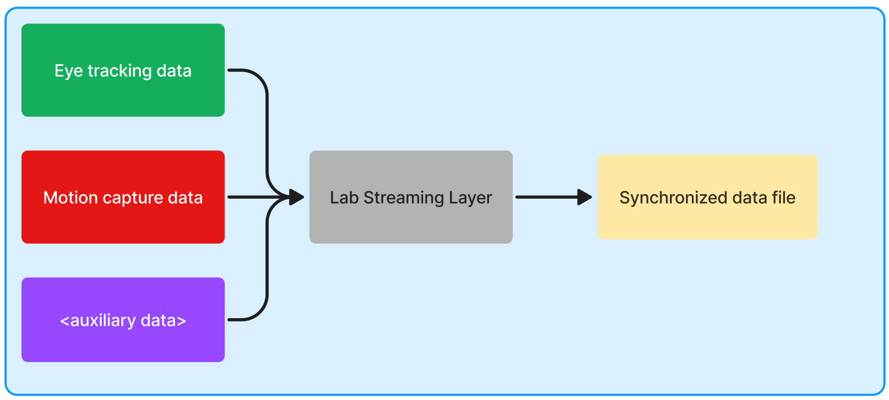


Figure 1.2: A high-level diagram demonstrating the core of Lab Streaming Layer's (LSL) functionality. Data on the left (eye tracking data, mouse data, and any number of other miscellaneous streams) are piped into LSL using an inlet function. LSL then timestamps the incoming data, and uses a drift correction algorithm to ensure the timing remains stable. The data is then put into a binary-encoded format which can be easily processed in many different environments.

is hard to separate cognitive goals from the eye movements that precede them; they are inextricably linked. Land *et al.* [67] found evidence that elements of cognitive planning were encoded in the eye movement signals. This enables the extrapolation and prediction of the eye movements necessary to perform tasks, ranging from basic, such as individual saccades, to intricate ones, such as making a sandwich or a cup of tea. Zooming in on the individual movements that make up the completion of a simple task such as sandwich making, we see that the eyes take on a clear assistive role for reaching. The role of the eyes in reach-to-grasp movements is well documented [35, 58, 68, 69], where the primary role of the eyes is to collect information that can be exploited. In general, the eyes collect information to create a motor plan that the hands execute. As mentioned earlier, there is a functional division between the two types of vision—vision-for-action (dorsal) and vision-for-perception (ventral) stream—in the primate cerebral cortex. To understand how the brain generates and utilizes motor plans, it is necessary to observe how we move in our environments.

But, moving outside of the laboratory is not an easy process. Recording eye movements “in the wild” (i.e. outside of the laboratory) is challenging, as many eye trackers lack portability and are difficult to use in non-lab settings. However, as stated earlier, many technological advances have made naturalistic data collection more feasible, and work needs to be done to show their usefulness in real-world settings. Previous research has shown that attention processes may differ between the lab and real-world environments [13, 70]. As a result, behavior in real-world settings may also differ. Foulsham *et al.* [70] conducted a study to determine if those who are immersed in the real world (e.g. out in public) pay attention to the same things as those who are not (e.g. those watching a video playback of the immersed person’s experience in the laboratory). This was an early study assessing the deployment of eye tracking outside of the lab, so much of the equipment was very bulky and required a large backpack to house the necessary equipment for data collection. An interesting finding came from how the participants tended to look at pedestrians, depending on how far away they were and if they were physically present. Those walking around outside (i.e. wearing the eye tracker) tended to look more at pedestrians further away rather than up close, and the opposite was found in the laboratory participants. Previous work has found that there are a multitude of reasons that humans will not look at others in certain social situations [1, 71], but when the social aspect is removed, humans are more likely to stare at others—something most people would find awkward. This work is a good example of why we collect real-world eye tracking data: human eye movement patterns are not the same between those in the real world, and those who are merely observing an approximation of the real world. Given that one of the primary roles of the eyes is to collect information for motor planning, it is reasonable that the resultant motor behaviour would also be modified. An important aspect then, is knowing how to design studies that capture these differences in humans.

Several experimental design approaches have been developed to promote more naturalistic behaviours during research, as reported by various studies [11, 12, 67, 72–74]. However, one particular approach stands out for its attempt to challenge the traditional methods and assumptions related to cognitive processes involved in common human behaviours. This approach is referred to as cognitive ethology [72, 73], and it questions some of the fundamental assumptions regarding cognitive process invariance while offering a framework for investigation. For example, are the same cognitive (and motor) processes used when someone reaches for a box of Kraft Dinner from their pantry as when they reach for an identical box but in a highly controlled laboratory experiment? The answer is not so clear, but some evidence suggests that even being watched performing a task can change your behaviour. An eye tracking study investigating eye contact during one-on-one interviews in-person or via a video feed showed that interviewer eye contact was able to modify the participant's viewing habits only in the in-person condition [75]. This suggests that there is something inherently different about an in-person interaction as compared to an artificial (i.e. video feed) interaction. These studies give a glimpse of the type of data that is being missed out on, suggesting that 'more natural' behaviours result in notably different cognition and eye behaviours. These studies provides credibility to cognitive ethology's prediction that lab-based studies may not generate completely accurate models which may be at best limited in predictive power, and at worst misleading. While this study may not be a perfect analog for comparing real life to laboratory studies, it provides evidence that there is, at the very least, value in studying naturalistic behaviours to augment current models and theories of attention.

Moving out of the laboratory requires us to re-evaluate many of the assumptions we make about data collection in general. Many cognitive scientists adhere to strict rules around the data collection environment itself; data must be collected in identical environments between participants with factors such as the lighting, auditory noise level, and the position of the participant in the room being controlled. The reasoning for this is simple: it reduces noise in the data. For some studies, this is critical, such as in electroencephalography (EEG) studies. EEG studies typically require quiet, isoluminant spaces to collect data as the sensors tend to be sensitive to this kind of environmental noise [76, 77]. But, why study these behaviours if they are not natural? Gramann *et al.* [78] argues that the primary function of the brain is to produce optimized motor outputs in dynamic environments, and understanding these natural dynamics of the brain necessitates a natural collection environment. In fact, a multitude of EEG studies have started to collect data in more real-world environments, challenging many of the original assumptions [78–83]. Taken together, these studies suggest that collecting (and analyzing) data known to be especially sensitive to noise in the real world is not only technically possible, but capable of being collected with current iterations of hardware. If even sensitive tools can collect data in real-world environments, why not more robust tools such as eye trackers and motion capture systems?

1.6 Aims and objectives

As I stated previously, moving outside of the lab is the best way to understand real, naturalistic, visuomotor human behaviour. But, it is fraught with the problems of losing experimental control, and most importantly for this thesis, the quality of the gaze and movement data you can collect. The primary aim of this thesis is to demonstrate that it is possible to deploy some of the tools visuomotor scientists use in real-world contexts to obtain data consistent with how humans actually behave.

Using eye-tracking and motion capture, in the following thesis I show the results of three distinct studies. These three studies all incorporate both eye and motion tracking in an attempt to quantify various behaviours. The studies are meant to be read in the context of a gradated movement from laboratory-based to the real world.

In the first study, I demonstrate that synchronized simultaneous collection of laboratory-grade eye- and body-tracking can be used to generate highly accurate 3D gaze vectors that have sub-centimetre accuracy in optimal conditions. Generating 3D gaze vectors is a challenging problem because there are no widely used established calibration routines. This study aimed to answer this question. To fully capture naturalistic behaviours, it was critical that the movements of the participant were unrestricted. We allowed free movement throughout the study using a portable eye tracker and motion capture. Gaze and body movements were recorded to determine where participants were looking and when they interacted with their environment. Participants completed four different calibration routines, where the pupil position data was used as an input to a model that predicts gaze position in 3D space. A validation procedure is used to assess model performance when the experimenters know exactly when and where the participant is supposed to be looking. Finally, participants complete a previously validated task known as the Pasta Task [14, 26, 84] to assess the model's performance on a real-world behaviour. Here, we assessed the performance of the resultant models generated form the calibration data and speculate on its use for future real-world tasks.

In the second study, I use a consumer-grade eye tracker and a mouse tracking to record the behaviour of participants as they navigate through a menu interface modeled after the popular video game, Mass Effect 3 [85]. This work was conducted as part of a Mitacs Accelerate internship with BioWare in Edmonton, Alberta. The original aim of this research was to identify significant performance metrics that could assist UX researchers at BioWare in enhancing user experience. In other words, I collected synchronized eye and mouse tracking data and analyzed the data to find when users were having trouble. In this field, typically qualitative measures such as interviews, surveys, and technical reports are used to uncover problematic areas of UX known as friction points. UX research is a field that benefits greatly from real-world data collection. I collected data from real users both in person as well as remotely over the internet, with little restrictions on their behaviours or computer setups. I use eye and hand tracking to attempt to quantify friction as a dynamic process. I demonstrate a methodology that allows for the detection of friction points based on gaze and movement signals. I show that many of the same patterns of friction that occur in the local cohort also occur in the remote cohort, suggesting that the increased noise (due to decreased experimental control) in the remote cohort did not overpower the signal.

In the third study, I attempted to move into the most challenging environment thus far: the clinic. The clinic can be a very chaotic space for recording high quality data. Here, I attempted to push the capabilities of eye tracking to its limit by collecting data from patients admitted to the University of Alberta hospital with a chief complaint of dizziness. Typically, neurological assessments are used to determine the etiology of the illness the patient is facing, which is performed in the clinic. If physicians are able to make accurate diagnoses in such a chaotic environment, perhaps a device designed to augment the diagnostic process should work in the same way. A portable device consisting of a laptop computer with a head-mounted eye tracker was used to present an easy-to-use stimulus to patients in a standard hospital bed. The stimulus was derived from a standard clinical neurological exam, and was designed to be easy to use by non-expert eye tracker users. Additionally, the output of the device was useful eye movement metrics that an attending physician could use to improve the efficacy of their clinical exam. Initially, we intended to collect reaching behaviour to be analyzed using markerless motion capture software, but this proved too difficult for the patients to complete. While the device did technically work, the patients were simply too sick to give data that was possible to analyze. The device was also tested on a control group (i.e. normative age-similar participants) to determine if the non-expert data collector biased the data quality. The control group gave high quality data, suggesting that such a device is feasible, but may be beyond the current limit of eye tracking technology for such a use case. Finally, I speculate on the value and potential use cases in clinical and non-clinical settings.

References

1. Laidlaw, K. E. W., Foulsham, T., Kuhn, G. & Kingstone, A. Potential social interactions are important to social attention. en. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 5548–5553. ISSN: 0027-8424, 1091-6490. http://dx.doi.org/10.1073/pnas.1017022108 (Apr. 2011).
2. Pickering, T. G. Clinical applications of ambulatory blood pressure monitoring: the white coat syndrome. en. *Clin. Invest. Med.* **14,** 212–217. ISSN: 0147-958X. https://www.ncbi.nlm.nih.gov/pubmed/1893653 (June 1991).
3. Pioli, M. R., Ritter, A. M., de Faria, A. P. & Modolo, R. White coat syndrome and its variations: differences and clinical impact. en. *Integr. Blood Press. Control* **11,** 73–79. ISSN: 1178-7104. http://dx.doi.org/10.2147/IBPC.S152761 (Nov. 2018).
4. Belletier, C. *et al.* Choking under monitoring pressure: being watched by the experimenter reduces executive attention. en. *Psychon. Bull. Rev.* **22,** 1410–1416. ISSN: 1069-9384, 1531-5320. http://dx.doi.org/10.3758/s13423-015-0804-9 (Oct. 2015).
5. Hayhoe, M. & Ballard, D. Eye movements in natural behavior. en. *Trends Cogn. Sci.* **9,** 188–194 (Apr. 2005).
6. Rayner, K. Eye movements and attention in reading, scene perception, and visual search. en. *Q. J. Exp. Psychol.* **62,** 1457–1506. ISSN: 1747-0218, 1747-0226. http://dx.doi.org/10.1080/17470210902816461 (Aug. 2009).
7. Eckstein, M. P. Visual search: a retrospective. en. *J. Vis.* **11.** ISSN: 1534-7362. http://dx.doi.org/10.1167/11.5.14 (Dec. 2011).
8. Hutton, S. B. Cognitive control of saccadic eye movements. en. *Brain Cogn.* **68,** 327–340. ISSN: 0278-2626, 1090-2147. http://dx.doi.org/10.1016/j.bandc.2008.08.021 (Dec. 2008).
9. Hayhoe, M. M. Advances in Relating Eye Movements and Cognition. en. *Infancy* **6,** 267–274. ISSN: 1525-0008, 1532-7078. http://dx.doi.org/10.1207/s15327078in0602_7 (Sept. 2004).
10. Hayhoe, M. M., Bensinger, D. G. & Ballard, D. H. Task constraints in visual working memory. en. *Vision Res.* **38,** 125–137. ISSN: 0042-6989. http://dx.doi.org/10.1016/s0042-6989(97)00116-8 (Jan. 1998).
11. Land, M. F. & Lee, D. N. Where we look when we steer. en. *Nature* **369,** 742–744. ISSN: 0028-0836. http://dx.doi.org/10.1038/369742a0 (June 1994).
12. Hayhoe, M. Vision Using Routines: A Functional Account of Vision. *Vis. cogn.* **7,** 43–64 (Jan. 2000).
13. Hayhoe, M. & Ballard, D. Eye movements in natural behavior. en. *Trends Cogn. Sci.* **9,** 188–194. ISSN: 1364-6613. http://dx.doi.org/10.1016/j.tics.2005.02.009 (Apr. 2005).

14. Lavoie, E. B. *et al.* Using synchronized eye and motion tracking to determine high-precision eye-movement patterns during object-interaction tasks. en. *J. Vis.* **18,** 18. ISSN: 1534-7362. http://dx.doi.org/10.1167/18.6.18 (June 2018).

15. Chapman, C. S. *et al.* Reaching for the unknown: multiple target encoding and real-time decision-making in a rapid reach task. en. *Cognition* **116,** 168–176. ISSN: 0010-0277, 1873-7838. http://dx.doi.org/10.1016/j.cognition.2010.04.008 (Aug. 2010).

16. Wispinski, N. J., Gallivan, J. P. & Chapman, C. S. Models, movements, and minds: bridging the gap between decision making and action. en. *Ann. N. Y. Acad. Sci.* ISSN: 0077-8923, 1749-6632. http://dx.doi.org/10.1111/nyas.13973 (Oct. 2018).

17. Wispinski, N. J. *et al.* Reaching for known unknowns: Rapid reach decisions accurately reflect the future state of dynamic probabilistic information. *Cortex* **138,** 253–265 (2021).

18. Gallivan, J. P., Chapman, C. S., Wolpert, D. M. & Flanagan, J. R. Decision-making in sensorimotor control. en. *Nat. Rev. Neurosci.* **19,** 519–534. ISSN: 1471-003X, 1471-0048. http://dx.doi.org/10.1038/s41583-018-0045-9 (Sept. 2018).

19. Gallivan, J. P. & Chapman, C. S. *Three-dimensional reach trajectories as a probe of real-time decision-making between multiple competing targets* 2014. http://dx.doi.org/10.3389/fnins.2014.00215.

20. Ouellette Zuk, A. A. *Going through the motions: Evaluating the impact of task, device and platform on mouse-tracking derived measures of decision difficulty* en. 2021. https://era.library.ualberta.ca/items/449f6cd6-ae12-4f9c-9e39-a0885f19b35b.

21. Freeman, J. B. & Ambady, N. MouseTracker: software for studying real-time mental processing using a computer mouse-tracking method. en. *Behav. Res. Methods* **42,** 226–241. ISSN: 1554-351X, 1554-3528. http://dx.doi.org/10.3758/BRM.42.1.226 (Feb. 2010).

22. Freeman, J. B., Dale, R. & Farmer, T. A. Hand in motion reveals mind in motion. en. *Front. Psychol.* **2,** 59. ISSN: 1664-1078. http://dx.doi.org/10.3389/fpsyg.2011.00059 (Apr. 2011).

23. Freeman, J. B. Doing psychological science by hand. en. *Curr. Dir. Psychol. Sci.* **27,** 315–323. ISSN: 0963-7214. http://dx.doi.org/10.1177/0963721417746793 (Oct. 2018).

24. Rheem, H., Verma, V. & Becker, D. V. Use of Mouse-tracking Method to Measure Cognitive Load. *Proc. Hum. Fact. Ergon. Soc. Annu. Meet.* **62,** 1982–1986. ISSN: 2169-5067. https://doi.org/10.1177/1541931218621449 (Sept. 2018).

25. Boser, Q. A. *et al.* Cluster-based upper body marker models for three-dimensional kinematic analysis: Comparison with an anatomical model and reliability analysis. en. *J. Biomech.* **72,** 228–234. ISSN: 0021-9290, 1873-2380. http://dx.doi.org/10.1016/j.jbiomech.2018.02.028 (Apr. 2018).
26. Valevicius, A. M. *et al.* Characterization of normative hand movements during two functional upper limb tasks. en. *PLoS One* **13,** e0199549 (June 2018).
27. Goodale, M. A. Transforming vision into action. en. *Vision Res.* **51,** 1567–1587. ISSN: 0042-6989, 1878-5646. http://dx.doi.org/10.1016/j.visres.2010.07.027 (July 2011).
28. Lashley, K. S. The mechanism of vision: XVIII. Effects of destroying the visual" associative areas" of the monkey. *Genet. Psychol. Monogr.* ISSN: 0016-6677. https://psycnet.apa.org/record/1949-01109-001 (1948).
29. Marr, D. & Hildreth, E. Theory of edge detection. en. *Proc. R. Soc. Lond. B Biol. Sci.* **207,** 187–217. ISSN: 0080-4649, 0950-1193. http://dx.doi.org/10.1098/rspb.1980.0020 (Feb. 1980).
30. Gregory, R. L. The intelligent eye. en. *Optom. Vis. Sci.* **48,** 851. ISSN: 1040-5488, 1538-9235. http://journals.lww.com/00006324-197110000-00016 (Oct. 1971).
31. Schneider, G. E. Two visual systems. en. *Science* **163,** 895–902. ISSN: 0036-8075. http://dx.doi.org/10.1126/science.163.3870.895 (Feb. 1969).
32. Ingle, D. Two visual systems in the frog. en. *Science* **181,** 1053–1055. ISSN: 0036-8075. http://dx.doi.org/10.1126/science.181.4104.1053 (Sept. 1973).
33. Mishkin, M. & Ungerleider, L. G. Contribution of striate inputs to the visuospatial functions of parieto-preoccipital cortex in monkeys. en. *Behav. Brain Res.* **6,** 57–77. ISSN: 0166-4328. http://dx.doi.org/10.1016/0166-4328(82)90081-x (Sept. 1982).
34. Mishkin, M., Lewis, M. E. & Ungerleider, L. G. Equivalence of parieto-preoccipital subareas for visuospatial ability in monkeys. en. *Behav. Brain Res.* **6,** 41–55. ISSN: 0166-4328. http://dx.doi.org/10.1016/0166-4328(82)90080-8 (Sept. 1982).
35. Goodale, M. A. & Milner, A. D. Separate visual pathways for perception and action. en. *Trends Neurosci.* **15,** 20–25. ISSN: 0166-2236. http://dx.doi.org/10.1016/0166-2236(92)90344-8 (Jan. 1992).
36. Milner, D. & Goodale, M. *The Visual Brain in Action* en. ISBN: 9780191545726. https://play.google.com/store/books/details?id=8CaQDwAAQBAJ (OUP Oxford, Oct. 2006).
37. Flanagan, J. R. & Johansson, R. S. Action plans used in action observation. en. *Nature* **424,** 769–771. ISSN: 0028-0836, 1476-4687. http://dx.doi.org/10.1038/nature01861 (Aug. 2003).

38. Zheng, B., Jiang, X., Bednarik, R. & Atkins, M. S. Action-related eye measures to assess surgical expertise. en. *BJS Open* **5.** ISSN: 2474-9842. http://dx.doi.org/10.1093/bjsopen/zrab068 (July 2021).

39. Gallivan, J. P. & Goodale, M. A. in *Handbook of Clinical Neurology* (eds Vallar, G. & Coslett, H. B.) 449–466 (Elsevier, Jan. 2018). https://www.sciencedirect.com/science/article/pii/B9780444636225000231.

40. Goodale, M. A., Króliczak, G. & Westwood, D. A. Dual routes to action: contributions of the dorsal and ventral streams to adaptive behavior. en. *Prog. Brain Res.* **149,** 269–283. ISSN: 0079-6123, 1875-7855. http://dx.doi.org/10.1016/S0079-6123(05)49019-6 (2005).

41. Chapman, C. S. & Goodale, M. A. Obstacle avoidance during online corrections. en. *J. Vis.* **10,** 17. ISSN: 1534-7362. http://dx.doi.org/10.1167/10.11.17 (Sept. 2010).

42. Dodge, R. & Cline, T. S. The angle velocity of eye movements. *Psychol. Rev.* **8,** 145–157. ISSN: 0033-295X, 1939-1471. https://psycnet.apa.org/fulltext/1926-02957-001.pdf (Mar. 1901).

43. Richardson, D. C. in *Encyclopedia of Biomaterials and Biomedical Engineering* 1060–1064 (CRC Press, May 2008). ISBN: 9780429154065. https://www.taylorfrancis.com/chapters/edit/10.1201/9780429154065-102/eye-tracking-characteristics-methods-daniel-richardson-michael-spivey.

44. Aserinsky, E. Effects of illumination and sleep upon amplitude of electro-oculogram. en. *AMA Arch. Ophthalmol.* **53,** 542–546. ISSN: 0096-6339. http://dx.doi.org/10.1001/archopht.1955.00930010550013 (Apr. 1954).

45. Jung, R. The registration of post-rotatory and optokinetic nystagmus and the optico-vestibular integration in humans. de. *Acta Otolaryngol.* **36,** 199–202. ISSN: 0001-6489. http://dx.doi.org/10.3109/00016484809123776 (May 1948).

46. Yarbus, A. L. Eye movements and vision Plenum. *New York.* ISSN: 0028-7369 (1967).

47. Tatler, B. W., Wade, N. J., Kwan, H., Findlay, J. M. & Velichkovsky, B. M. Yarbus, eye movements, and vision. en. *Iperception* **1,** 7–27. ISSN: 2041-6695. http://dx.doi.org/10.1068/i0382 (July 2010).

48. *Gazepoint* en. https://www.gazept.com. Accessed: 2022-7-27. Aug. 2013. https://www.gazept.com.

49. Wisiecka, K. *et al. Comparison of Webcam and Remote Eye Tracking* in *2022 Symposium on Eye Tracking Research and Applications* (Association for Computing Machinery, Seattle, WA, USA, June 2022), 1–7. ISBN: 9781450392525. https://doi.org/10.1145/3517031.3529615.

50. Kassner, M., Patera, W. & Bulling, A. *Pupil: an open source platform for pervasive eye tracking and mobile gaze-based interaction* in *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication* (Association for Computing Machinery, Seattle, Washington, Sept. 2014), 1151–1160.

51. Machıén, L. *et al.* The habitual nature of food purchases at the supermarket: Implications for policy making. en. *Appetite* **155,** 104844. ISSN: 0195-6663, 1095-8304. http://dx.doi.org/10.1016/j.appet.2020.104844 (Dec. 2020).

52. Muybridge, J. The Horse in Motion. en. *Nature* **25,** 605–605. ISSN: 0028-0836. https://www.nature.com/articles/025605b0 (Apr. 1882).

53. Ellard, C. G., Goodale, M. A. & Timney, B. Distance estimation in the Mongolian gerbil: the role of dynamic depth cues. en. *Behav. Brain Res.* **14,** 29–39. ISSN: 0166-4328. http://dx.doi.org/10.1016/0166-4328(84)90017-2 (Oct. 1984).

54. Smeets, J. B. & Brenner, E. A new view on grasping. en. *Motor Control* **3,** 237–271. ISSN: 1087-1640. http://dx.doi.org/10.1123/mcj.3.3.237 (July 1999).

55. Hommel, B. *et al.* No one knows what attention is. en. *Atten. Percept. Psychophys.* **81,** 2288–2303. ISSN: 1943-3921, 1943-393X. http://dx.doi.org/10.3758/s13414-019-01846-w (Oct. 2019).

56. Gallivan, J. P., Barton, K. S., Chapman, C. S., Wolpert, D. M. & Flanagan, J. R. Action plan co-optimization reveals the parallel encoding of competing reach movements. en. *Nat. Commun.* **6,** 7428. ISSN: 2041-1723. http://dx.doi.org/10.1038/ncomms8428 (July 2015).

57. Magnuson, J. S. Moving hand reveals dynamics of thought. en. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 9995–9996. ISSN: 0027-8424. http://dx.doi.org/10.1073/pnas.0504413102 (July 2005).

58. Goodale, M. A., Pelisson, D. & Prablanc, C. Large adjustments in visually guided reaching do not depend on vision of the hand or perception of target displacement. en. *Nature* **320,** 748–750. ISSN: 0028-0836. http://dx.doi.org/10.1038/320748a0 (1986).

59. Singhal, A., Monaco, S., Kaufman, L. D. & Culham, J. C. Human fMRI reveals that delayed action re-recruits visual perception. en. *PLoS One* **8,** e73629. ISSN: 1932-6203. http://dx.doi.org/10.1371/journal.pone.0073629 (Sept. 2013).

60. Mathis, A. *et al.* DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. en. *Nat. Neurosci.* **21,** 1281–1289. ISSN: 1097-6256, 1546-1726. http://dx.doi.org/10.1038/s41593-018-0209-y (Sept. 2018).

61. Lugaresi, C. *et al.* MediaPipe: A Framework for Building Perception Pipelines. arXiv: 1906.08172 `[cs.DC]`. http://arxiv.org/abs/1906.08172 (June 2019).

62. Song, J.-H. & Nakayama, K. Hidden cognitive states revealed in choice reaching tasks. en. *Trends Cogn. Sci.* **13,** 360–366. ISSN: 1364-6613. http://dx.doi.org/10.1016/j.tics.2009.04.009 (Aug. 2009).

63. Voudouris, D., Smeets, J. B. J. & Brenner, E. Fixation Biases towards the Index Finger in Almost-Natural Grasping. en. *PLoS One* **11,** e0146864. ISSN: 1932-6203. http://dx.doi.org/10.1371/journal.pone.0146864 (Jan. 2016).

64. Voudouris, D., Smeets, J. B. J., Fiehler, K. & Brenner, E. Gaze when reaching to grasp a glass. en. *J. Vis.* **18,** 16. ISSN: 1534-7362. http://dx.doi.org/10.1167/18.8.16 (Aug. 2018).

65. Williams, H. E., Chapman, C. S., Pilarski, P. M., Vette, A. H. & Hebert, J. S. Gaze and Movement Assessment (GaMA): Inter-site validation of a visuomotor upper limb functional protocol. en. *PLoS One* **14,** e0219333 (Dec. 2019).

66. SCCN. *Lab Streaming Layer* Mar. 2021.

67. Land, M., Mennie, N. & Rusted, J. The roles of vision and eye movements in the control of activities of daily living. en. *Perception* **28,** 1311–1328. ISSN: 0301-0066. http://dx.doi.org/10.1068/p2935 (1999).

68. Prime, S. L. & Marotta, J. J. Gaze strategies during visually-guided versus memory-guided grasping. en. *Exp. Brain Res.* **225,** 291–305. ISSN: 0014-4819, 1432-1106. http://dx.doi.org/10.1007/s00221-012-3358-3 (Mar. 2013).

69. Johansson, R. S., Westling, G., Bäckström, A. & Flanagan, J. R. Eye-hand coordination in object manipulation. en. *J. Neurosci.* **21,** 6917–6932. ISSN: 0270-6474, 1529-2401. https://www.ncbi.nlm.nih.gov/pubmed/11517279 (Sept. 2001).

70. Foulsham, T., Walker, E. & Kingstone, A. The where, what and when of gaze allocation in the lab and the natural environment. en. *Vision Res.* **51,** 1920–1931. ISSN: 0042-6989, 1878-5646. https://www.sciencedirect.com/science/article/pii/S0042698911002392 (Sept. 2011).

71. Goffman, E. *Behavior in Public Places* en. ISBN: 9781439108697. https://play.google.com/store/books/details?id=EM1NNzcR-V0C (Simon and Schuster, June 2008).

72. Kingstone, A., Smilek, D., Birmingham, E., *et al.* Cognitive ethology: Giving real life to attention research. *Measuring the mind,* 341–357. https://books.google.com/books?hl=en&lr=&id=qTwVDAAAQBAJ&oi=fnd&pg=PA341&dq=Cognitive+ethology+Giving+real+life+attention+research+Kingstone+Smilek&ots=PfAQeFSBX5&sig=jz8-sBMnA6KfFHKbkNnbE5TMx7Q (2005).

73. Kingstone, A., Smilek, D. & Eastwood, J. D. Cognitive Ethology: a new approach for studying human cognition. en. *Br. J. Psychol.* **99,** 317–340. ISSN: 0007-1269. http://dx.doi.org/10.1348/000712607X251243 (Aug. 2008).

74. Matthis, J. S., Yates, J. L. & Hayhoe, M. M. Gaze and the Control of Foot Placement When Walking in Natural Terrain. en. *Curr. Biol.* **28,** 1224–1233.e5. ISSN: 0960-9822, 1879-0445. http://dx.doi.org/10.1016/j.cub.2018.03.008 (Apr. 2018).

75. Freeth, M., Foulsham, T. & Kingstone, A. What affects social attention? Social presence, eye contact and autistic traits. en. *PLoS One* **8,** e53286. ISSN: 1932-6203. http://dx.doi.org/10.1371/journal.pone.0053286 (Jan. 2013).

76. Luck, S. J. *An Introduction to the Event-Related Potential Technique, second edition* en. ISBN: 9780262324069. https://play.google.com/store/books/details?id=y4-uAwAAQBAJ (MIT Press, June 2014).

77. Bigdely-Shamlo, N. *et al.* Hierarchical Event Descriptors (HED): Semi-Structured Tagging for Real-World Events in Large-Scale EEG. en. *Front. Neuroinform.* **10,** 42. ISSN: 1662-5196. http://dx.doi.org/10.3389/fninf.2016.00042 (Oct. 2016).

78. Gramann, K., Ferris, D. P., Gwin, J. & Makeig, S. Imaging natural cognition in action. en. *Int. J. Psychophysiol.* **91,** 22–29. ISSN: 0167-8760, 1872-7697. http://dx.doi.org/10.1016/j.ijpsycho.2013.09.003 (Jan. 2014).

79. Gramann, K., Jung, T.-P., Ferris, D. P., Lin, C.-T. & Makeig, S. Toward a new cognitive neuroscience: modeling natural brain dynamics. en. *Front. Hum. Neurosci.* **8,** 444. ISSN: 1662-5161. http://dx.doi.org/10.3389/fnhum.2014.00444 (June 2014).

80. Scanlon, J. E. M., Redman, E. X., Kuziek, J. W. P. & Mathewson, K. E. A ride in the park: Cycling in different outdoor environments modulates the auditory evoked potentials. en. *Int. J. Psychophysiol.* **151,** 59–69. ISSN: 0167-8760, 1872-7697. http://dx.doi.org/10.1016/j.ijpsycho.2020.02.016 (May 2020).

81. Robles, D. *et al. Surrounding Traffic Matters: Increases in Traffic Volume Are Related to Changes in EEG Rhythms in Urban Cyclists* en. May 2022. https://www.biorxiv.org/content/10.1101/2022.05.27.493782.

82. Jungnickel, E., Gehrke, L., Klug, M. & Gramann, K. in *Neuroergonomics* (eds Ayaz, H. & Dehais, F.) 59–63 (Academic Press, Jan. 2019). ISBN: 9780128119266. https://www.sciencedirect.com/science/article/pii/B9780128119266000105.

83. Scanlon, J. E. M., Jacobsen, N. S. J., Maack, M. C. & Debener, S. Does the electrode amplification style matter? A comparison of active and passive EEG system configurations during standing and walking. en. *Eur. J. Neurosci.* **54,** 8381–8395. ISSN: 0953-816X, 1460-9568. http://dx.doi.org/10.1111/ejn.15037 (Dec. 2021).

84. Williams, H. E., Chapman, C. S., Pilarski, P. M., Vette, A. H. & Hebert, J. S. Gaze and Movement Assessment (GaMA): Inter-site validation of a visuomotor upper limb functional protocol. en. *PLoS One* **14,** e0219333. ISSN: 1932-6203. http://dx.doi.org/10.1371/journal.pone.0219333 (Dec. 2019).

85. BioWare. *Mass Effect 3* Mar. 2012. https://www.bioware.com/games/#mass-effect-3.

Chapter 2

Generating accurate 3D gaze vectors using synchronized eye tracking and motion capture

A version of this work was previously published as: S.A. Stone, Boser, Q.A., Dawson T.R., Vette, A.H., Hebert, J.S., Pilarski, P.M., & Chapman, C.S. Generating accurate 3D gaze vectors using synchronized eye tracking and motion capture. *Behavior Research Methods*, doi:10.3758/s13428-022-01958-6.

Abstract

Assessing gaze behaviour during real-world tasks is difficult; dynamic bodies moving through dynamic worlds make gaze analysis difficult. Current approaches involve laborious coding of pupil positions. In settings where motion capture and mobile eye tracking are used concurrently in naturalistic tasks, it is critical that data collection be simple, efficient, and systematic. One solution is to combine eye tracking with motion capture to generate 3D gaze vectors. When combined with tracked or known object locations, 3D gaze vector generation can be automated. Here we use combined eye and motion capture and explore how linear regression models generate accurate 3D gaze vectors. We compare spatial accuracy of models derived from four short calibration routines across three pupil data inputs: the efficacy of calibration routines were assessed, a validation task requiring short fixations on task-relevant locations, and a naturalistic object interaction task to bridge the gap between laboratory and “in the wild” studies. Further, we generated and compared models using spherical and cartesian coordinate systems and monocular (Left or Right) or binocular data. All calibration routines performed similarly, with the best performance (i.e., sub-centimetre errors) coming from the naturalistic task trials when the participant is looking at an object in front of them. We found that spherical coordinate systems generate the most accurate gaze vectors with no differences in accuracy when using monocular or binocular data. Overall, we recommend one-minute calibration routines using binocular pupil data combined with a spherical world coordinate system to produce the highest quality gaze vectors.

2.1 Introduction

The majority of laboratory examinations of eye gaze are highly constrained and reliant on the assumption that gaze behaviors are task-invariant [86]. That is, many laboratory tasks do not reflect naturalistic behaviours. Common sense says that where someone is looking is dependent upon both eye and head movements [87, 88], meaning head position must be accounted for when calculating and analyzing gaze. Most studies investigating hand-eye coordination circumvent this problem by restricting head movements through the use of a chin rest [89]. In the real world, we are free to gaze at objects throughout our full field of view, or even anywhere in our 3D space, provided we can turn and move. But, in the lab, the areas the participant can interact with are typically severely limited, such as restricting gaze to a computer monitor or tabletop [90, 91]. Controlling for such environmental variables lets researchers ask specific questions about the motor and neural mechanisms that govern hand-eye coordination but fail to ask how gaze performs in natural settings. When collecting data outside of the laboratory, it is simply not feasible nor ecologically valid to restrict movement of the head or restrict gaze to the interaction with a limited amount of space. Additionally, real-world data tends to be much more difficult to process and analyze because of the permissive setting in which it is collected; free movement of the body is encouraged, as it more closely reflects natural behaviour.

Collecting data outside of the laboratory—or "in the wild"—is challenging [89]; determining fixations from dynamic bodies moving through dynamic worlds is a non-trivial problem to solve. A few studies have collected data while performing simple every-day activities [12, 26, 92–94]. For example, Land & Hayhoe [92] found that eye behaviours were similar across different use cases, such as during making a cup of tea or preparing a sandwich. They found eye movements could be broken down into four systematic categories: locating (the target), directing (the hands to the target), guiding (the hands during movement), and checking (if the condition has been satisfied). These general rules of interaction help inform us of potential systematic analyses that can be performed on the data. Data recorded "in the wild" also tend to be harder to parse into fractional chunks for analysis; Lappi [89] describes some of these common issues when collecting real-world natural gaze behaviours. In his review, Lappi suggests that complex eye movement behaviours are built from combinations of primitive eye behaviours such as fixations, saccades, and pursuits. These primitive building blocks can be used as indices to break complex tasks into digestible blocks that can be analyzed more similarly to controlled lab-based experiments.

Over the last decade eye tracking technology has become cheaper and easier to use. Traditional eye tracking headsets tended to be bulkier and required the head position to be fixed, whereas newer eye trackers such as the Pupil Labs Core [50] are more portable and do not require a fixed head. One common consideration of designing an eye tracking study is the time-consuming manual labour required for cleaning and analysis [26, 65, 94, 95]. Much of this manual labour is centred around two primarily video based categorization steps: 1) the cleaning of the pupil data, most of which is difficult to automate because of the nature of data quality from individual participants and 2) the assignment of fixations to objects in the world on the "world camera", an outward facing camera attached to a head mounted eye tracker. This portion of analysis is so time consuming that many researchers will only analyze a subset of data rather than the whole [96–98]. For example, Parr *et al.* [98] were only able to analyze every third trial of their prosthetic hand-eye coordination task. A major concern is that a subset of data does not always represent the population-level statistics of the entire dataset—effects could be driven by outliers. Secondly, this leaves open the possibility of incorrect coding, leading to lower quality data that may contain additional errors, influencing statistical tests. Optimizing the volume of data analysis possible would have great benefits for statistical power and data reliability.

Motion capture (mo-cap) is a technique used to record human movement in 3D space [99] and, when combined and synchronized with eye tracking, offers a solution for automating real-world gaze analysis. Human movement science greatly benefits from this technology, as it allows for the quantification of movements during reach-to-grasp [17, 100] or reach-to-point [101–103] behaviours. Additionally, mo-cap technology comes in many forms, including infrared-based or the burgeoning field of markerless-motion-capture [104], both of which are typically capable of integrating with eye tracking headsets. Tracking gaze during movement grants insight into the strategies that different populations may use when completing the same task. For example, research into gaze strategies during reach-to-grasp behaviours has uncovered key strategic differences in normative [95] versus prosthetic arm users [65, 94, 105], where prosthesis users tend to move much slower, fixate longer, and do not "look ahead" to the intended target location after grasping.

While the synchronized collection of eye tracking and mo-cap data is not trivial, tools such as Lab Streaming Layer (LSL; SCCN [106]) have made this process much easier. However, once you have two synchronized data streams, it is not easy to determine where someone is looking based on raw data. Here we explore a technique requiring the experimenter to collect a separate eye-calibration data file, specifically for the purposes of building a model that will map head and pupil positions to a three dimensional (3D) gaze vector in a common world coordinate system. A 3D gaze vector is a line that extends from the head out into 3D space to predict where the participant is looking in world-space [107–109]. During these eye-calibration trials, participants are asked to focus on a tracked mo-cap marker (in our task on the tip of a calibration "wand") as it moves through space, typically for about a minute. In the 2D eye tracking space, there does exist some guidelines and recommendations to generate gaze points. However, to our knowledge, despite the increasing number of studies that use 3D gaze vectors to assess behaviour, no standardized and very few recommended calibration routines exists. That is, how should you best move the tracked "wand"-marker through space? In addition, what data should be used to build predictive models, including which coordinate frame(s) or binocular / monocular data, depending on how pupil data were recorded.

With the goal of providing researchers interested in naturalistic tasks recommendations and guidelines for expected accuracy, we generated and assessed 3D gaze vector models from all possible combinations of: four different calibration routines, two coordinate frames, and three sets of pupil data inputs. Our results describe an approach that is capable of generating accurate (sub-centimetre and below one visual degree in the best case) 3D gaze vectors (GVs) using the position of the pupils and the 3D location of the participant's head in space. To create the GVs, we use a linear regression algorithm to train models based on input pupil positions time-synchronized to the 3D location of a calibration wand. Then, we assess their spatial accuracy across a variety of data sets.

2.2 Methods

2.2.1 Equipment

Eye tracking data were collected using a Pupil Labs Core (200Hz; [50]) USB eye tracking headset. Lab Streaming Layer (LSL; [106]) was used to synchronize eye tracking and mo-cap data. The official Pupil Labs LSL plugin was used in conjunction with the Pupil Capture software to directly send data into the LSL datastream. Mo-cap data were collected using an OptiTrack mo-cap system (two systems were used throughout the study as the lab was upgraded: initially a 12-camera Flex 13 system, 120Hz; then a 14-camera Prime 13-W system, 200Hz). The OptiTrack systems were calibrated using the included Motive program to have a spatial accuracy of 0.1mm or less. A custom program was written in C# to pass frame data from the OptiTrack Motive application to the LSL datastream for synchronization. Rigid clusters of reflective markers were fixed to the participant and objects in the environment to track the position and orientation of the Head, Right Hand (centred approximately dorsally), Task Cart, Side Cart, Pasta Box (in Task data), and a Calibration Wand (in calibration data). Marker clusters were also fixed to the participant's pelvis, trunk, upper arms, forearms, and left hand in as described by Boser *et al.* [110], but these data were not used in the current study. It is worth noting that theoretically any combination of eye tracker and mo-cap system could be used, provided they collect time series data as synchronized 2D pupil positions (in eye camera coordinates) and 3D marker position (in mo-cap).

2.2.2 Participants

Twenty-one undergraduate and graduate students from the Department of Psychology research pool at the University of Alberta participated in this study. All participants were right-handed, had normal or corrected-to-normal vision, and were naive to the tasks. Eight participants were collected using the OptiTrack Flex 13 system at 120 Hz, and 13 were collected on the OptiTrack Prime 13-W system at 200 Hz. One participant was removed due to recording errors (poor tracking quality), for a total of twenty participants. This study was approved by the University of Alberta Health Research Ethics Board under protocol Pro00087329 and ethical protocols were in adherence to the 1964 Declaration of Helsinki.

2.2.3 Procedure

Each test of data quality consisted of 3 sets of Calibration/Validation trials and 2 sets of 10 Task trials, proceeding in the following order:

1. Calibration/Validation set
2. Task set
3. Calibration/Validation set
4. Task set
5. Calibration/Validation set

Each Calibration/Validation set included four Calibration trials (one of each type described below) and one Validation trial presented in a pseudo-random order. Each Task set included 10 repetitions of the previously published Pasta Box task (see [26] for a full description of the task parameters). In short, the Pasta Box task requires the participant to move a rectangular box of pasta between three key locations: the Side Cart, the Green shelf, and the Blue shelf. In between each of the reaches, the participant must touch the Home position (see Fig. 2.4 and section 2.2.3 for a visual representation of the task and relevant spaces). Each trial takes approximately 15 seconds to complete. In total, participants performed 12 Calibration trials (3 repetitions of each of 4 types), 3 Validation trials and 20 Task trials. Not all participants had usable data for every trials; we discuss dealing with missing data and removal in section 2.2.4.

Calibration Trials

Participants were asked to track the position of a single spherical mo-cap marker (14 mm diameter) with their eyes for about one minute per trial. The participant could move their head freely while tracking the marker. The marker was placed at the tip of a 40 cm wand which moved through the task space in one of four Calibration routines:

1. *Experimenter Sweep (ES):* The experimenter moved the wand in slow S-shaped curves along each of the room-coordinate axes (parallel to floor, left/right, parallel to floor in/out, parallel to wall up/down).

2. *Self Sweep (SS):* Replicating ES but with the participant holding the wand and replicating the movements.

3. *Experimenter Paint (EP):* The experimenter moved the wand to each of the relevant locations in the Pasta Box task (minus Neutral, see below) and explored small (10-20 cm in each dimension) volumes at these locations.

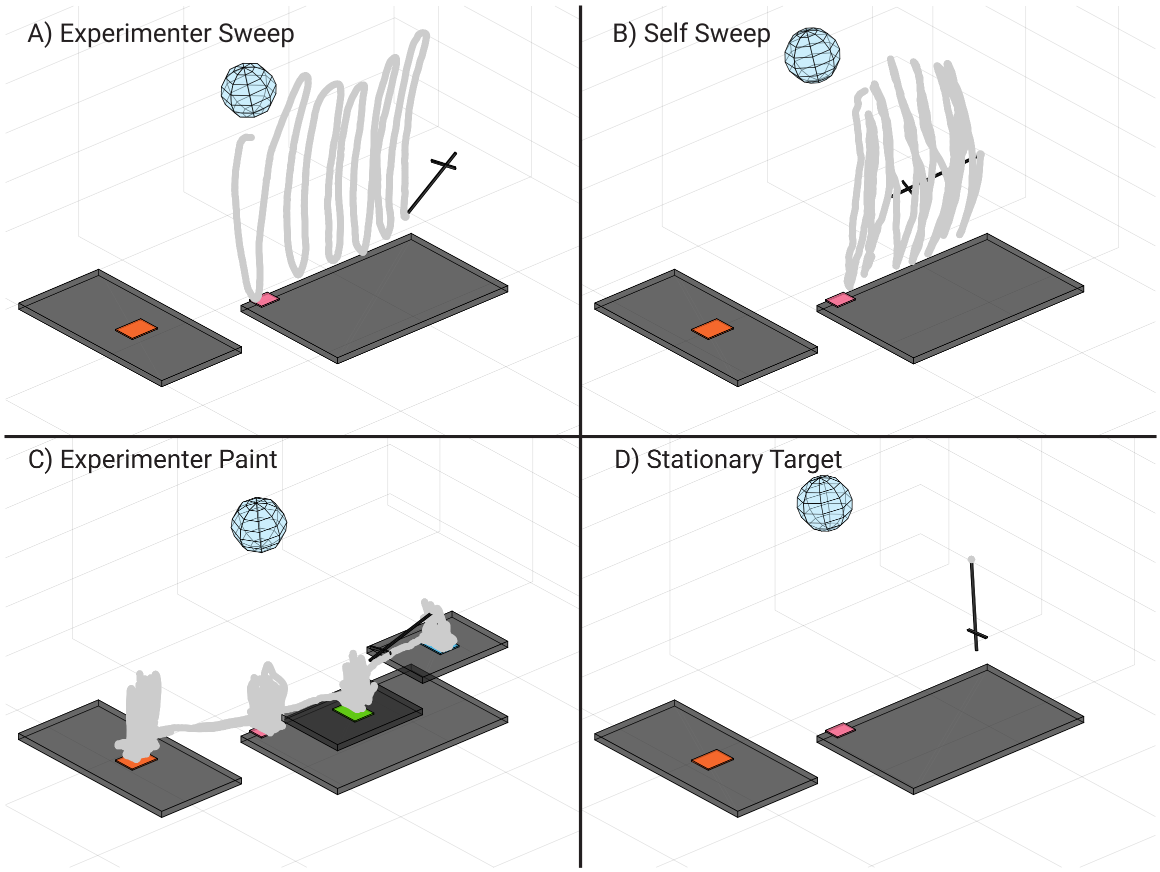


Figure 2.1: The Calibration routines used in the present study, with traces in grey showing example wand movements over time. Each routine takes approximately one minute to complete. The black inverted 't' shaped object is the calibration wand used in all routines. In all quadrants, the blue sphere represents the participant's head position, with the Orange target to the participant's right and the wand being directly in front of the participant. A) The Experimenter Sweep (ES) routine. The experimenter stands to the participant's left and waves the wand in s-shaped patterns through space, covering all three dimensions roughly equally (only up/down movements shown in figure). B) The Self Sweep (SS) routine. Identical in procedure to the ES routine, but the participant themselves carry out the wand movements. C) The Experimenter Paint (EP) routine. The experimenter stands to the right of the participant and moves the wand for approximately 15 seconds in small volumes at four locations relevant to the later Task trials: the Side Cart, the Home position, the Green Shelf and the Blue Shelf. D) The Stationary Target (ST) routine. The participant locks their gaze on the wand, which is fixed to the table. The participant moves their head up, then down, then centers, then left, then right (i.e., in the form of a cross), then rotates their head in swirl-like motions while maintaining fixation on the tip of the wand.

4. *Stationary Target (ST):* The wand was fixed to the table directly in front of the participant (~60 cm away), who was asked to maintain fixation on the wand-tip while nodding their head up and down, returning to centre, then turning it left and right, then rotating it in a clockwise then counterclockwise spiral.

The intention for each of these trials was to create calibration routines with a diversity of different coverages in terms of both task and pupil-position space (see Fig. 2.1 for the wand movements, and Fig. 2.2 for example corresponding pupil positions).

Validation Trials

Participants were asked to fixate on 5 stationary targets (see Fig. 2.4 for locations) presented at Task-relevant locations for ~5 s, in a specific sequence, and at least 2 times each. An auditory beep signalled the start of the first fixation and beeped every 5 seconds thereafter to signal a switch to the next Task-relevant location in this order:

Neutral → Side Cart → Blue Shelf → Home → Blue Shelf → Green Shelf → Home → Green Shelf → Side Cart → Home → Neutral.

This order of 11 fixations mirrors the order these locations are visited during the actual Task trials.

Task Trials

The set-up for the Pasta Box task is shown in Fig. 2.3. Participants began each Task trial with their hand on the Home position and their eyes fixating on the Neutral target, marked by a mo-cap marker. A beep then cued them to initiate an object interaction sequence consisting of three movements:

1. Reach and grasp the Pasta Box at the Side Cart, move it to Green Shelf then return hand to Home;

2. Reach and grasp the Pasta Box at Green Shelf, move it to Blue Shelf then return the hand to Home;

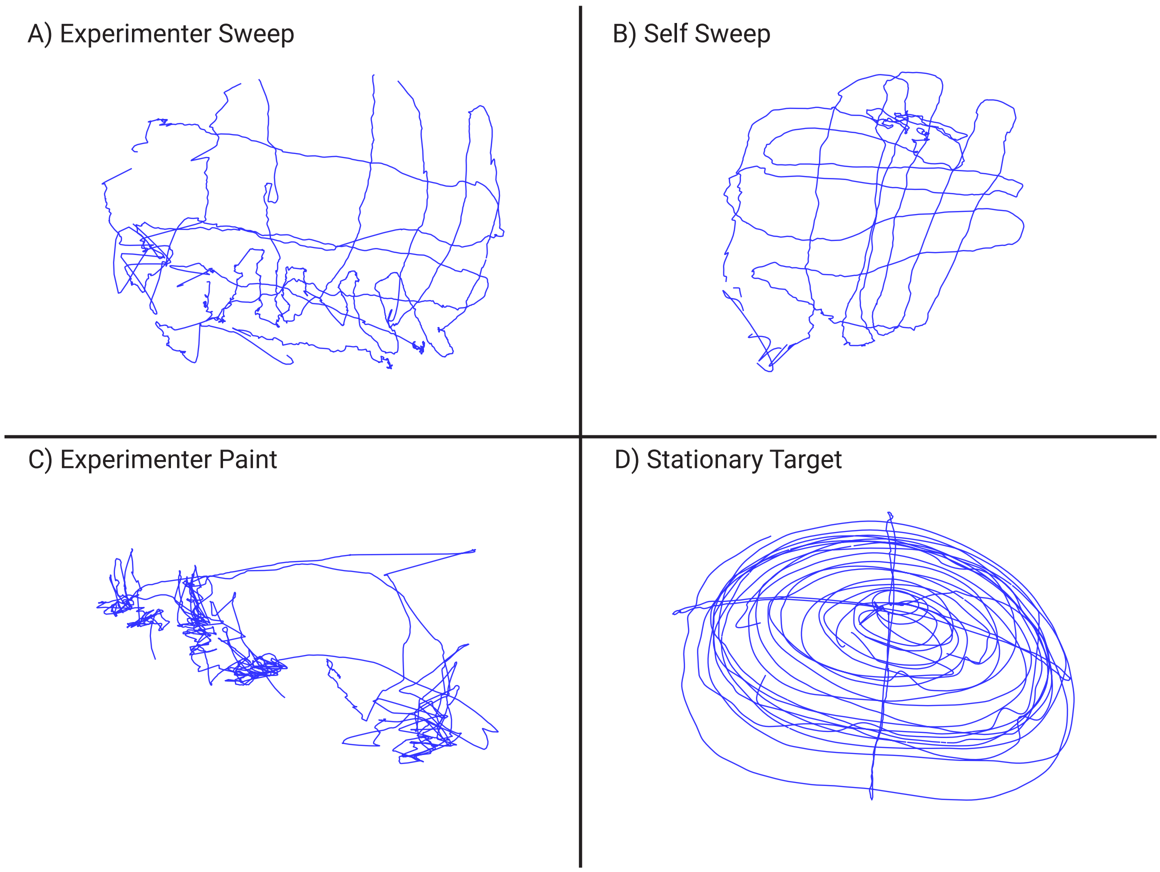


Figure 2.2: Example corresponding gaze patterns associated with each of the calibration routines. Pupil position from one eye is shown over the course of the entire calibration. A) The Experimenter Sweep (ES) routine: the gaze seems to be slightly jittery because the participant has to constantly adjust to the experimenter's wand position. B) The Self Sweep (SS) routine: the gaze pattern is much more smooth, because the participant is moving the wand while simultaneously fixating on the tip. C) The Experimenter Paint (EP) routine: gaze locks to four different locations, which slightly overlap because the participant was free to move their head and likely tends toward central fixation on each location. D) The Stationary Target (ST) routine: the head is moved in a cross-like movement (up, down, centre, left, right) then in swirl-like movements for approximately one minute.

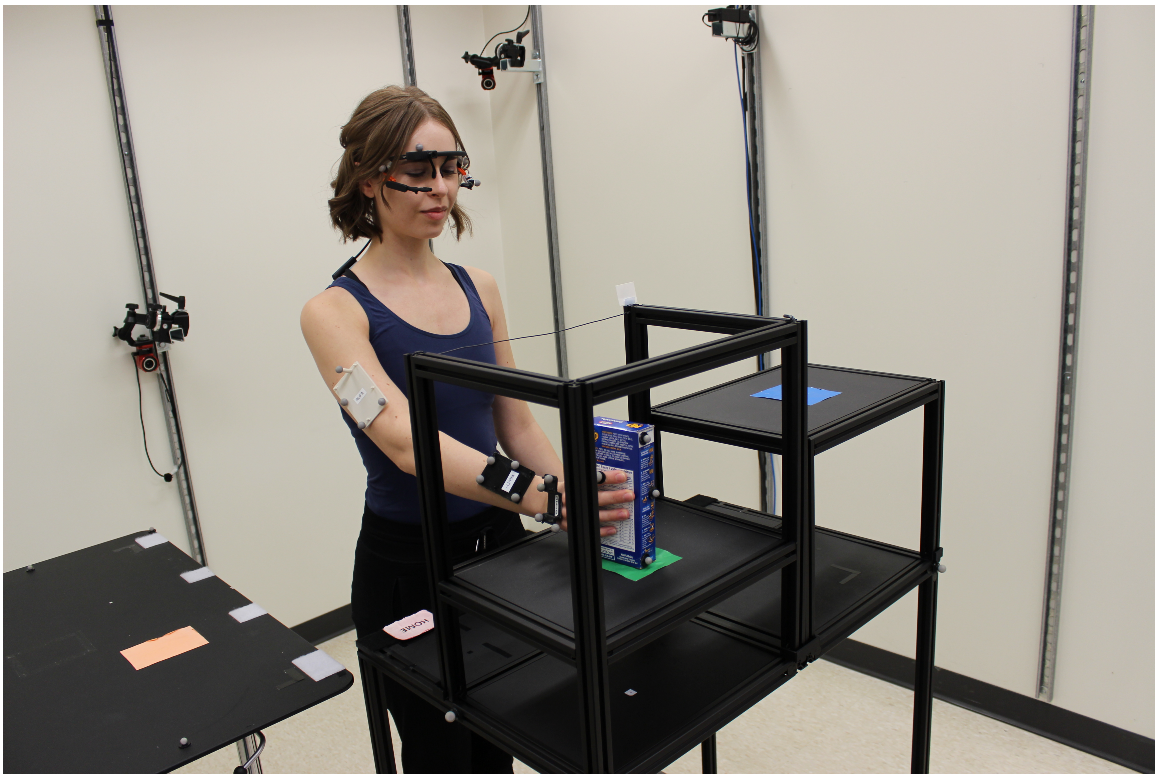


Figure 2.3: A still shot of a participant midway through completing one trial of the Pasta Task. From left-to-right there are four coloured rectangular markers which are the Side Cart (orange), Home (pink), Green Shelf (green), and Blue Shelf (blue) locations respectively. For each trial, the participant picks up the pasta box on the Side Cart, and moves it to the Green Shelf, then the Blue Shelf, and finally back to the Side Cart. Between each grasp, the participant touches the Home position with their right hand. Note that the participant is wearing mo-cap marker plates on their body, but for the purposes of the present study, only the right hand markers were used for pre-processing and analysis.

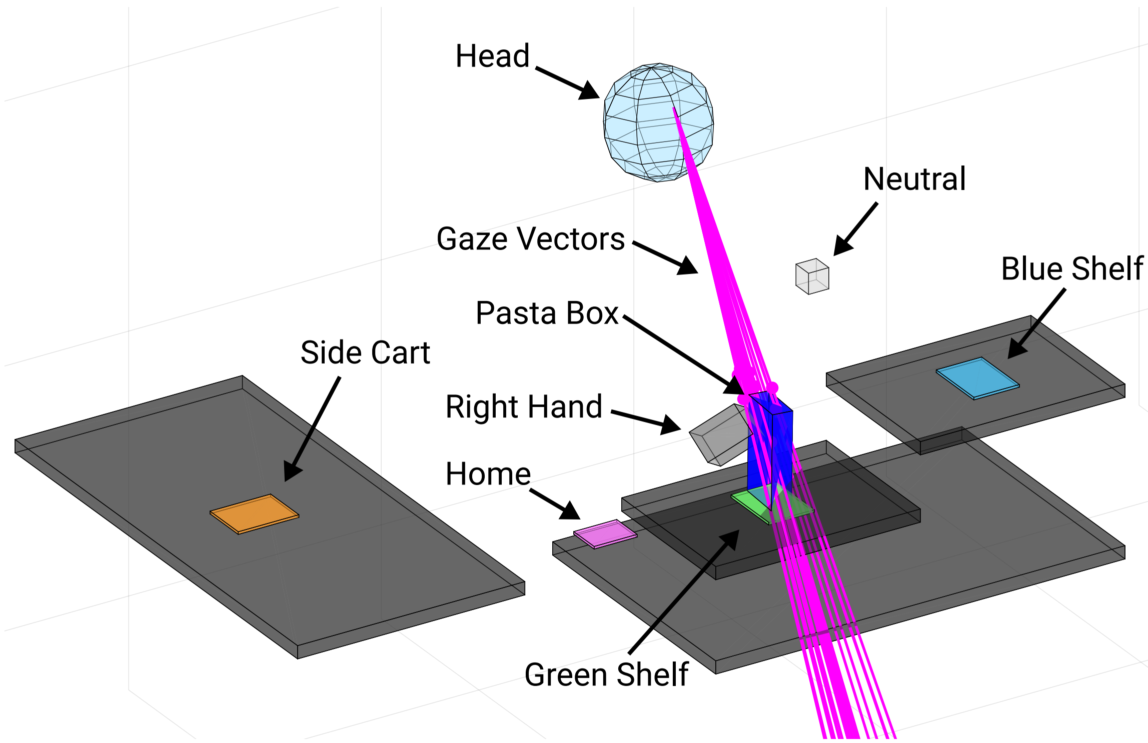


Figure 2.4: The locations, objects, and markers critical for all three tasks. The five locations are shown: Neutral, Side Cart, Home, Green Shelf, and Blue Shelf. For the Pasta Box task, the participant moved the box from location to location (see 2.2.3). The Head rigid body was used to determine the origin of the resulting gaze vectors. The Right Hand's velocity profile was used to determine when the participant picked up or dropped off the pasta box (see 2.2.4). All 72 gaze vectors generated are shown in pink, with most being close to the target object (pasta box), and some performing rather poorly.

3. Reach and grasp the Pasta Box at Blue Shelf, move it to the Side Cart then return the hand to Home. At the end of the task the participant also returns their gaze to the neutral marker.

The task was demonstrated to each participant visually. The participant was given as many practice trials as they felt necessary to be comfortable with the sequence of movements.

2.2.4 Data Processing

Pre-processing

Mo-cap data were exported from Motive and run through custom MATLAB scripts to check for marker name consistency and remove residual sections of noisy data (marker displacements of more than 5 mm between frames, and islands of data less than 100 ms in duration). Mo-cap and eye tracking data were then synchronized to the mo-cap frame rate using the common timestamps in the LSL datastream files. The combined data were imported into our custom software platform for integrated analysis of eye and mo-cap data; the Gaze and Movement Assessment Tool (GaMA; [65]). Within GaMA, raw pupil position data was cleaned by: 1) Removing any data points outside of pupil camera bounds (<0 or >1); 2) Removing any data points more than 4 standard deviations away from the mean position; 3) Removing any data points with velocities greater than 6 (meaning the pupil was travelling across the entire camera 6 or more times per second). After this removal, any gaps < 50 ms were filled using the *inpaint_nans* [111] function in MATLAB then, any remaining islands of data < 50 ms were deleted. Finally, the pupil data were filtered in MATLAB using a 4^{th} order zero-lag low-pass Butterworth filter with a cutoff frequency of 10 Hz. A 10 Hz cutoff was chosen because the demands of the tasks do not depend on eye dynamics with movements more than 10 times / second. Also within GaMA, the mo-cap data were filtered using a 4^{th} order zero-lag low-pass Butterworth filter with a cutoff frequency of 6 Hz. Rigid bodies, represented as both a position and rotation, were defined using the clusters of markers attached to the participant's head and hand, as well as objects in the environment. For the Task trial data, virtual objects were also created to represent the position, orientation and extent of the objects in the environment (Task Cart, Side Cart, Pasta Box).

Gaze Vector Modelling

The cleaned eye and motion data were then used to generate predictions of the direction the participant was looking in 3D space, or "gaze-in-world" vectors, herein referred to as GVs. The process of generating a single GV consists of two steps:

1. Generate eye gaze models using data from a specific Calibration trial
2. Use the eye gaze models to predict the GV direction at each frame in a given trial

In step 1, Calibration data are used to fit three eye gaze models. Each model takes pupil position data as input and predicts a single coordinate of the 3D gaze fixation point relative to the Head rigid body coordinate system in the 3D mo-cap space. For example: one model might use pupil position data to predict only the x-coordinate of the fixation point relative to the head, a second, separate model would be used to predict only the y-coordinate, etc. Each eye gaze model was generated using the built-in MATLAB function `fitlm` with the `'quadratic'` model specification and robust fitting using the `'bisquare'` weight function.

In this study we explored three options for model input (pupil input data from right eye only $[x_r,y_r]$, left eye only $[x_l,y_l]$, or binocular data $[x_l,y_l,x_r,y_r]$), as well as two options for expressing the fixation point relative to the Head coordinate system (Cartesian [x,y,z] coordinates, or Spherical $[r, \theta, \phi]$ coordinates). We anticipated that using the Spherical coordinate system would increase accuracy of the GV direction because it isolates depth of fixation to the 'r' model, whereas in Cartesian, all three models are influenced by depth of fixation.

In step 2, once the eye gaze models were generated for a given Calibration trial and set of parameters (left/right/both eyes $\times$ cartesian / spherical coordinate system), they were used to predict the coordinates of the fixation point relative to the head at each frame in a given Calibration, Validation, or Task trial. The known transformation between the Head rigid body coordinate system and global mo-cap coordinate system is then used to calculate the position of the fixation point relative to the global coordinate system. The GV is represented by the line originating at the head rigid body origin (mid forehead), passing through the fixation point, extending infinitely forward and away from the head in the direction of the fixation point (see Fig 2.5). It is important to note that only the direction of the GV was used in subsequent analysis, the distance from the head to the predicted fixation point was not considered.

Dependent Measures

We conducted separate analyses to assess GV accuracy in each of the three types of data collected (Calibration, Validation and Task). For each analysis, we collapsed across the 3 repetitions of a given Calibration type by finding which of the repetitions performed the "best" on that data. This involved eliminating abnormally poor GVs (those whose average distance from the target of analysis were 30 cm or more away), then taking the remaining GV with the lowest average distance to targets (see below). One participant was removed from the Calibration dataset, three from the Validation dataset, and two from the Task dataset because of average errors above 30cm. An advantage of this approach was this it allowed participants to be included for analysis that may have had errors in recording one repetition of Calibration data. As linear distance error does not account for the perspective of the participant, we also calculated the visual angle error simultaneously for each trial. The visual angle error accounts for the distance between the subject's eyes and the target object.

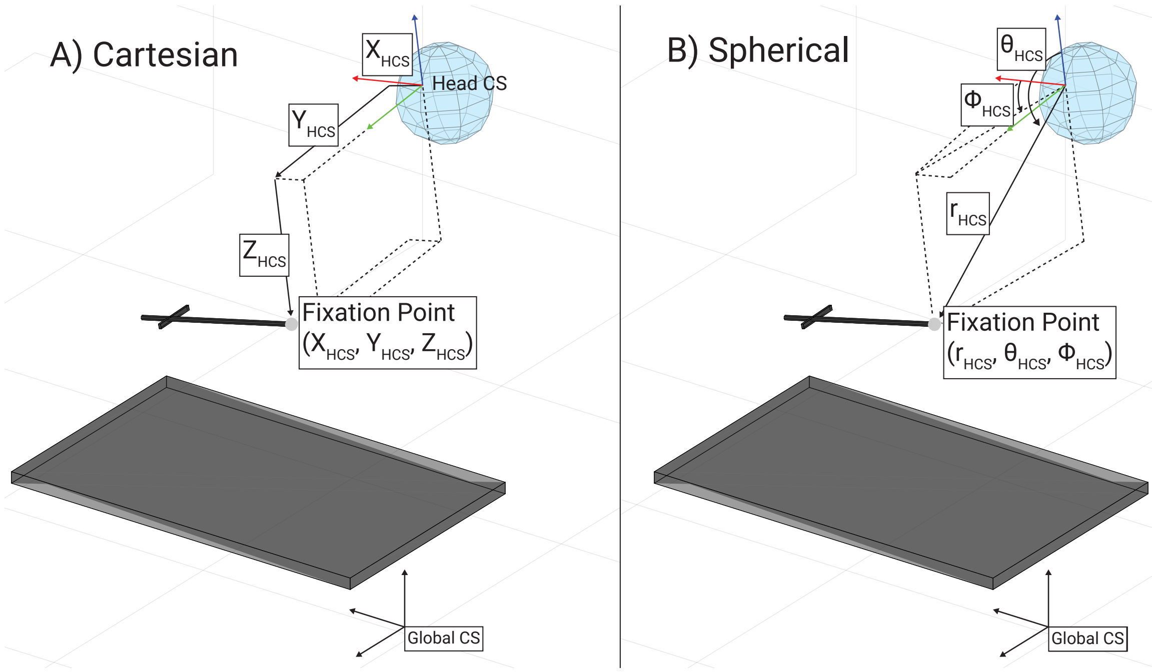


Figure 2.5: A visual demonstration of the differences between the Cartesian and Spherical coordinate systems used. The tip of the black wand is the gaze fixation point in both coordinate systems. A) The Cartesian coordinate system: coordinates are represented by coordinate triplets of [x,y,z]. Here, the wand tip is only represented by its offset from the origin. A consequence is that the depth of the wand is embedded in all of the dimensions. B) The Spherical coordinate system: coordinates are encoded as triplets of [r,θ, ϕ]. Here, the wand tip is described in terms of the angular and euclidean distance from the head. As a consequence, euclidean distance (i.e., depth) can be sequestered to a single coordinate. As our goals did not require depth estimates, we were able to train our models without the depth information.

Calibration Trials The dependent measure for Calibration trials was the mean 3D distance between a given GV and the Wand Tip over the entire trial. To reiterate from above, distances were always calculated as the minimum distance between the GV line and the Wand Tip point, meaning the depth of fixation along the GV was not a determinant of accuracy.

For each of the 12 Calibration trials we generated all 72 possible GVs (12 Calibration files $\times$ 2 Coordinate systems $\times$ 3 pupil data inputs). For each coordinate system and input data combination, we compared the three repetitions of GVs of the same type (e.g., across the 3 ES GVs that were Cartesian and with Both eyes) across all 12 Calibration trials to find the one that performed the best. Note that trials were not tested on the data used to train the model. First, we eliminated GV outliers (those with GV to Wand Tip distances $>$ 30 cm), then we took the median performance for each of the 3 repetitions across the remaining Calibration trials. The repetition with the lowest median performance was then selected as the Best and used for the remaining analyses.

Validation Trials The dependent measure for Validation trials was the mean 3D distance between a GV and each of the Task Relevant target locations. To extract these Task Relevant target looks, we isolated 1 second epochs of stable-gaze data in the 5 s between the cueing-beeps. For example, between the first and second beep, participants were instructed to look at the Neutral target. Within this 5 s window, we use a modified moving mean algorithm to find the 1 s of data where 1) there was at least 50% of detected pupil and 2) the L or R pupil data has the lowest velocity. This process generates 11 stable-gaze epochs, three for the Home location and two each for the remaining four Task-relevant locations. For each of the five Task-relevant locations, the reported distance is the mean over these stable-gaze epochs.

Similar to the Calibration trials, but across the 3 Validation trials, we selected the best GV within a set of repetitions by comparing their performance across the 5 locations. First, we eliminated outliers (over 30 cm mean distance from any location), then took the one with the minimum median distance across all Validation trials and the five locations as the best GV.

Task Trials The dependent measure for the Task trials was the mean 3D distance between a given GV and the nearest bounding box face of the Neutral (4 cm cube) or Pasta Box (9 x 4 x 18 cm; see Fig. 2.4) object at specific locations and times during the interaction task. Eye gaze behaviour is well understood for this task, as described in Lavoie *et al.* [94] and Williams *et al.* [65]. Following the same procedure as in this earlier work, each task trial was segmented into specific movements and movement-phases (Reach, Grasp, Transport and Release) using detailed procedures described elsewhere [94]. For this analysis, we isolated looks toward the Neutral marker at the start of the trial and looks toward the Pasta Box each time it was being grasped (just prior to object pickup) and released (just after object dropoff). Previous work using this identical task shows that there are fixations to these objects around these times on almost every trial [94]. These were single frame events that occurred once (for the look to Neutral) or twice (for the looks toward the Pasta Box at the Side Cart, Green Shelf and Blue Shelf locations) per location. Distances to locations with two looks were averaged.

Similar to the Calibration and Validation trials, across the 20 possible Task trials we selected the best GV within a set of repetitions by comparing their performance across the 4 locations (note no interactions occurred at the Home location so it was not included in the Task trial analysis). First, we eliminated outliers (over 30 cm mean distance from any object), then took as the best GV the one with the minimum median distance across all Task trials and the four locations.

2.3 Results

Statistical analysis was performed in JASP 0.16.1 [112]. Repeated-measures ANOVAs (rmANOVAs) were used to analyze the three trial types, which used the same participant pool but were statistically independent from one another. We conducted statistical analysis on two independent sets of data: a linear distance error (centimetres) and a visual angle error (degrees) to account for distance. All results below are reported with the linear distance error (LD) first and visual angle error (VA) second. We opted to use a conservative statistical approach, correcting α for the number of tests run in each family as described by Cramer *et al.* [113]. Each of the trial types were considered a family for this analysis. All p values were Greenhouse-Geisser corrected if sphericity was violated and more than two levels existed in the factor.

2.3.1 Factors for rmANOVA

For clarity, here we lay out all of the factors and their levels input into each rmANOVA. The Coordinate factor describes the type of coordinate frame used. The Eye factor describes whether monocular (left or right) or binocular data were used. The Calibration factor describes the routine when testing on Calibration data. The PredictedCalibration factor describes what data were *input* into the model to calculate the errors. The Location factor describes the specific location that the participant was to interact with.

Levels All Coordinates had two levels: Cartesian and Spherical. Eye had three levels: Right, Left, and Both. Calibration had four levels: ExperimenterSweep, Paint, Self, and Stationary (see 2.2.3). PredictedCalibration had four levels: ExperimenterSweep, Paint, Self, and Stationary. Location had five levels in the Validation trials: Neutral, SideCart, Home, GreenShelf, and BlueShelf but only four levels in the Task trials: Neutral, SideCart, GreenShelf, and BlueShelf (see Fig. 2.4).

2.3.2 Calibration Trials

Here, we ran an rmANOVA on a 2 (Coordinate) × 3 (Eye) × 4 (Calibration) × 4 (PredictedCalibration) design.

A significant main effect of Coordinate was detected (LD: F(1,1) = 72.984, $p <$ 0.001, η^2 = 0.024; VA: F(1,1 = 100.586, $p <$ 0.001, η^2 = 0.020), where a model generated using Spherical data had lower error than with Cartesian data (see Fig. 2.6). A significant main effect of Calibration was detected (LD: F(1,1.971) = 11.894, p < 0.001, η^2 = 0.050; VA: F(1,2.058) = 12.393, $p <$ 0.001, η^2 = 0.047), with Stationary data on average performing best. A significant main effect of PredictedData was detected (LD: F(1,1.870) = 8.660, $p <$ 0.001, η^2 = 0.128; VA: F(1,2.294) = 6.995, $p <$ 0.001, η^2 = 0.119), with Stationary data being predicted more accurately in a Spherical coordinate system. A significant Coordinate × PredictedData interaction was detected (LD: F(1,1.259) = 18.377, $p <$ 0.001, η^2 = 0.015; VA: F(1,1.215) = 24.979, $p <$ 0.001, η^2 = 0.016), where Stationary data were the hardest to predict when predicted by non-Stationary models, but performed well when predicted by a Stationary Calibration model. A significant Coordinate × Calibration interaction was detected (LD: F(1,2.589) = 20.498, $p <$ 0.001, η^2 = 0.012; VA: F(1,2.114) = 26.167, $p <$ 0.001, η^2 = 0.010), with Stationary data again being hard to predict, unless it is predicted by a Stationary model. A significant Coordinate × Calibration × PredictedData interaction was detected (LD: F(1,4.338) = 4.942, $p <$ 0.001, η^2 = 0.006; VA: F(1,3.834) = 7.644, $p <$ 0.001, η^2 = 0.006), driven by the performance of Stationary data on non-Stationary Calibration models (see Fig. 2.6B and D).

All other tests were either not significant or were rejected because they did not meet Cramer's adjusted α criterion [113].

2.3.3 Validation Trials

We used an rmANOVA on a 2 (Coordinate) × 3 (Eye) × 4 (Calibration) × 5 (Location) design.

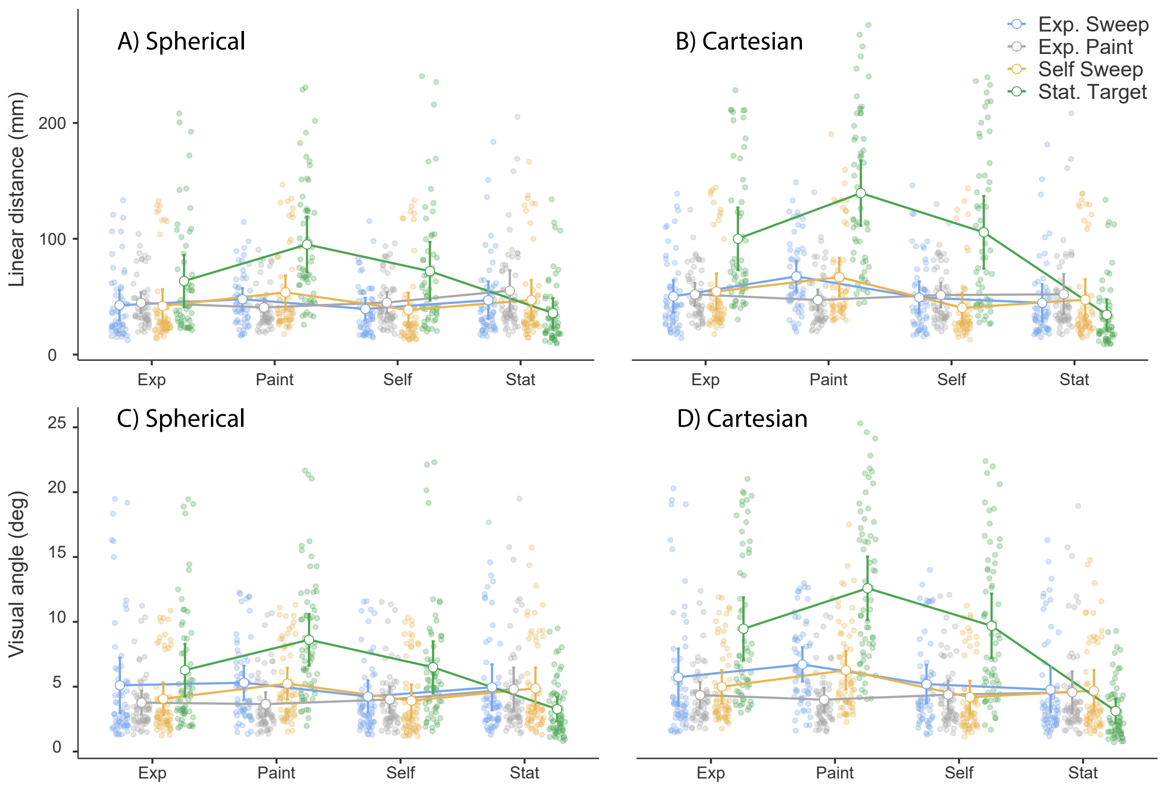


Figure 2.6: Plots showing the average performance of the models in linear distance (LD) and visual angle (VA) errors for the Calibration trials. The legend in the top right denotes which Calibration data was used during assessment. The top row are plots of LD errors (in mm) and the bottom row shows VA error (in degrees). The left plots are Spherical coordinate data, and the right plots are Cartesian coordinate data. 95% confidence intervals are used around each mean, with the observed scores that make up that mean plotted as translucent points. In all plots, the X axis denotes the Calibration routine used to train the model. A) Mean error of the Spherical LD models tested. Error is shown in mm at each location. B) Mean error of the Cartesian LD models tested. Error is shown in mm at each location. C) Mean error of the Spherical VA models tested. Error is shown in degrees at each location. D) Mean error of the Cartesian VA models tested. Error is shown in degrees at each location.

A significant main effect of Coordinate was detected (LD: F(1,1) = 25.928, $p < 0.001$, $\eta^2 = 0.008$; VA: F(1,1) = 33.716, $p < 0.001$, $\eta^2 = 0.003$), where a model generated using Spherical data had lower error than with Cartesian data. A significant Coordinate × Calibration interaction was detected (LD: F(1,2.501) = 7.838, $p < 0.001$, $\eta^2 = 0.006$; VA: F(1,3) = 7.274, $p < 0.001$, $\eta^2 = 0.002$), where the Spherical models tended to outperform Cartesian models, except when testing on Stationary data (see 2.7B).

All other tests were either not significant or were rejected because they did not meet Cramer's adjusted α criterion.

2.3.4 Task Trials

For the Task data, we were concerned with the performance of the GVs on real-world data. Here, we ran an rmANOVA on a 2 (Coordinate) × 3 (Eye) × 4 (Calibration) × 4 (Location) design.

A significant main effect of Coordinate system was detected (LD: F(1,17) = 21.475, $p < 0.001$, $\eta^2 = 0.008$; VA: (F(1,17) = 18.748, $p < 0.001$, $\eta^2 = 0.006$), where Spherical models had lower errors than Cartesian models (see Fig. 2.8). A significant main effect of Location was detected (LD: F(1,3) = 37.102, $p < 0.001$, $\eta^2 = 0.202$; VA: F(1,3) = 20.550, $p < 0.001$, $\eta^2 = 0.083$), where the SideCart location was the most difficult to predict, resulting in the highest errors overall (see Fig. 2.8A). A significant Coordinate × Calibration interaction was detected (LD: F(1,3) = 7.396, $p = 0.001$, $\eta^2 = 0.006$; VA: F(1,3) = 11.177, $p = 0.001$, $\eta^2 = 0.004$), with Spherical data outperforming Cartesian data in all cases except for Stationary data.

All other tests were either not significant or were rejected because they did not meet Cramer's adjusted α criterion.

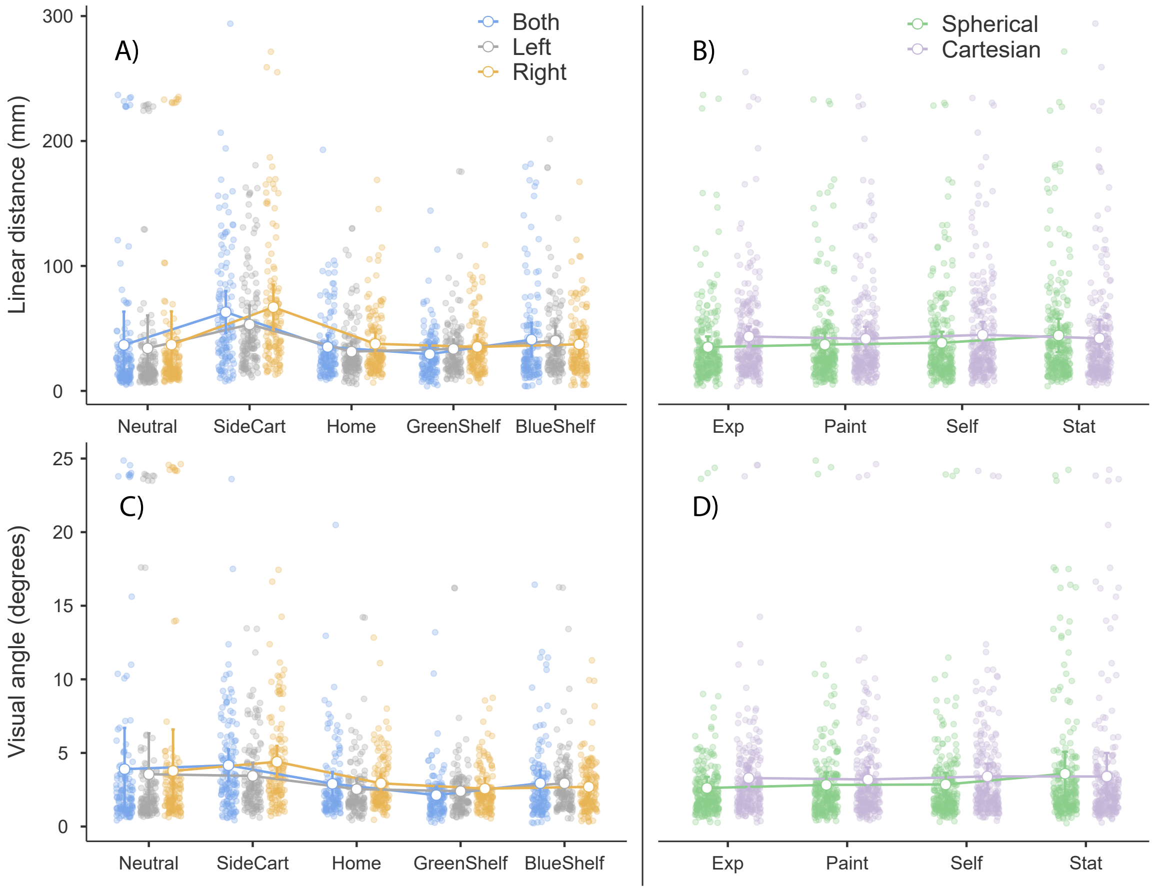


Figure 2.7: Plots showing the average performance of the models in linear distance (LD) and visual angle (VA) errors for the Validation trials. The plots on the left side are the average LD and VA errors (A, C) for each location in the Validation task (X axis). The legend here indicates what type of pupil input data was used. The plots on the right side the average LD and VA errors (B, D) for each of the Calibration types used as inputs to the models (X axis). The legend in the top right indicates whether a Spherical or Cartesian model was used. 95% confidence intervals are used around each mean, with the observed scores that make up that mean plotted as translucent points. A) Mean error generated at each of the locations (along the X axis) is shown for each type of Eye data used. Error is shown in millimetres at each location. B) Mean error generated each of the Calibration routines used is shown for Spherical and Cartesian models. Error shown in millimetres for each Calibration routine. C) Mean error generated at each of the locations (along the X axis) is shown for each type of Eye data used. Error is shown in degrees at each location. D) Mean error generated each of the Calibration routines used is shown for Spherical and Cartesian models. Error shown in degrees for each Calibration routine.

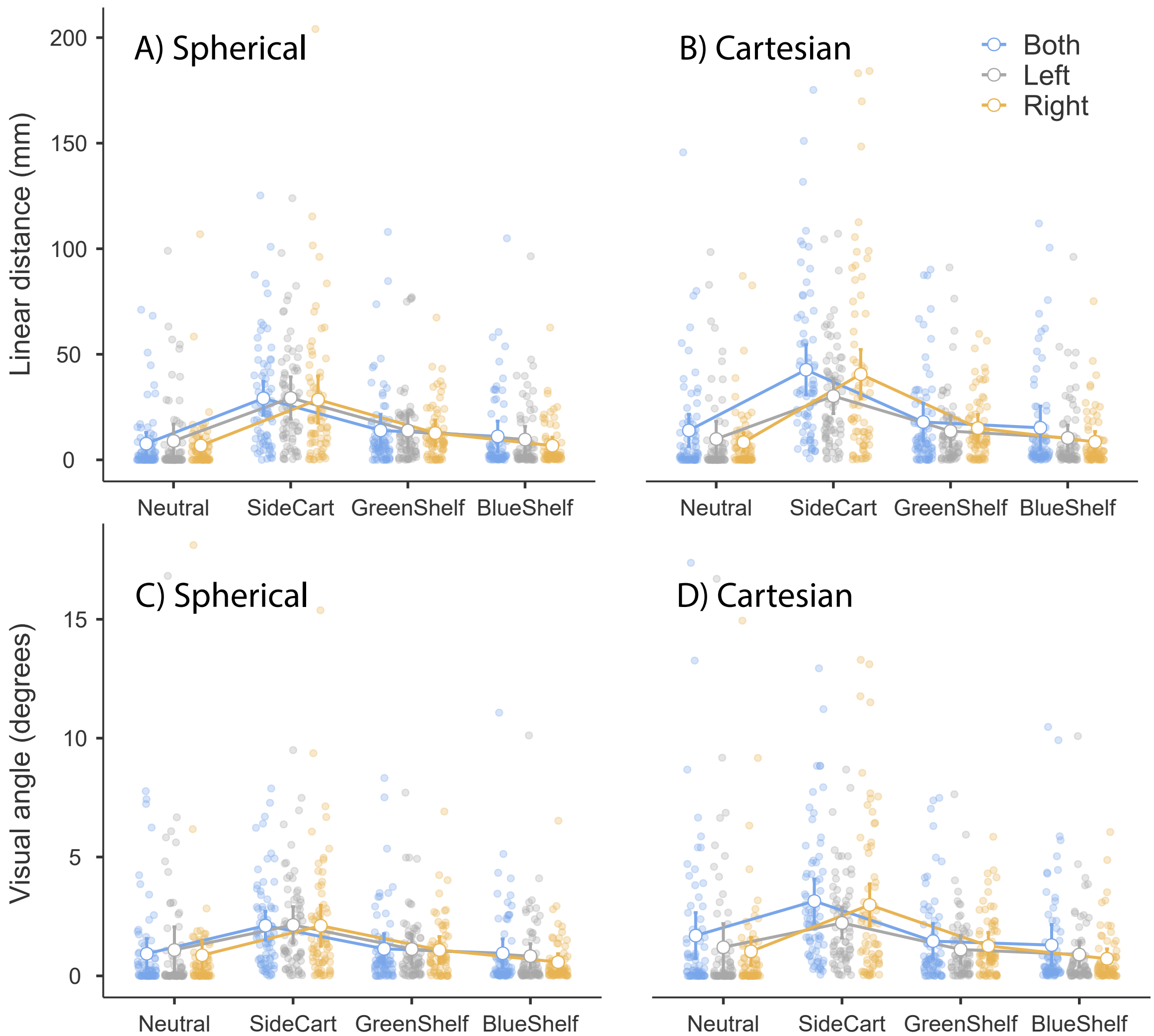


Figure 2.8: Plots showing the average performance at each location, split by the pupil input data in linear distance (LD) and visual angle (VA) errors for the Task trials. 95% confidence intervals are used around the mean points. The X axis denotes the Location during the Pasta Box task for all plots. A) Mean LD error for a Spherical coordinate system at each Location. Note that errors are below a centimetre when the participant is fixating on the Neutral marker (i.e., directly in front) or at the pasta box on the Blue Shelf. B) Mean LD error for a Cartesian coordinate system at each Location. C) Mean VA error for the Spherical coordinate system at each Location. Note that errors are below one degree of visual angle when fixating on either the Neutral marker or at the pasta box on the Blue Shelf. D) Mean VA error for a Cartesian coordinate system at each Location.

2.4 Discussion

Here we describe a method for generating 3D GVs using combined eye tracking and mo-cap data. We achieve this by collecting Calibration data where the eyes are continuously fixated on a tracked mo-cap marker, and using it to train a set of linear models to predict the 3D coordinates of the gaze fixation point. Within this method we explored four different Calibration routines (ES, SS, EP, ST), three options for eye input into the model (binocular, left, right), and two options for model coordinate system (spherical and cartesian).

We describe a set of four one-minute calibration procedures and their performance relative to one another in three different, but related analyses. All Calibration procedures were similar in that the goal of each was for the participant to fixate on a specific mo-cap marker for the duration of the procedure. We propose a simple model-based assessment (MATLAB's fitlm function) that allows us to give a recommendation for the best Calibration procedure based on the average GV error from a known location. First, we assessed a model trained on a Calibration routine's eye and mo-cap data calculating its error when testing on all other Calibration procedures' input data. Second, the participant completed several Validation trials. During these trials, the participant fixated on different areas of interest for long (˜5 s) periods of time to effectively emulate eye gaze behaviour during our Pasta Box task [26, 65, 94, 110], allowing us to assess error at each area of interest used during the task. Finally, participants performed a real-world task where they were instructed to perform the described Pasta Box task. Here, we demonstrate that our analysis techniques extend to data that were not recorded for the express purpose of being put through this analysis pipeline. That is, we can assess real-world task data and calculate performance metrics to best determine which Calibration procedure to use.

With respect to the type of coordinate system used to predict the gaze fixation point relative to the head, the results demonstrate that using a Spherical coordinate system generally results in a GV with a more accurate direction than using a Cartesian coordinate system as well as lower error overall. This result is aligned with our prediction, as we expected that the depth of fixation would be difficult to model based on pupil position data alone. It is worth noting that the reduction in error appeared to be systematic across all Calibration routines used—Spherical outperforms Cartesian. Although we did not assess gaze depth in the current study, when a Cartesian coordinate system is used, both the depth of fixation and direction of gaze are partially represented in all three eye gaze models (the x, y, and z coordinates). Whereas, using spherical coordinates confines the depth to one model (the 'r' coordinate) which does not impact the direction of the GV. In future work we intend to further explore the accuracy of the depth of the fixation point. However, the present work indicates that when only the direction of gaze is of interest, a spherical coordinate system should be used to generate GVs. When comparing each Calibration procedure on their ability to predict Calibration data, all models perform relatively well, with Stationary performing the best. However this is driven by the fact that all Calibration procedures (excluding Stationary) appear to have a difficult time predicting Stationary data. The Stationary routine was intended to allow the eyes to explore the maximal range of trackable pupil-space (see Fig 2.2D) while the actual target remained constant in space, potentially leading to a more robust model. The Stationary Calibration takes advantage of the compensatory vestibular-ocular response (VOR; [114]), in that the eyes and head move, but the gaze target remains static. The approach for the Stationary Calibration is one of quantity over quality; during the Stationary routine, data is collected from almost all accessible areas of the pupil, but not a lot of time is spent at each location nor are many of these locations generally useful during the Pasta Box task.

While it may be tempting to conclude that Stationary performs best overall, the data actually collected during Validation and Task trials do not reflect this same level of pupil space exploration. It is also important to consider that data collected during actual trials do not typically result in the pupil being located in positions on the eye consistent with the Stationary routine. Therefore, despite the advantage that the Stationary routine appears to show for Calibration data, the fact that it did not perform better during the more ecologically valid Validation and Task trials leads us to recommend using a Calibration routine that reflects the dynamics of eye exploration necessary during task completion. Anecdotally, explaining the Paint Calibration procedure was the simplest to perform and is extensible to any task, while the Experimental Sweep procedure was the easiest to keep consistent between sessions. Therefore, one of these two would be our recommendation for ease and consistency without sacrificing performance for ecologically valid data.

The Validation task was designed to mimic the behaviours that occur during a typical Pasta Box trial while still giving control over where the participant is looking and when. During a real trial, it is much more difficult to intrinsically know where the participant should be looking. These results are in line with the Calibration results, suggesting that Spherical coordinates result in more accurate GVs. One of the challenges the model faces is when the participant turns to fixate on the Side Cart, which results in higher error. Side Cart error appears to be worse when using data from Both pupils, and performs best when using monocular data, notably from the Left eye. One possible explanation for this is that the Left eye is always in view of the cameras when fixating on the target at the Side Cart, whereas the Right eye is potentially lost for a short duration. It is possible that using a 'hybrid' approach with monocular pupil data input, constantly switching to the 'better' eye, could result in superior performance. However, this is to be investigated and cannot currently be stated for certain. Regardless, it does suggest that collecting data from both eyes gives the most flexibility and opportunity to maximize data quality across sessions and even within a task. The Validation dataset functions as a 'sanity check' to ensure that the performance of the model is at least in line with our expectations: instead of tracking a moving marker (e.g., the wand during Calibration trials), the participant is fixating on a single static marker at a Task-relevant location. Performance appears to be similar to the Calibration trials analysis, suggesting the Validation dataset has done its job.

The Task results demonstrate that performance of the model has been effective on real-world data using a well-documented task [26, 65, 94, 110]. Previous work has shown that normative participants tend to fixate on the object they are about to interact with (or about to stop interacting with) for several hundred milliseconds [5, 12, 115]. Assessing performance on a real-world task is challenging because the behaviours of the participant are not controlled beyond simple verbal instructions (e.g., pick up the pasta box and move it to a new location) or visual demonstrations. However, we can use the principals described by Lappi [89] and [12] to find points in time when we expect the participant to initiate a reaching behaviour, such as a fixation on the object to be interacted with. With the identified fixation, we assessed error at this time point as the minimum distance between the 3D GV and the Pasta Box. Overall, performance of the model looked good; errors were remarkably low (see Fig. 2.8A & C). The average error for a Spherical coordinate system was below a centimetre and under a degree of visual angle for Task trials. We were surprised to find that error was lowest in the Task trials as they were the least-controlled in terms of participant instruction. However, when the participant turned their head, the error was significantly higher than at other areas.

Currently, there do not seem to be any standardized calibration procedures that also allow for the assessment of performance during real-world task use. Here we show a methodology that allows anyone with access to an eye tracker that outputs pupil locations in 2D space and a motion tracker in 3D space to generate GVs that can have as low as sub-centimetre error. While we did not find that any particular Calibration routine's data significantly outperformed any other, we found that using a Spherical coordinate system generated significantly less error on average when compared to a Cartesian coordinate system. Further, we suggest using a calibration routine that reflects the actual behaviours of the participant during task completion. For example, if the task involves looking at and reaching towards specific areas, a calibration routine that includes eye and hand movements towards those locations should generate higher quality models, or at minimum match the task demands and therefore be easier to employ.

2.5 Conclusion

We found that, when recording synchronized eye and mo-cap data for the purpose of producing accurate 3D gaze vectors, there are a few useful rules of thumb:

1. For fixations to real objects positioned in front of participants, gaze vectors generated using this approach will result in an average error of about 1-2 cm. If within peripersonal space (around 60 cm distance), this corresponds to about 1 visual degree.

2. A spherical coordinate system will on average produce more accurate gaze vectors (when depth is not considered).

3. Locations that require a head turn typically result in an accuracy falloff, adding about 2-3 cm of error in our data.

4. The best way to minimize error is to ensure quality data by making sure the eye tracker is properly fitted and the cameras are getting sufficient coverage of the eyes.

5. Binocular data, while not always the most accurate, gives the option to use either or both of the eyes when generating gaze vector models.

6. The calibration routine used should reflect the locations in space that the participant will be interacting with. More data is not always better.

Data Availability and Open Practices

All data and analysis scripts used in the present study are available at the following link: https://osf.io/znvwb/. This study was not preregistered prior to data collection or analysis.

References

5. Hayhoe, M. & Ballard, D. Eye movements in natural behavior. en. *Trends Cogn. Sci.* **9,** 188–194 (Apr. 2005).

12. Hayhoe, M. Vision Using Routines: A Functional Account of Vision. *Vis. cogn.* **7,** 43–64 (Jan. 2000).

17. Wispinski, N. J. *et al.* Reaching for known unknowns: Rapid reach decisions accurately reflect the future state of dynamic probabilistic information. *Cortex* **138,** 253–265 (2021).

26. Valevicius, A. M. *et al.* Characterization of normative hand movements during two functional upper limb tasks. en. *PLoS One* **13,** e0199549 (June 2018).

50. Kassner, M., Patera, W. & Bulling, A. *Pupil: an open source platform for pervasive eye tracking and mobile gaze-based interaction* in *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication* (Association for Computing Machinery, Seattle, Washington, Sept. 2014), 1151–1160.

65. Williams, H. E., Chapman, C. S., Pilarski, P. M., Vette, A. H. & Hebert, J. S. Gaze and Movement Assessment (GaMA): Inter-site validation of a visuomotor upper limb functional protocol. en. *PLoS One* **14,** e0219333 (Dec. 2019).

86. Kingstone, A., Smilek, D. & Eastwood, J. D. Cognitive ethology: A new approach for studying human cognition. *British Journal of Psychology* **99,** 317–340 (2008).

87. Land, M. F. The coordination of rotations of the eyes, head and trunk in saccadic turns produced in natural situations. *Experimental brain research* **159,** 151–160 (2004).

88. Tomasi, M., Pundlik, S., Bowers, A. R., Peli, E. & Luo, G. Mobile gaze tracking system for outdoor walking behavioral studies. *Journal of vision* **16,** 27–27 (2016).

89. Lappi, O. Eye movements in the wild: Oculomotor control, gaze behavior & frames of reference. *Neuroscience & Biobehavioral Reviews* **69,** 49–68 (2016).

90. Sperandio, I., Kaderali, S., Chouinard, P. A., Frey, J. & Goodale, M. A. Perceived size change induced by nonvisual signals in darkness: the relative contribution of vergence and proprioception. en. *J. Neurosci.* **33,** 16915–16923 (Oct. 2013).

91. Neggers, S. F. & Bekkering, H. Ocular gaze is anchored to the target of an ongoing pointing movement. en. *J. Neurophysiol.* **83,** 639–651 (Feb. 2000).

92. Land, M. F. & Hayhoe, M. In what ways do eye movements contribute to everyday activities? en. *Vision Res.* **41,** 3559–3565 (2001).

93. Land, M., Mennie, N. & Rusted, J. The roles of vision and eye movements in the control of activities of daily living. en. *Perception* **28,** 1311–1328 (1999).

94. Lavoie, E. B. *et al.* Using synchronized eye and motion tracking to determine high-precision eye-movement patterns during object-interaction tasks. en. *J. Vis.* **18,** 18 (June 2018).

95. Parr, J., Vine, S. J., Harrison, N. & Wood, G. Examining the spatiotemporal disruption to gaze when using a myoelectric prosthetic hand. *Journal of motor behavior* **50,** 416–425 (2018).

96. Wilson, M. R., Vine, S. J. & Wood, G. The influence of anxiety on visual attentional control in basketball free throw shooting. *Journal of Sport and Exercise Psychology* **31,** 152–168 (2009).

97. Vickers, J. N. & Williams, A. M. Performing under pressure: the effects of physiological arousal, cognitive anxiety, and gaze control in biathlon. en. *J. Mot. Behav.* **39,** 381–394. ISSN: 0022-2895. http://dx.doi.org/10.3200/JMBR.39.5.381-394 (Sept. 2007).

98. Parr, J. V. V., Vine, S. J., Harrison, N. R. & Wood, G. Examining the Spatiotemporal Disruption to Gaze When Using a Myoelectric Prosthetic Hand. en. *J. Mot. Behav.* **50,** 416–425. ISSN: 0022-2895, 1940-1027. http://dx.doi.org/10.1080/00222895.2017.1363703 (July 2018).

99. Van der Kruk, E. & Reijne, M. M. Accuracy of human motion capture systems for sport applications; state-of-the-art review. en. *EJSS* **18,** 806–819 (July 2018).

100. Chapman, H. *et al.* Posterior parietal cortex control of reach-to-grasp movements in humans. *European Journal of Neuroscience* **15,** 2037–2042 (2002).

101. Rossit, S. *et al.* No neglect-specific deficits in reaching tasks. *Cerebral Cortex* **19,** 2616–2624 (2009).

102. Rossit, S. *et al.* No Neglect-Specific Deficits in Reaching Tasks. *Cerebral Cortex* **19,** 2616–2624. ISSN: 1047-3211. eprint: https://academic.oup.com/cercor/article-pdf/19/11/2616/1073442/bhp016.pdf. https://doi.org/10.1093/cercor/bhp016 (Mar. 2009).

103. Whitney, D., Westwood, D. A. & Goodale, M. A. The influence of visual motion on fast reaching movements to a stationary object. *Nature* **423,** 869–873 (2003).

104. Mathis, A. *et al.* DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. *Nature neuroscience* **21,** 1281–1289 (2018).

105. Hebert, J. S. *et al.* Quantitative eye gaze and movement differences in visuomotor adaptations to varying task demands among upper-extremity prosthesis users. *JAMA network open* **2,** e1911197–e1911197 (2019).

106. SCCN. *Lab Streaming Layer* Mar. 2021.

107. Scheel, C. & Staadt, O. *Mobile 3D gaze tracking calibration* in *2015 12th Conference on Computer and Robot Vision* (2015), 176–183.

108. Browatzki, B., Bülthoff, H. H. & Chuang, L. L. A comparison of geometric-and regression-based mobile gaze-tracking. *Frontiers in human neuroscience* **8,** 200 (2014).

109. Ohno, T. & Mukawa, N. *A free-head, simple calibration, gaze tracking system that enables gaze-based interaction* in *Proceedings of the 2004 symposium on Eye tracking research & applications* (2004), 115–122.

110. Boser, Q. A. *et al.* Cluster-based upper body marker models for three-dimensional kinematic analysis: Comparison with an anatomical model and reliability analysis. *Journal of biomechanics* **72,** 228–234 (2018).

111. D'Errico, J. inpaint_nans. *MATLAB Central File Exchange* (2012).

112. JASP Team. *JASP (Version)[Computer software]* 2021. https://jasp-stats.org/.

113. Cramer, A. O. *et al.* Hidden multiplicity in exploratory multiway ANOVA: Prevalence and remedies. *Psychonomic bulletin & review* **23,** 640–647 (2016).

114. Zangemeister, W. H. & Stark, L. Active head rotations and eye-head coordination. *Annals of the New York Academy of Sciences* **374,** 540–559 (1981).

115. Prime, S. L. & Marotta, J. J. Gaze strategies during visually-guided versus memory-guided grasping. *Experimental brain research* **225,** 291–305 (2013).

Chapter 3

Unconscious frustration: dynamically assessing user experience using eye and mouse tracking

A version of this work was previously published as: S.A. Stone, Chapman, C.S. Unconscious frustration: using eye and mouse tracking to dynamically assess user experience on a menu interface. *Proceedings of the ACM on Human-Computer Interaction*, doi:10.1145/3591137.

Abstract

Eye-tracking has become easier to deploy in user experience (UX) studies to get a sense of where users attend to during interactions. Additionally, mouse tracking grants insights into the cognition driving the user's behaviours and end goals, as can measuring the coordination between the eye and mouse-cursor. We created a menu navigation task based on a popular video game to assess two populations: a local cohort, and a remote cohort. We used two different eye trackers (monitor-mounted hardware, and a webcam-based algorithm; local used both simultaneously, remote used webcam only) with concurrent mouse tracking to detect friction in the UX. We found that both eye trackers had similar performance and revealed a previously undetected friction point. We argue this friction point was only detected because of the use of quantified, coordinated unconscious behaviours (eye and hand movements). The methods demonstrated are easily integrated into current UX studies with minimal cost.

3.1 Introduction

3.1.1 Background: user experience

Measuring user experience (UX) is difficult and many of the methods employed in UX research (UXR) are qualitative [116, 117]. These approaches have obvious value, but make the critical assumption that the end-user is consciously aware of their own experiences. However, users can have a poor experience and not have cognitive access to the reasons why. For example, a user could be unconsciously focused on an irrelevant but visually salient part of a menu but not be able to recognize that distracted behaviour upon later reflection. As a result, users can be confused for reasons they may not be able to articulate. Using traditional qualitative methods, this unconscious frustration is hard (or impossible) to measure. While quantitative research has been conducted in UXR, few solid methodologies exist in the field [118–121]. Here, we investigate if a systematic quantitative approach to measuring unconscious behaviour by simultaneously recording eye gaze and mouse movements improves our interpretation of user behaviour.

3.1.2 Background: tracking the eyes and hands

As highlighted in the example above, where you look can be tightly linked to your experience but often falls below a threshold of awareness. Therefore, tracking a person's gaze is an important tool for measuring a person's unconscious experience. Fortunately, eye tracking has become more portable and affordable in recent years. An eye tracker records the position of the pupils, which can be transformed into a projected gaze position. Eye movements can inform us about the cognitive processes used to interact with our environment [12, 13, 122]. For example, Land & Hayhoe analyzed eye movements while participants completed a simple tea- or sandwich-making task. In general, the eye movements encoded the necessary steps to complete the task, and a generalized model was capable of predicting the order of behaviours. The increased portability of newer eye trackers has made it easier to extend eye-tracking studies outside of the laboratory. For example, monitor-mounted gaze tracking devices can be had for as little as two hundred US dollars and the advent of webcam eye tracking has made it possible to collect data from almost anyone in the world right from their own home. Due to this widespread availability, webcam eye tracking is specifically useful in the field of UXR as it grants access to the context UX researchers are most interested: naturalistic interaction. However, what is not clear is whether a webcam eye tracker can be used as a drop-in replacement for a more conventional monitor-mounted eye tracker given that it is likely to have lower temporal and spatial resolution.

While eye tracking can provide information about what a person may be seeking or distracted by, eyes typically are not used as an input device. Therefore, to understand the chain from information to input that ultimately results in a particular experience, we also need to measure how people interact with their workspace. For digital UX, the most common input device is a computer mouse, whose movements can be easily collected during interaction. Measuring hand movements grants insight into the dynamics of cognition (for a review, see [22]). Freeman *et al.* argue that motor and cognitive systems in the brain are not independent, but rather deeply intermingled. Movements are continuously updated through visual cognition over time [27, 35, 36, 58]. Recording even simple hand movements can therefore offer deep insight into cognitive processes and how they dynamically unfold [15, 16, 19, 123, 124]. Simultaneous eye and mouse tracking may therefore grant deeper insight into the UX assessment process than previous methods.

3.1.3 Related works: eye and hand tracking in UXR

A primary goal of UXR is to understand what produces high quality experiences that promote end-user adoption (for a detailed review, see [125]). A major barrier to adoption is when a user experiences a friction point - an interaction that confuses or slows down the user. A useful test bed for friction points, and therefore a good candidate for us to test the collection of eye and hand movement data, is a user interface (UI). UIs are one of the primary ways that we interact with computer software, and how users navigate through UIs presents many opportunities for both conscious and unconscious friction points to arise. Because the eyes and the hands are the primary medium for users to interact with the UI, recordings will encode points of friction, but it can be tricky to interpret and analyze the data. One method is to collapse the data across time to get a static look at user performance. For example, when interacting with web pages, users tend to look at the areas they are about to interact with [118–121, 126]. An intuitive way to visualize static data is known as a heatmap, which requires collapsing across time to calculate averaged statistics. But, static data interpretations like heatmaps fail to capture real human behaviours which are typically dynamic, with reach and gaze trajectories being updated in real-time [27, 127].

Dynamic measures, then, may be best suited to accurately detect friction points, especially those of which the user is not aware. One way to collect coarse dynamic measures is by splitting the data into natural phases. Here, a phase can coincide with the user's current goal. For example, if a user was attempting to send an email on a web page, the task could ostensibly be split into three phases: 1) locating the button to compose an email, 2) composing the email and 3) locating and clicking on the send button. Friction can be detected in any or all of the phases, but if an analysis collapses across the entire task the researcher will be puzzled as to *when* the actual friction occurred. Navigating through a UI is similar and their structure therefore offers a useful scaffolding onto which we can structure an analysis of behaviour. A user will have a goal of the menu object they are trying to get to, and getting there will require the user to complete several 'sub-goals' before reaching their destination. For our purposes, we can split the eye and mouse time-series data into these phases to better identify unconscious points of friction.

Importantly, this rich time series data contains information not only about where a person looks and how they move, but also can be used to measure the tight coupling that typically exists between the eye and the hand. That is, within each phase of a task, an intuitive notion is that coordination between the eyes and the hands is necessary for effective interaction [128]. What is less intuitive is an effective way to calculate this relationship dynamically. One tool quantifies this eye-mouse coordination by calculating the velocities of both the hands and the eyes (i.e. the mouse cursor and gaze position) and checks 1) if they are going in the same direction (i.e. similar vectors) and 2) if the velocities are close to one another. This method generates what is known as a Tlead value [129], which approximates how far ahead the eyes are of the mouse cursor (or vice versa). Here, we can use the Tlead value as a measurement of the relationship that exists between the eyes and the hands dynamically. These types of measures are best used in dynamic environments, which most UX tends to exist in.

3.1.4 Assessing dynamic behaviours

In the following study, we investigated the utility of tracking mouse and gaze position using both a monitor-mounted eye tracker and a webcam eye tracker during naturalistic, video game-based UI menu navigation. We did this in three distinct ways: 1) splitting the task into smaller 'phases' such that each phase could be analyzed individually, 2) assessing the hand-eye coordination relationship dynamically over these phases and, 3) collecting data in two groups: a local and a remote cohort which had either simultaneous hardware and webcam-based eye tracking (local) or only webcam (remote)

We predicted that user friction could be detected through eye and hand movements when a UI interaction was broken into task relevant phases. Additionally, we were interested in the limitations of webcam-based eye tracking, and whether its shortcomings would prevent its effective use as a UXR tool. To address these questions, we collected data from two cohorts. The first cohort (Local) was studied in-person using two types of eye trackers (dedicated hardware and webcam), whereas the second cohort (Remote) completed the task remotely through the Amazon MTurk service using webcam eye tracking only. At the beginning of each trial, participants were given a task to complete through a simple text prompt. We experimentally induced friction by cuing participants about their specific goal using either a Direct prompt or an Indirect prompt. Direct prompts were explicit about which menu items to interact with and in what order, while Indirect prompts were more vague, presenting the task at a much higher level. We predicted that the Indirect prompts would be harder for the participants to complete, resulting in longer completion times and more mouse and gaze movement. Our goal was that this intentional manipulation would validate our measures such that any *other* changes could be taken as signs of naturally occurring friction. To address our second aim of determining the utility of webcam eye trackers, the dedicated hardware and webcam eye trackers were directly compared in the Local cohort to get a sense of the overall accuracy and performance of each. We predicted that even in the face of reduced temporal and spatial accuracy, webcam eye trackers would be capable of detecting any friction points identified using the higher-quality eye tracker. Finally, the Remote cohort allowed us to test if the effects found in the Local cohort could be replicated using only a webcam and where we did not control the collection environment. We predicted that we would see similar friction points in both the Remote and Local cohorts, while also not introducing a significant amount of noise into the data.

3.2 Methods

3.2.1 Participants

Ethical approval was granted by the University of Alberta Human Research Ethics Board under protocol Pro00087329 and ethical protocols were in adherence to the

1964 Declaration of Helsinki.

Ten BioWare employees were recruited for the local-cohort (all male, mean age: 30.6 ± 5.2 years). All participants gave informed consent to participate in the current study. Three participants had to be removed from the data pool (one had unusable data, and two had recording errors). We acknowledge that this is a gender biased sample, but it was a convenience sample and likely reflects the actual gender imbalance in this industry.

Thirty-eight subjects participated in the remote-cohort using the Amazon MTurk and Prolific platforms (19 females, mean age: 30.6 ± 9.5 years). All remote subjects gave informed consent to participate in the current study. No participants were removed from the dataset. All remote participants were paid $7.50 USD for their participation in this study.

3.2.2 Equipment

For the local cohort, two different eye trackers were used. The first was a Tobii Eye Tracker 4C: a consumer-grade monitor-mounted eye tracking solution, which offers data collection at 90 Hz at a cost of around $200 CAD. The second was a consumer-grade webcam (Logitech C270), capable of collecting data at a resolution of 1280×720 pixels at 30 frames per second. The purpose of this webcam was not to find the most powerful or feature-rich device, but rather to use a webcam that a typical person may own. Additionally, a standard optical wired desktop mouse was used (Dell MS116).

For the remote cohort, subjects required their own computer with a physical mouse (instead of a trackpad) and a webcam. We did not discriminate on the quality of the webcam, as we were interested to collect data from a wide range of computer setups. As such, we do not know the average resolution or collection frequency (i.e. frames per second) of the webcams used in the remote portion of data collection. What we can report is that the average sampling rate of the webcam eye tracker algorithm [130] was 15.17Hz (± 8.56Hz), which reflects not just the capabilities of the webcam, but also a participant's computer specifications and internet connection speed.

Software

For the Local cohort, custom software was written in C# to access and collect data from the Tobii eye tracker using the official Tobii SDK. Additional programs were also written to access and collect data about the mouse position and mouse clicks over time, again in C#[1]. Because there are multiple data streams that are being recorded simultaneously, it is important to ensure that all of these data streams are synchronized using a common source for time-stamping. We used Lab Streaming Layer [66], a data stream synchronization library designed specifically for this purpose.

For the webcam-based eye tracker (Local and Remote cohorts), we used an implementation built by Labvanced [130] that is capable of tracking gaze positions on the screen in their platform. At the beginning of each trial, a short (30s) calibration task appeared on the screen where the participant was guided to fixate on points in a circle. Through the use of websockets, a custom program written in Python triggered the recording of gaze positions on the screen, giving access to the timing information so we could later synchronize the webcam gaze data back to the mouse and other gaze (Tobii) data (Local cohort only).

For the Remote cohort, we relied on the built-in Labvanced mouse tracker, which sampled data at 60hz and was automatically synchronized with the Labvanced eye tracking data.

3.2.3 Task

The menu navigation task used for this study was designed in Labvanced. Labvanced is an online site-as-a-service that allows users to create and deploy psychology experiments [130]. Because we were interested in replicating an authentic user experience,

[1]All of the software used to collect the data are open source and available at the following URLs: https://github.com/scottastone/TobiiGazeLSL and https://github.com/scottastone/MouseLSLGUI

we recreated the main menu layout from a popular video game created by BioWare: Mass Effect 3 (ME3; [85]). The Labvanced task is an abstract version of the menus displayed in ME3 without any of the complex visual elements present in the original game (e.g such as the bright colours, or starry backgrounds; see Figure 3.1). We wanted users to navigate to common menu destinations, where typical users make adjustments such as to the game audio, in-game settings, or graphical settings. We replicated the following menu options, herein referred to as *goal frames*: Accomplishments, Gameplay, Graphics, Mouse, Narrative, and Sound.

The general procedure of the task is as follows: the participant is given a prompt where they are given a goal to complete. Once acknowledged, the participant then navigates through the menu to the goal frame where they perform a target interaction (e.g. check a box, find a piece of information, adjust a scroll bar, etc), before navigating back to the main menu to click on *"Exit game"*. This will trigger the beginning of the next trial. For an example visual of the general flow of the task, see Figure 3.1. In the example, the user is told to navigate to the *Graphics* goal frame and to turn on 'Antialiasing'.

Task design

Two types of prompts were presented to the participant throughout the task: a *Direct* or *Indirect* prompt.

A Direct prompt is a clear and concise set of instructions that guides the user towards a specific end goal by providing the names of intermediary menu buttons that they need to click on or adjust in order to reach the target.. An example of an Direct prompt is: *"Go to Extras - Options - Gameplay and turn Hints on."*. This prompt is meant to give direct instructions about which buttons to click on, while taking away any extra information.

The Indirect prompt aims to mimic the thought process of a video game player who wants to access the settings menu to make necessary adjustments. For example,

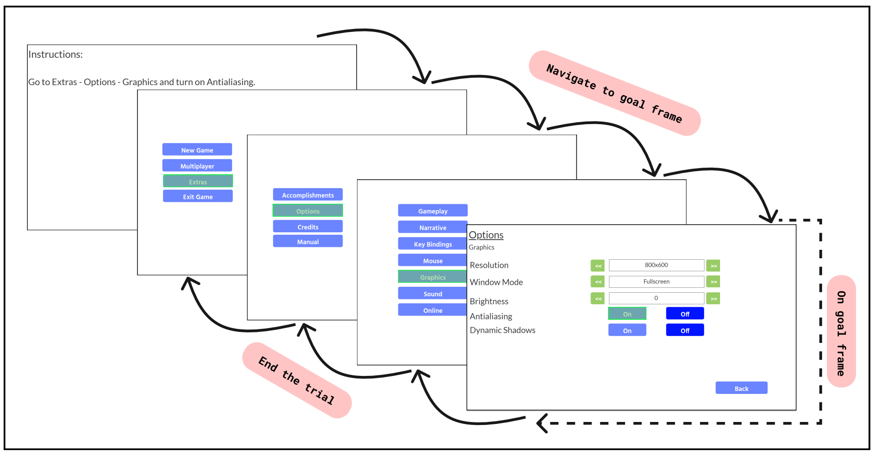


Figure 3.1: A waterfall representation of the order of frames that the participant will encounter in a typical trial. At the top left is the first frame of the trial the user will see, known as the prompt. The prompt contains all of the instructions necessary to complete the trial. Each frame contains a goal target, which is highlighted in green in this figure only for clarity. Upon clicking continue (bottom centre of the left-most frame; not shown), the participant will enter the next frame (moving clockwise), starting the trial. The user continues to navigate through the menu until they reach the goal frame (bottom right frame). This is the *"Navigate to the goal frame"* phase. The participant then completes the given task on the goal frame (*"On goal frame"*). Next, the participant will work to *"End the trial"* by moving backwards (moving counter-clockwise) in the menu towards the first frame they encountered following the prompt. Here, they must click "Exit game" to start the next trial.

one prompt might be *"You notice you are having trouble hearing the characters' dialogue in game. Go and turn on the subtitles."*. These prompts are necessarily more abstract in that they give less direct information to the player, but enough information to complete the task. As previously discussed, this manipulation was introduced specifically because we expected the Indirect prompts to be more difficult for the end user. In general, if any friction occurred during the task, the Indirect prompt should exacerbate that friction.

This task design was chosen to allow the participant to more fully explore the task space without being explicitly told to do so. An Indirect prompt does exactly this. Some of this time spent exploring will not be fruitful, which we interpret as unwanted friction. Conversely, we expected Direct prompts to have less exploration time, as the interim buttons needed to reach the goal frame are given directly. However, some goal frames will require lots of exploration (such as Accomplishments) because providing information is their primary role, rather than being used to make adjustments (i.e. Graphics goal frame). The key difference here then, is understanding the context in which the goal frames *should* be used in order to detect friction within them.

3.2.4 Procedure

For the Local cohort, the participant was brought into the room and seated in a comfortable computer chair. They were given the opportunity to adjust the height, lumbar, and arm heights to ensure they were comfortable for the duration of the experiment. They were then positioned in front of the computer monitor (27" Dell, 2560x1440 pixels, 60Hz), approximately 40cm away. If the participant was wearing a mask, they were asked to remove it, as the webcam eye tracking algorithm does not work with one on. Each participant was given up to 10 practice trials on the task to familiarize themselves with the procedure, though none of the participants needed all 10. Upon ensuring that the participant was comfortable and understood the task, the experimenter began the study and left the room. The participant was given a total of

96 trials (48 Direct prompts and 48 Indirect prompts). The task took approximately one hour to complete.

Remote cohort subjects received on-screen instructions to seat themselves comfortably and ensure they had ample lighting in the room. Collecting data remotely meant we had no guarantees of environment or posture, nor could we know what hardware was being used by the participant (e.g. monitor size and resolution were not collected). Each participant was given up to 10 practice trials on the task to familiarize themselves with the general flow of the study. After completing webcam calibration, the participant was given a total of 104 trials (52 Direct prompts and 52 Indirect prompts). The task took approximately one hour to complete. For the remote cohort only, we re-balanced the task to include more equal representation of each goal frame. As such, instead of having a total of 96 trials, participants completed a total of 104 trials[2]. The basic presentation of the prompt and overall flow of the task however remained identical and took the same amount of time.

3.3 Dependent Variables

There are four general categories of data collected: *time*, *mouse*, *gaze*, and *coordination*. Within each category subdivided the data into three measures defined by distinct phases: navigating to the goal frame, on the goal frame, and ending the trial (see Figure 3.1 to see how the phases were split). Please note that the results of *mouse* data are not reported in the main text, as its dynamics are captured by the *coordination* measure. For full reporting of the mouse data, refer to Appendix B.

Time

All time data are reported in seconds, split into each of the three phases.

[2]All of the prompts can be found at the data repository: https://osf.io/f49xg/

Gaze

Gaze distance (e.g. the cumulative amount of movement) in each phase was calculated. To test for the predicted accuracy drop off in webcam data we calculated two additional gaze measures for the interactions on the goal frame: time in goal target and minimum distance to goal target. Time in goal target is the amount of time the gaze was within the bounds of the goal target, which was the piece of information that needed to be interacted with on the goal frame. Minimum distance is the smallest distance between any gaze sample and the bounds of the goal target, with this value achieving 0 if the gaze was within the goal target at any time. A lower accuracy system should have a larger minimum distance and less time in target. All gaze data are reported in standardized units and was sampled at 90 hz (Tobii data) and approximately 20 hz (Webcam data), which were up-sampled to 90hz for analysis. The units were standardized by converting all pixel coordinates to fit the Labvanced coordinate space of 800×450 units.

Coordination

The amount of time the eyes and hands moved together was quantified using Tlead [129] and examined across each phase. Tlead calculations necessitate data that are sampled at identical frequencies, meaning the gaze and mouse data must be resampled to a common sampling rate. While the calculation of Tlead returns three possible values: NaN proportion (meaning one or both of the data streams are decoupled from one another), positive proportion (gaze leading the cursor), and negative proportion (cursor leading gaze), we were primarily interested in the amount of coupling observed. Therefore, for our analysis, we collapsed Tlead into the percent of time the eyes and hands were either coupled (e.g. signed; non-NaN) or decoupled (NaN). Importantly, not all interactions require tight eye-hand coupling; sometimes our eyes collect information in one space, but our hands work in another. Our novel use of Tlead allows us to quantify tasks that require tight coupling versus those that force eye-hand decoupling.

3.4 Results

3.4.1 Statistical test designs

All statistics were calculated using Jamovi 2.3.18 [131]. Repeated measures ANOVA (rmANOVA) were sphericity-corrected using the Greenhouse-Geisser estimate of the F statistic where necessary.

For the Local cohort, because two different eye trackers were used concurrently, tests that investigate gaze-based measures have a $6 \times 2 \times 2$ rmANOVA design: 6 GoalFrames (accomplishments, gameplay, graphics, mouse, narrative, sound), 2 Conditions (Direct, Indirect) and 2 Eyetrackers (Webcam and Monitor-mounted). Mouse and time measures were identical regardless of the eye tracker used, so the rmANOVA design only required a 6×2 design (GoalFrame $\times$ Condition). To test coordination measures, we used an rmANOVA with a $6 \times 2 \times 2$ design (GoalFrame $\times$ Condition $\times$ Eyetracker).

Since we were interested in comparing the performance of the Remote and Local cohorts, we also used a mixed ANOVA with a $6 \times 2 \times 2$ design (GoalFrame $\times$ Condition) for within subjects and the subject Cohort (Local-webcam, Remote) as the between subjects factor to compare across all of our measures. For these tests, only results with significant main effects of Cohort or interactions involving Cohort will be reported.

3.4.2 Statistics: Time

Time: to navigate to the goal frame

These data are the average time that it took a participant to locate the intended goal frame given by the prompt. In the Local cohort, a main effect of GoalFrame was detected (F(1,1.411) = 15.274, $p = 0.003$, $\eta^2 = 0.403$), with Narrative taking the longest for participants to find. The Remote cohort showed the same pattern with no interactions involving Cohort. A main effect of Cohort was detected (F(1,1) = 4.503, $p = 0.039$, $\eta^2 = 0.041$), where the Remote cohort took longer, which we interpret to be an expertise effect between the cohorts.

Time: on the goal frame

This is the average amount of time the participant spent to complete the task on the goal frame given by the prompt. In the Local cohort, a main effect of Condition was detected (F(1,1) = 35.743, $p < 0.001$, $\eta^2 = 0.014$), with Indirect prompts taking longer than Direct prompts. This suggests that Indirect prompts do in fact take longer to complete, agreeing with our earlier prediction. A main effect of GoalFrame was detected (F(1,1.791) = 160.472, $p < 0.001$, $\eta^2 = 0.898$), with Accomplishments taking the longest time to complete the task, followed by Sound. A significant GoalFrame $\times$ Condition interaction was detected (F(1,13.104) = 10.236, $p = 0.002$, $\eta^2 = 0.024$), where users tended to take longer when given an Indirect prompt on all goal frames with the exception of Narrative. The Remote cohort showed the same pattern and the cohort comparison showed no significant main effects or interactions involving Cohort.

Time: to end trial

This is the average time it takes for the user to end the trial after they make the intended manipulation on the goal frame. In the Local cohort, a main effect of Condition was detected (F(1,1) = 7.889, $p = 0.031$, $\eta^2 = 0.074$), where participants took longer to end the trial on Indirect prompts than Direct prompts. The Remote cohort showed the same pattern and the cohort comparison showed no significant main effects or interactions involving Cohort.

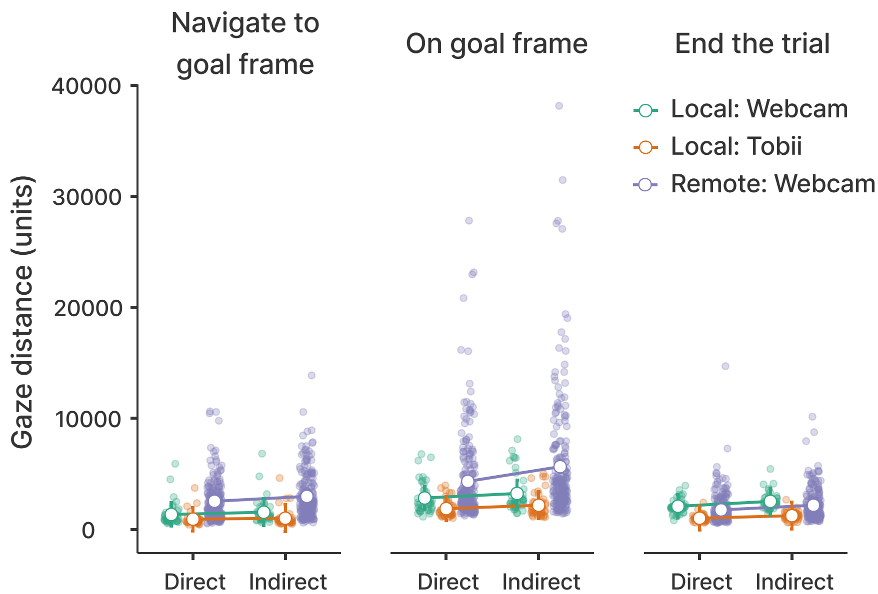


Figure 3.2: The eye distance traveled over the course of an entire trial split by the phase. Data for both the Local and Remote cohorts are shown, split by the type of eye tracker used. The green circles and lines are webcam eye tracker data from the Local cohort, the orange circles and lines are the Tobii eye tracker data from the Local cohort, and the purple circles and lines are the webcam eye tracker data from the Remote cohort. The X axis shows which kind of prompt was presented to the participant. The Y axis shows the average distance traveled in units. Each data point is scatter-plotted under each mean. 95% confidence intervals are plotted around the means. A significant main effect of Condition demonstrates that Indirect prompts result in more distance being traveled. A significant interaction between Condition $\times$ Eyetracker shows the webcam accumulates more error over time.

3.4.3 Statistics: Gaze

Gaze: distance to navigate to the goal frame

This is the average distance in pixels that the gaze traveled on the screen when the participant was navigating to the goal frame. See Figure 3.3A. In the Local cohort, a significant main effect of GoalFrame was detected (F(1,1.530) = 16.732, $p < 0.001$, $\eta^2 = 0.306$), with Narrative goal frames requiring the most gaze distance to find. A significant main effect of Eyetracker was detected (F(1,1) = 8.819, $p = 0.025$, $\eta^2 = 0.067$), with the webcam eye tracker overestimating the distance traveled relative the monitor-mounted system. A significant GoalFrame × Eyetracker interaction was detected (F(1,1.983) = 6.539, $p = 0.012$, $\eta^2 = 0.001$), where goal frames that required more time to complete also resulted in higher estimates for distance traveled for webcam eye tracker data relative to the monitor-mounted eye tracker. This suggests the webcam eye tracker likely overestimates the distance traveled and this effect scales with time. The Remote cohort showed the same pattern with no interactions involving Cohort. A main effect of Cohort was detected (F(1,1) = 5.233, $p = 0.027$, $\eta^2 = 0.069$), where Remote participants moved their eyes more, likely due to the expertise differences between the cohorts.

Gaze: distance on the goal frame

This is the average distance in pixels that the gaze traveled on the screen when the participant was on the intended goal frame completing the task outlined by the given prompt. See Figure 3.3B. In the Local cohort, a significant main effect of Condition was detected (F(1,1) = 17.499, $p = 0.006$, $\eta^2 = 0.016$), where Indirect prompts required more eye movements than Direct prompts. Again, this supports our earlier prediction that Indirect prompts are more difficult, and will thus take longer to complete with more eye movements. A significant main effect of GoalFrame was detected (F(1,2.151) = 60.987, $p < 0.001$, $\eta^2 = 0.566$), with Accomplishments requiring the most gaze movements overall. A significant main effect of Eyetracker was detected (F(1,1) = 16.356, p = 0.007, η^2 = 0.134), where the webcam eye tracker overestimated the distance traveled relative to the monitor-mounted eye tracker. A significant GoalFrame × Condition interaction was detected (F(1,2.482) = 5.090, p = 0.016, η^2 = 0.034), where Indirect prompts resulted in higher distances traveled, with the exception of on Mouse and Narrative goal frames. A significant GoalFrame × Eyetracker interaction was detected (F(1,1.251) = 14.540, p = 0.004, η^2 = 0.039), where goal frames that took longer to complete had disproportionately higher distances traveled for the webcam eye tracker versus the monitor-mounted eye tracker. A Condition × Eyetracker interaction was detected (F(1,1) = 11.591, p = 0.014, η^2 = 0.000), where the webcam eye tracker resulted in a higher travel distance difference (compared to the monitor-mounted eye tracker) on Indirect prompts. For the cohort comparison, a significant main effect of Cohort was detected (F(1,1) = 4.857, p = 0.033, η^2 = 0.041), where Remote participants had more gaze movements on the goal frame. A significant GoalFrame × Cohort interaction was detected (F(1,1.142) = 7.607, p = 0.006, η^2 = 0.046), where Remote participants had more gaze movements on the Accomplishments goal frame. A significant Condition × Cohort interaction was detected (F(1,1) = 9.685, p = 0.003, η^2 = 0.002), where Remote participants had more gaze movements when given a Indirect prompt, again likely attributed to the Remote cohort's inexperience with the menu layout. A visual summary of these results can be seen in Figure 3.3B.

Gaze: distance to end trial

This is the average distance in pixels that the gaze traveled on the screen when the participant had completed the task outlined by the prompt and was on their way to end the current trial. In the Local cohort, a significant main effect of Condition was detected (F(1,1) = 20.480, p = 0.004, η^2 = 0.028), where Indirect prompts resulted in more gaze distance traveled. A significant main effect of Eyetracker was detected (F(1,1) = 25.193, p = 0.002, η^2 = 0.504), where the webcam eye tracker traveled

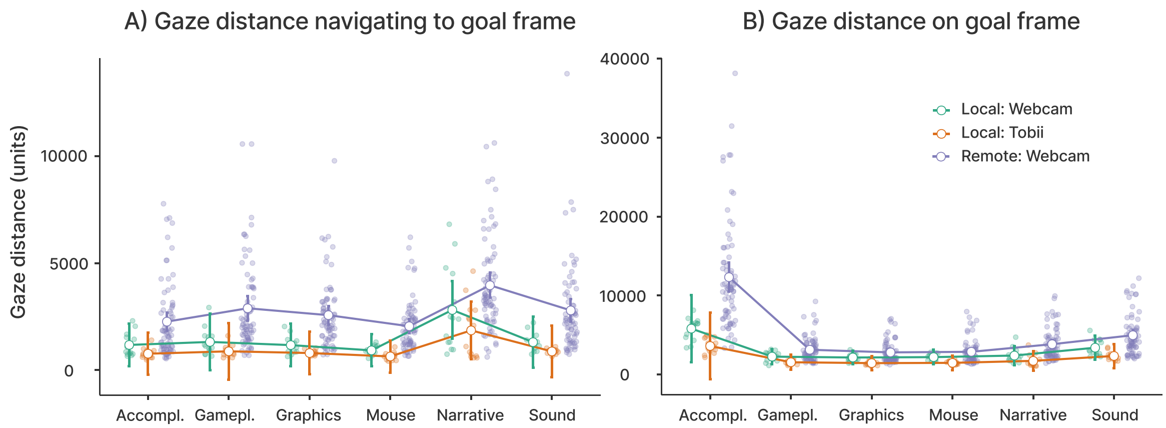


Figure 3.3: Plots of the gaze distance in two phases: navigation and on the goal frame. Data for both the Local and Remote cohorts are shown, split by the type of eye tracker used. The green circles and lines are webcam eye tracker data from the Local cohort, the orange circles and lines are the Tobii eye tracker data from the Local cohort, and the purple circles and lines are the webcam eye tracker data from the Remote cohort. The X axis has each of the goal frames (with Accompl. and Gamepl. for Accomplishments and Gameplay, respectively). The Y axis is the distance traveled in standardized units. The underlying data are scatter-plotted under each mean. 95% confidence intervals are plotted around the means. A) The average gaze distance traveled navigating to the goal frame. Here, we can see an increase in the amount of gaze movement required to enter the Narrative goal frame, regardless of the participant pool or eye tracker used. B) The average gaze distance traveled while on the intended goal frame. Here, we can see that Accomplishments required more gaze movements relative to other frames. The Remote cohort looked around the most, which is indicative of their relative inexperience with the menu and variable collection environment.

a further distance. A significant Condition × Eyetracker interaction was detected (F(1,1) = 9.421, p = 0.022, η^2 = 0.004), where the web cam eye tracker traveled more during the Indirect prompts. The Remote cohort showed the same pattern and the cohort comparison showed no significant main effects or interactions involving Cohort.

Gaze: minimum distance to goal target

This is the minimum distance between the gaze and the goal target(s) on the goal frame throughout an entire trial. A significant main effect of Eyetracker was detected (F(1,1) = 21.146, p = 0.004, η^2 = 0.486), where the webcam eye tracker had higher distances to any of the goal targets on screen, confirming the webcam eye tracker was overall less spatially accurate. A main effect of Cohort was detected (F(1,1) = 7.589, p = 0.009, η^2 = 0.092), where the Remote cohort had a lower minimum distance. A significant Cohort × Condition interaction was detected (F(1,1) = 4.907, p = 0.032, η^2 = 0.006) where the Local cohort had a larger decrease in minimum distance between conditions than the Remote cohort.

Gaze: on target time

This is the average amount of time the gaze data were within the bounds of any of the intended targets on the goal frame outlined by the prompt. A value of 0 indicates that gaze was never within the bounds of the goal targets on the goal frame. In the Local cohort, a significant main effect of GoalFrame was detected (F(1,2.317) = 2.317, p < 0.001, η^2 = 0.073), where gaze was within the bounds of the goal targets on the Narrative goal frame. A significant main effect of Condition was detected (F(1,1) = 7.342, p = 0.035, η^2 = 0.005), where Indirect prompts lead to a longer dwell time on the goal targets. A significant main effect of Eyetracker was detected (F(1,1) = 69.095, p < 0.001, η^2 = 0.663), where the webcam eye tracker had significantly less time spent within the target bounds. A significant GoalFrame × Eyetracker interaction was detected (F(1,2.324) = 13.451, p < 0.001, η^2 = 0.064), where the time spent on the Narrative goal frame using the monitor-mounted eye tracker had disproportionately more time spent within the target goal target bounds than other goal frames. A significant Condition × Eyetracker interaction was detected (F(1,1) = 6.840, p = 0.040, η^2 = 0.004) where the amount of time gaze data were within the bounds of the goal target was equal across Condition, whereas the monitor-mounted eye tracker had higher values for Indirect prompts. For the cohort comparison, a main effect of Cohort was detected (F(1,1) = 4.269, p = 0.045, η^2 = 0.036), where the Remote cohort looked at targets for longer, likely reflecting their inexperience with the menu system. A significant GoalFrame × Cohort interaction was detected

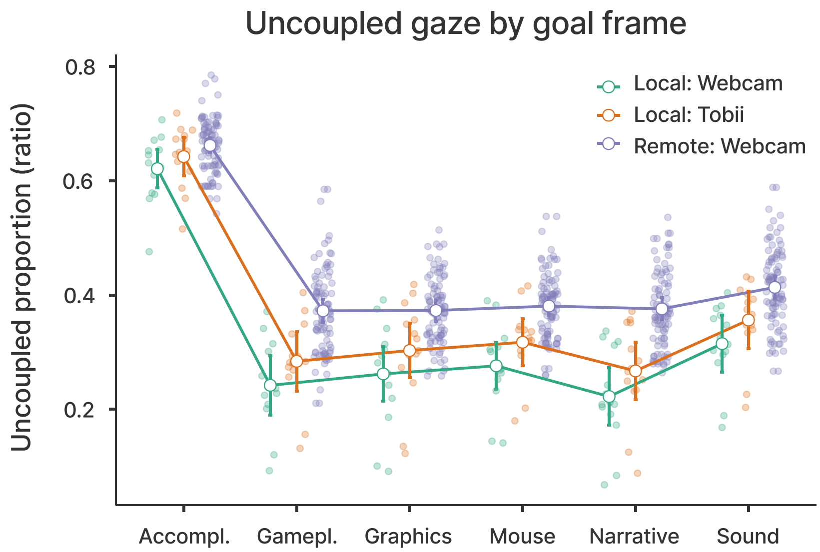


Figure 3.4: The average proportion of $\text{Tlead}_{\text{NaN}}$ values for each type of goal frame. Data for both the Local and Remote cohorts are shown, split by the type of eye tracker used. The green circles and lines are webcam eye tracker data from the Local cohort, the orange circles and lines are the Tobii eye tracker data from the Local cohort, and the purple circles and lines are the webcam eye tracker data from the Remote cohort. The X axis has each of the goal frames (with Accompl. and Gamepl. for Accomplishments and Gameplay, respectively). The Y axis is proportion of $\text{Tlead}_{\text{NaN}}$ values as a ratio. The underlying data are scatter-plotted under each mean. 95% confidence intervals are plotted around the means. Here, we can see that there are significantly more $\text{Tlead}_{\text{NaN}}$ values on the Accomplishments goal frame, suggesting that the eyes and hands are dissociated from one another for this task. The eyes are drawn to search on one part of the screen while the hands click relatively stationary in another.

(F(1,3.755) = 2.491, p = 0.049, η^2 = 0.018), where the Remote cohort looked at targets on most goal frames more, with the exception of Narrative and Sound.

3.4.4 Statistics: Coordination

Coordination: Tlead

This is the percent of time the eyes and hands were uncoupled. In the Local cohort, a significant main effect of GoalFrame was detected (F(1,2.059) = 241.902, $p < 0.001$, $\eta^2 = 0.762$), where the the eyes and hands were disproportionately uncoupled on the Accomplishments goal frame relative to the others. A significant main effect of Eye-tracker was detected (F(1,1) = 209.679, $p < 0.001$, $\eta^2 = 0.016$), where data collected with the Webcam were more coupled. A significant GoalFrame × Eyetracker interaction was detected (F(1,2.144) = 6.413, $p = 0.011$, $\eta^2 = 0.000$), where the Webcam eye tracker had a larger difference between coupled and uncoupled proportions for all goal frames. See Figure 3.4 for a visual of the results. For the cohort comparison, no main effect of Cohort was detected. A significant GoalFrame × Cohort interaction was detected ((F(1,2.521) = 3.881, $p = 0.016$, $\eta^2 = 0.006$), which seems to be driven by the Local cohort showing slightly more coupling on the Narrative goal frame. A significant Condition × Cohort interaction was detected ((F(1,1) = 16.663, $p < 0.001$, $\eta^2 = 0.003$), where the Remote cohort had a slightly larger difference between the prompts given.

3.5 General Discussion

3.5.1 Main findings

We collected eye, hand and coordination measures in two cohorts of participants as they completed a simple UI navigation task. For the Local cohort, we investigated a consumer-grade monitor-mounted eye tracker and directly compared it to a simultaneously recorded webcam-based eye tracking algorithm. Because both datasets were recorded concurrently in the Local cohort, we were able to directly compare performance. To our surprise, we found that the webcam produced sufficient data quality to be directly comparable to the more sophisticated monitor-mounted eye tracker in revealing important features of a user's experience. While the temporal and spatial resolution was much lower in the webcam data, we were still able to detect a friction point in the menu design that has not, to our knowledge, been discovered previously. For the Remote cohort we looked exclusively at webcam data and did a between groups comparison to the Local data to investigate if our findings would hold once the experimental control over hardware and environment was removed.

In general, we found similar results across both cohorts. In terms of the initial prompt used, an Indirectly worded prompt resulted in worse performance than a Direct one (see Figure 3.2). That is, users took longer and moved both their mouse and eyes more with an Indirect prompt. In both cohorts, we also found two candidate points of friction. The first candidate friction point was on the Accomplishments goal frame, where participants spent a significant amount of time locating given accomplishments. The second candidate friction point was navigating to the Narrative goal frame. A conventional analysis using only time-based measures would have concluded that these two delays were both indicators of friction. However, our unique analysis of eye and mouse movements, and especially their coordination, paints quite a different picture. For the Accomplishments frame, we saw an increase in eye-hand decoupling (see Figure 3.4), indicating much more eye movement than mouse movement. In context of the task given, this was in fact a reasonable result; the Accomplishments task required repeatedly clicking a button with the mouse while reading and searching through text descriptions that appeared with each click. This required lots of gaze movements and few mouse movements, suggesting that the eyes and hands had distinct roles. As a result, this is actually the expected behaviour rather than a point of friction. By comparison, the delay when navigating to the Narrative frame was accompanied by more eye and hand movements, which remained coupled. The extended search time and increased eye and hand movements indicate confused and inefficient exploration (eye) and exploitation (hand). Only the collection and analysis of gaze and movement behaviours was able to disambiguate these cases and identify a true point of friction.

3.5.2 Local versus Remote cohorts

While the level of congruence between the two cohorts was remarkable, especially considering the lack of experimental control over the conditions in which the Remote cohort was tested, one key difference between cohorts was the level of experience. The Local cohort were all employees of the company that developed the video game the UI task was based on, whereas the Remote cohort did not necessarily have any experience with video games. Collecting data from less-experienced users can help pinpoint potential issues that the more experienced group may have simply adapted to over time. For example, a new video game player may not be familiar with game-specific terminologies used in the menus (e.g. graphical options such as vertical synchronization, or mouse sensitivity adjustments). Our findings support this conclusion, showing that the Remote cohort was significantly slower and had more gaze and hand movements than the Local cohort, but *only* when they were operating on the goal frame (see Figure 3.3). That is, they exhibited no difference in the mechanics of the task (moving the mouse, clicking buttons, exiting each trial) and their inexperience was only evident when task relevant knowledge was expected to play a significant factor. Interestingly, regardless of the level of experience, the friction point described above naturally emerged in the data for users in both cohorts. As researchers, we are able to pinpoint specifically when the friction was occurring, and provide actionable insight into how to fix the problem. Here, we saw that all users tended to have issues finding the Narrative goal frame, and one potential reason for this could be that the name of the menu is ambiguous or may not reflect its contents for most users. A UX researcher could test this hypothesis by altering the menu's name and testing if the friction still exists. It is also possible to imagine expertise-dependent adjustments UX designers may wish to use, such as tailoring an experience better suited for a novice as compared to an expert user. Additionally, it is important for the UX researcher to contextualize the utility of friction. Some friction can actually be useful to ensure a user is paying attention to critical components and can lead to higher user satisfaction [132].

Recording gaze and mouse data gives insights into the end-user's behaviour, most of which is unconsciously controlled [22, 27, 35, 36, 57, 133]. Typically, when assessing user interfaces, researchers use qualitative approaches such as interview-style questions aimed at probing the end-user's conscious experience [125]. We argue that while a qualitative approach might lend to some insights about the overall experience, some friction points cannot be uncovered simply because users themselves may not be aware they were experiencing friction. Here, we argue that the methods employed in the present study demonstrate that friction can be detected through the use of easy-to-deploy hardware and software at a minimal cost. It is important to note that this study did not exhaustively compare the performance of hardware and software eye trackers (for a study investigating this, please see Wisiecka *et al.* [134]), so it is important that the researcher knows the inherent limitations of the implementation they choose. Additionally, UX researchers can augment their current approaches by adding our methodology at little (i.e. monitor-mounted eye tracker) to no (i.e. webcam eye tracker) cost, both financially and methodologically.

The webcam eye tracker was less accurate and had a lower temporal resolution than the dedicated eye tracker, but this did not impede data analysis or interpretation for our design. However, if high accuracy or temporal resolution is a necessity for the experimental design, webcam eye tracking may not be a suitable choice. For example, researchers interested in speed-accuracy trade-offs would likely benefit from higher-powered systems. When looking at the minimum gaze distance to any of the targets, we found that the webcam was consistently about 20 pixels (or 6.25 units[3]) away from the goal target in the Local cohort. An offset or threshold can boost accuracy, but target size matters. Improved eye-tracking algorithms may reduce the need for dedicated systems in future studies as spatial and temporal accuracy increases.

3.5.3 Moving beyond the laboratory

In general, we look at the data in the Remote group as an exercise in a cognitive ethology-based approach towards data collection [73]. Cognitive ethology calls for moving beyond some laboratory-based assumptions (e.g. cognitive process invariance and control) to better understand the naturalistic variance that occurs in real-world

[3]The collection environment was 800x450 units, and we can extrapolate pixels if we know the monitor resolution, which we did for the Local cohort.

complex tasks [72, 73, 135, 136]. In the Remote cohort, the noise in the collection environment did not appear to overwhelm the signal as evidenced by the similar findings between the studies. Arguably, some scientific findings lose sight of the real-world implications of their research; the environment in which the data were collected is not similar to how people actually behave. This approach challenges many of the control assumptions we as scientists make when conducting our studies. Is it critical that the users sit exactly 40cm from the screen? Does the lighting of the room need to be identical between participants? Our results suggest that, at least in this case, that may not be necessary. It is hard to know what the trade-off between data quality and real-life applicability truly is unless we are conducting ecologically valid studies in the first place. Perhaps it is reasonable to sacrifice data quality for the sake of having a closer understanding of the cognitive processes in the environment they are naturally practiced. This is particularly true for UX researchers interested in getting a deeper insight into how and where users *actually* use and encounter their products.

3.5.4 Privacy

Naturally, as we move towards more seamless data collection (i.e. the participant may not even realize their gaze is being tracked), privacy concerns arise. Many webcam eye tracking algorithms are collected and processed on the local machine (e.g. WebGazer [120, 137] and Labvanced [130]) or obtains the user's consent prior to recording [130] and thus require the end-user to be aware of its presence. The current iteration of the algorithm used in the present task also required inter-stimulus calibration periods, making it obvious to the end-user that their gaze was being recorded. However, in the future it is likely that gaze tracking models will become advanced enough to not require frequent re-calibrations and as a result may become functionally invisible to the end user. We believe that the end-user should always be made aware of (and asked for consent for) the recording of their data, as eye movement patterns can be used to identify unique individuals [138]. This is a major concern that will likely require legislation to capably handle.

3.5.5 Conclusion

Many qualitative methods are used simply because they make logical sense; who else would know better about the user's experience than the user themselves? We provide evidence that there's more than one way to find friction in a design, and it is not mutually exclusive to the current methods used. Eye and mouse tracking can provide a wealth of knowledge to the UX researcher at very little cost. The presented methods may currently be suitable for UX researchers, with the caveat that there is no standardized (or easy) way to analyze the data. Analysis requires intimate knowledge of data cleaning methods from eye and mouse trackers. Future use of eye and mouse tracking in UXR should include efforts to, ironically enough, improve the experience of the UX researchers through standardized tools for cleaning and analysis.

Acknowledgements

Thanks to everyone at BioWare for their enthusiasm and support. Thanks to Mitacs for the financial support.

References

12. Hayhoe, M. Vision Using Routines: A Functional Account of Vision. *Vis. cogn.* **7,** 43–64 (Jan. 2000).

13. Hayhoe, M. & Ballard, D. Eye movements in natural behavior. en. *Trends Cogn. Sci.* **9,** 188–194. ISSN: 1364-6613. http://dx.doi.org/10.1016/j.tics.2005.02.009 (Apr. 2005).

15. Chapman, C. S. *et al.* Reaching for the unknown: multiple target encoding and real-time decision-making in a rapid reach task. en. *Cognition* **116,** 168–176. ISSN: 0010-0277, 1873-7838. http://dx.doi.org/10.1016/j.cognition.2010.04.008 (Aug. 2010).

16. Wispinski, N. J., Gallivan, J. P. & Chapman, C. S. Models, movements, and minds: bridging the gap between decision making and action. en. *Ann. N. Y. Acad. Sci.* ISSN: 0077-8923, 1749-6632. http://dx.doi.org/10.1111/nyas.13973 (Oct. 2018).

19. Gallivan, J. P. & Chapman, C. S. *Three-dimensional reach trajectories as a probe of real-time decision-making between multiple competing targets* 2014. http://dx.doi.org/10.3389/fnins.2014.00215.

22. Freeman, J. B., Dale, R. & Farmer, T. A. Hand in motion reveals mind in motion. en. *Front. Psychol.* **2,** 59. ISSN: 1664-1078. http://dx.doi.org/10.3389/fpsyg.2011.00059 (Apr. 2011).

27. Goodale, M. A. Transforming vision into action. en. *Vision Res.* **51,** 1567–1587. ISSN: 0042-6989, 1878-5646. http://dx.doi.org/10.1016/j.visres.2010.07.027 (July 2011).

35. Goodale, M. A. & Milner, A. D. Separate visual pathways for perception and action. en. *Trends Neurosci.* **15,** 20–25. ISSN: 0166-2236. http://dx.doi.org/10.1016/0166-2236(92)90344-8 (Jan. 1992).

36. Milner, D. & Goodale, M. *The Visual Brain in Action* en. ISBN: 9780191545726. https://play.google.com/store/books/details?id=8CaQDwAAQBAJ (OUP Oxford, Oct. 2006).

57. Magnuson, J. S. Moving hand reveals dynamics of thought. en. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 9995–9996. ISSN: 0027-8424. http://dx.doi.org/10.1073/pnas.0504413102 (July 2005).

58. Goodale, M. A., Pelisson, D. & Prablanc, C. Large adjustments in visually guided reaching do not depend on vision of the hand or perception of target displacement. en. *Nature* **320,** 748–750. ISSN: 0028-0836. http://dx.doi.org/10.1038/320748a0 (1986).

66. SCCN. *Lab Streaming Layer* Mar. 2021.

72. Kingstone, A., Smilek, D., Birmingham, E., *et al.* Cognitive ethology: Giving real life to attention research. *Measuring the mind,* 341–357. https://books.google.com/books?hl=en&lr=&id=qTwVDAAAQBAJ&oi=fnd&pg=PA341&dq=Cognitive+ethology+Giving+real+life+attention+research+Kingstone+Smilek&ots=PfAQeFSBX5&sig=jz8-sBMnA6KfFHKbkNnbE5TMx7Q (2005).

73. Kingstone, A., Smilek, D. & Eastwood, J. D. Cognitive Ethology: a new approach for studying human cognition. en. *Br. J. Psychol.* **99,** 317–340. ISSN: 0007-1269. http://dx.doi.org/10.1348/000712607X251243 (Aug. 2008).

85. BioWare. *Mass Effect 3* Mar. 2012. https://www.bioware.com/games/#mass-effect-3.

116. Law, E. L.-C., van Schaik, P. & Roto, V. Attitudes towards user experience (UX) measurement. *Int. J. Hum. Comput. Stud.* **72,** 526–541. ISSN: 1071-5819. https://www.sciencedirect.com/science/article/pii/S1071581913001304 (June 2014).

117. Hinderks, A., Schrepp, M., Domıénguez Mayo, F. J., Escalona, M. J. & Thomaschewski, J. Developing a UX KPI based on the user experience questionnaire. *Comput. Stand. Interfaces* **65,** 38–44. ISSN: 0920-5489. https://www.sciencedirect.com/science/article/pii/S0920548918301247 (July 2019).

118. Huang, J., White, R. & Buscher, G. *User see, user point: gaze and cursor alignment in web search* in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Association for Computing Machinery, Austin, Texas, USA, May 2012), 1341–1350. ISBN: 9781450310154. https://doi.org/10.1145/2207676.2208591.

119. Papoutsaki, A., Laskey, J. & Huang, J. *SearchGazer: Webcam Eye Tracking for Remote Studies of Web Search* in *Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval* (Association for Computing Machinery, Oslo, Norway, Mar. 2017), 17–26. ISBN: 9781450346771. https://doi.org/10.1145/3020165.3020170.

120. Papoutsaki, A. *Scalable Webcam Eye Tracking by Learning from User Interactions* in *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems* (Association for Computing Machinery, Seoul, Republic of Korea, Apr. 2015), 219–222. ISBN: 9781450331463. https://doi.org/10.1145/2702613.2702627.

121. Hauger, D., Paramythis, A. & Weibelzahl, S. *Using Browser Interaction Data to Determine Page Reading Behavior* in *User Modeling, Adaption and Personalization* (Springer Berlin Heidelberg, 2011), 147–158. http://dx.doi.org/10.1007/978-3-642-22362-4_13.

122. Land, M. F. & Hayhoe, M. In what ways do eye movements contribute to everyday activities? en. *Vision Res.* **41,** 3559–3565. ISSN: 0042-6989. http://dx.doi.org/10.1016/s0042-6989(01)00102-x (2001).

123. Gallivan, J. P., Chapman, C. S., Wolpert, D. M. & Flanagan, J. R. Decision-making in sensorimotor control. *Nature Reviews Neuroscience* **19,** 519–534 (2018).

124. Stone, S. A. *et al.* Generating accurate 3D gaze vectors using synchronized eye tracking and motion capture. en. *Behav. Res. Methods.* ISSN: 1554-351X, 1554-3528. http://dx.doi.org/10.3758/s13428-022-01958-6 (Sept. 2022).

125. Nunnally, B. & Farkas, D. *UX Research: Practical Techniques for Designing Better Products* en. ISBN: 9781491951262. https://play.google.com/store/books/details?id=sfOeDQAAQBAJ ("O'Reilly Media, Inc.", Nov. 2016).

126. Guo, Q., Jin, H., Lagun, D., Yuan, S. & Agichtein, E. *Mining touch interaction data on mobile devices to predict web search result relevance* in *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval* (Association for Computing Machinery, Dublin, Ireland, July 2013), 153–162. ISBN: 9781450320344. https://doi.org/10.1145/2484028.2484100.

127. Westwood, D. A. & Goodale, M. A. Perceptual illusion and the real-time control of action. en. *Spat. Vis.* **16,** 243–254. ISSN: 0169-1015. http://dx.doi.org/10.1163/156856803322467518 (2003).

128. Ballard, D. H., Hayhoe, M. M., Li, F. & Whitehead, S. D. Hand-eye coordination during sequential tasks. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **337,** 331–339 (1992).

129. Deng, S., Chang, J., Kirkby, J. A. & Zhang, J. J. Gaze–mouse coordinated movements and dependency with coordination demands in tracing. *Behav. Inf. Technol.* **35,** 665–679. ISSN: 0144-929X. https://doi.org/10.1080/0144929X.2016.1181209 (Aug. 2016).

130. Finger, H., Goeke, C., Diekamp, D., Standvoß, K. & König, P. *LabVanced: a unified JavaScript framework for online studies* in *International Conference on Computational Social Science (Cologne)* (2017). https://www.researchgate.net/profile/Caspar-Goeke/publication/322273524_LabVanced_A_Unified_JavaScript_Framework_for_Online_Studies/links/5a4f7ac64585151ee284d8c2/LabVanced-A-Unified-JavaScript-Framework-for-Online-Studies.pdf.

131. jamovi. *jamovi (version 2.2.5)[Computer software]* 2022. https://www.jamovi.org/.

132. Mejtoft, T., Hale, S. & Söderström, U. *Design friction* in *Proceedings of the 31st European Conference on Cognitive Ergonomics* (2019), 41–44.

133. Bridgeman, B. Conscious vs Unconscious Processes: The Case of Vision. *Theory Psychol.* **2,** 73–88. ISSN: 0959-3543. https://doi.org/10.1177/0959354392021004 (Feb. 1992).

134. Wisiecka, K. *et al. Comparison of Webcam and Remote Eye Tracking* in *2022 Symposium on Eye Tracking Research and Applications* (2022), 1–7.

135. Chisholm, J. D. *et al. A Cognitive Ethology Study of First- and Third-Person Perspectives* 2014. http://dx.doi.org/10.1371/journal.pone.0092696.

136. Smilek, D., Birmingham, E., Cameron, D., Bischof, W. & Kingstone, A. Cognitive Ethology and exploring attention in real-world scenes. en. *Brain Res.* **1080,** 101–119. ISSN: 0006-8993. http://dx.doi.org/10.1016/j.brainres.2005.12.090 (Mar. 2006).

137. Yang, X. & Krajbich, I. *Webcam-based online eye-tracking for behavioral research* Dec. 2020. http://dx.doi.org/10.31234/osf.io/qhme6.

138. Holland, C. & Komogortsev, O. V. *Biometric identification via eye movement scanpaths in reading* in *2011 International Joint Conference on Biometrics (IJCB)* (Oct. 2011), 1–8. http://dx.doi.org/10.1109/IJCB.2011.6117536.

Chapter 4

Eye tracking in the clinic: an investigation of an easy-to-use automated clinical eye movement assessment

Abstract

Vertigo is a type of dizziness that makes people feel as if they are moving or spinning when they are not. It is caused by problems in the vestibular system, usually in the inner ear or in the brain. Vertigo is a major cause of visits to emergency departments (around 2-3 % of all visits). Accurate diagnosis is important to guide treatment, as patients with vertigo from stroke (i.e. central vertigo) benefit from rapid therapy and stroke prevention treatment, whereas vertigo caused by inner ear problems (peripheral vertigo) can benefit from other specific therapies. However, challenges in the evaluation of vertigo exist: 1) the physician exam may not be sensitive or standardized enough to detect subtle but crucial indicators, and 2) the expertise needed to perform and interpret the assessment may not be present. Eye-tracking technology can augment clinical assessment of the patient, providing additional information to the physician. Here, we assess the feasibility of leveraging advances in eye-tracking technology to develop a system for assessing vertigo, thereby improving the diagnostic process in the emergency department and the clinic. The intended goal of this project is to determine what it will take to create a system capable of collecting and analyzing eye movements in patients with vertigo in acute care settings to improve the speed and accuracy of vertigo diagnosis. While we did find that an uncontrolled environment is capable of generating high quality data, it is important to consider the medical state of the participant to ensure continuous data collection. We found that our device was capable of collecting and calculating eye tracking metrics that are usable by a trained clinician for augmenting diagnosis. The device was easily deployed by a prior medical professional with no eye tracking training, suggesting that it can be deployed in many different use cases. Overall, we discuss the findings, limitations, and future directions for such a device to be successfully deployed in an uncontrolled environment.

4.1 Introduction

Vertigo is a sub-type of dizziness caused by vestibular system imbalance. Accurate diagnosis of vertigo often requires the expertise of a neurologist [139, 140]. Vertigo is a major burden for patients and the health system, with more than 4 million visits in US emergency departments, making up an estimated 2-3% of all visits [141–145]. Broadly speaking, there are two main categories of vertigo: peripheral and central, classified based on etiology. Peripheral vertigo may be caused by variable pathology within the vestibular organs, which induce sensations of movement. Examples of peripheral vertigo include benign paroxysmal positional vertigo, vestibular neuronitis and Menniere's disease. These diseases, while symptomatically similar across patients, tend to be difficult to diagnose because they present so similarly. Central vertigo, a much more serious diagnosis, is caused by injury to brain areas such as the brainstem, cerebellum, or thalamus. Central lesions (i.e. stroke) can lead to poor outcomes or death if left undiagnosed and untreated. Clearly, this is not something a well-trained physician would want to mix up. Current best practices for vertigo diagnosis include a clinical examination (e.g. the HINTS test: head-impulse, nystagmus, test-of-skew [146–148]) and high resolution brain imaging. When dealing with vertigo patients, central and peripheral causes can be easily conflated, making it much more difficult to provide an accurate diagnosis. However, discerning a deadly etiology of vertigo (e.g. brainstem stroke) from a non-life threatening peripheral cause therefore relies on the availability of computed tomography (CT) or magnetic resonance imaging (MRI) scanners, and unfortunately, even these modalities have a high early false negative rate for central vertigo [146]. If a tool can easily supplement clinical examinations with relevant metrics, it has the potential to enhance patient outcomes, making it a valuable asset.

The current gold standard for diagnosing vertigo is the HINTS exam: a three-step bedside oculomotor examination [146] consisting of: 1) the head impulse, 2) nystagmus assessment and, 3) test of skew. For the head impulse test, the patient is asked to fixate on a stationary landmark in the room. The examiner, using their hands, will jerk the patient's head in a pseudo-random order to evoke gaze corrections. For the nystagmus assessment, the patient fixates on a moving object (such as the examiner's finger) while holding their head still as the object is moved laterally back and forth several times. Typically, the examiner will pause at the extremes of gaze, as nystagmus can be exaggerated at these positions. For the test of skew, each eye is covered individually, and any changes of gaze or re-fixations are noted, as this can be a sign of a pathological etiology. The key output of the HINTS test is evidence of the vertigo being either central or peripheral in origin, where evidence is more clear in high- or medium-risk patients, but caution should be applied in low-risk populations [149]. When performed correctly, the HINTS exam has been shown to have an almost 98% sensitivity for ischemic stroke. However, despite its efficacy, the HINTS exam is not always used. The HINTS exam may not be performed simply due to a perceived lack of time [148, 149], or due to concerns of causing additional damage (e.g. vertebral dissections, [150]). A study by McDowell & Moore [148] found that only 5% of vertigo patients had the head impulse test performed, a discrepancy the authors attribute to physician unfamiliarity or a preference for relying on patient history. Given that the bulk of the HINTS exam is assessing what the eyes are doing during the perturbations, it should be possible to non-invasively record the eyes. In using such a device, you should be able to achieve the high sensitivity of the HINTS exam while overcoming its barriers to usage, thereby improving vertigo diagnoses.

Promisingly, recent advancements in eye tracking technology has resulted in an increased availability of mobile eye trackers, potentially suitable for use in the clinic. Eye trackers are capable of producing an objective assessment of eye movements, including direction, frequency, and speed. Eye trackers have historically been very expensive (~$50,000+) bulky laboratory equipment that require specialized training to use and analyze the resultant data. Recently, affordable (<~$5,000) portable eye trackers such as the Pupil Labs Core [50] have become available, permitting use outside of the laboratory and specialized testing centers. These eye trackers are capable of recording pupil positions at high frequencies and resolutions, making them comparable to the expensive systems but much more portable. Such a system, working within the confines of a high stress emergency department, could be used to augment the vertigo clinical exam. Currently, very few studies have attempted to investigate the efficacy of such a system, owing to the fact that there are a number of barriers to its deployment. First, eye trackers are complex devices that usually require advanced training to use, and many medical professionals do not have this training. Second, objective assessments may be difficult as emergency room consults tend to be high-stress dynamic situations, and uncontrolled environments can make assessment that much harder. Finally, vertigo patients are likely to be difficult to collect eye movement data from for a multitude of reasons, including the fact that they may be unable to open their eyes for long enough periods of time to record high quality data. In the current study, we address these issues in turn to test the feasibility of deploying mobile eye-tracking a diagnostic aid for vertigo when it presents in the emergency department.

It is also important to acknowledge a barrier that is largely outside of our control—the willingness of clinical experts to adopt new technology. While the use of automation and systematic assessments has improved the efficacy of many health care systems over the past half century, many physicians are still wary of new technology being used for diagnostic purposes, especially when it comes to liability [151, 152]. Although these concerns are valid, there are cases where the benefits of a new approach are so evident that it can overcome hesitancy. For example, digital blood pressure monitors are preferred over aneroid (mechanical) monitors, despite being slightly less accurate [153]. This is because human interpretation can factor in the general error when evaluating the results. Furthermore, the minor perceived error is outweighed by the improved efficiency gained from using a digital device, effectively rendering the error insignificant. It is crucial to take this perspective into account when utilizing tools like eye trackers in clinical spaces. By being aware of their limitations and not relying solely on the device for diagnosis, greater efficiency can be achieved. We believe that eye trackers, when used and deployed similarly to how we use digital blood pressure monitors, are capable of providing valuable diagnostic information while requiring minimal training. One problem though, is eye trackers typically require extensive training for effective use.

To effectively deploy a rapid clinical assessment tool, we need to recognize that it is untenable to expect every end-user to be extensively trained on research-typical eye tracker use. Because such a tool would necessarily be used in high-stress scenarios (e.g. ambulatory and emergency department care), the end-user (i.e. data collector) will not have time to do any level of troubleshooting if anything were to go wrong. As such, a goal of the present study was to specifically recruit a person without any eye tracking training to act as the primary data collector. We chose to use a regular healthcare worker with no prior eye tracking experience to challenge the claims of many modern eye tracker manufacturers who claim ease of use, portability, and general use cases outside of the laboratory. If this is truly the case, we should be able to see high quality data collected by a non-expert in non-laboratory settings. However, if data collection proves difficult, the data quality should be relatively low.

Assuming a task can be designed that can be run by a non-expert, it is still important that the task captures the key features of impairment that the typical clinical exam is testing for. In the case of a typical neurological evaluation of a patient, this involves multiple eye movement tests [154, 155]. The objective of these tests is to identify any impairments, as they can help pinpoint the location of brain damage, if it is present. For instance, a person who cannot smoothly pursue a moving object may have damage to the cerebro-ponto-cerebellar pathway [156]. Although this encompasses many brain regions, determining that the problem is neurological can be critical for patient care and potentially life-saving. While many of these tests are easy to administer, some can be challenging to assess accurately. For instance, a physician's observation of smooth pursuit eye movements may not capture important quantitative information such as average velocity or aberrant eye movements. Additionally, a physician can be distracted by their environment and miss important details without any specific fault. By contrast, an eye tracker can specifically capture this information. Therefore, in addition to targeting the deployment of our tasks to a non-expert, it was equally important that we designed tasks that matched key clinical features. Of course, the clinic is unlike the laboratory-based settings that eye tracking is collected in.

As we think about moving into the clinic, it is important to be cognizant of the critical differences that may affect good experimental design. For instance, clinics are much more chaotic than laboratories—who tend to control many of the environmental aspects of the room. While some studies have assessed eye tracker use outside of the lab [157, 158], the purpose of many of these studies was to collect naturalistic data from a normative population. For example, Hessels *et al.* equipped participants with an eye tracker while they navigated through several crowds to investigate eye gaze behaviours and how social affordances (e.g. eye contact) can modulate gaze behaviours. They found that, despite a lack of control on the environment, it was still possible to maintain a significant degree of control over the experiment while allowing the subject to maintain many degrees of freedom. This suggests that eye trackers (and perhaps many other tools of the lab) could be deployed outside of highly-controlled laboratory spaces. However, some environments may prove to be particularly challenging such as clinics or emergency departments.

To summarize, with the advancement in technical and physical capabilities of modern eye trackers, deployment in clinics should be possible. However, there are many challenges to working in a busy clinical environment. For example, eye tracking experience is rare outside of research laboratories, meaning a the device should be usable by a non-expert. Another challenge arises from even being able to quantify clinically relevant measures. An assessment device should be able to quantify aspects of a neurological exam in a way that is intuitive the attending physician such that it can be used to augment their diagnosis. However, uncontrolled settings such as clinics and emergency departments can mean high quality data is difficult to collect for reasons beyond the data collector's control. High traffic, loud noises, and salient distractors can make it difficult for both the data collector and the patient to give their full attention to the task. Clearly, a diagnostic rapid assessment tool should overcome these barriers if it is to be considered for use in a clinical setting.

To test how well such a system can be deployed in a busy clinical environment, we explored collecting diagnostic eye tracking data in the clinic—an uncontrolled environment, and with data collection performed by a health care worker with no prior expertise in eye tracking. Here, we created a simple digitized test battery akin to what an attending physician would administer to a suspected vertigo patient. The primary goal was to collect data and calculate useful metrics to help differentiate between peripheral and central vertigo. We assessed the test battery on control participants, and attempted to validate the data on vertigo patients. The patient group were all patients who were admitted to the University of Alberta Hospital with a chief complaint of dizziness (see Table 4.1 for a description of the patients). The data were collected shortly after the patient was admitted to the hospital. A control group allowed us to see if the collection was affected by the data collector. We expected that data quality would be high for both groups, with the patient group having slightly noisier data. In order to determine feasibility of such a system, we aimed collect data outside of laboratory settings in both the control and patient groups.

4.2 Materials and methods

To overcome the difficulties of implementing a rapid assessment tool in a clinical setting, we developed a neurological test battery that was derived from clinical practice and designed for ease of use and deployment. We devised a portable test battery that can be set up at a patient's hospital bed. The test was run on a laptop computer and consisted of 5 sub-tasks that utilized a mobile eye tracker to record eye movements. Data were collected by a single retired health care professional to assess usability and portability.

4.2.1 Task & system

We created a task that was designed to be deployed on a portable laptop. The task itself contained 5 sub tasks akin to the neurological tests a physician would conduct on a patient. Using this system, we recorded the pupils of the participant for the duration of the experiment using a mobile eye tracker. The system itself was designed to be portable, such that all of the equipment could be stored in a backpack so it was ready to go when needed. Additionally, it was designed to be easy to use with minimal training required for use.

4.2.2 Equipment

We use a Pupil Labs Core eye tracking headset (pupil-labs.com; [50]), a low-cost, light-weight, mobile, open-source hardware and software solution that records the pupils at up to 200 Hz. The headset is designed to allow for free movement of the head and does not require a chin-rest bar. The Core software records the data and allows for pupil positions to be passed into Lab Streaming Layer (LSL; https://github.com/sccn/liblsl). LSL is a middle-layer software that handles synchronization between data streams to ensure all data use a common timestamping source and do not drift over time. LSL is capable of taking an arbitrary number of data streams (in our case, the eye tracking data and stimulus labels) and synchronizing them, ensuring that all of the data are timestamped and do not drift relative to each other. LSL was used to align timestamps from when stimuli appear on screen to the timestamps of the eye tracking data. This step is critical to ensure eye tracking data are synchronized to when stimuli appear on screen. Importantly, we only recorded the position of the pupils, and not gaze data (i.e. the position in space being looked at).

The task was deployed on a laptop computer. The specific laptop was a Dell XPS 15 7590 with a i7-9750H processor and a 15" screen (1920x1080 resolution). We chose this laptop because it had a reasonably powerful processor, which was necessary for running the eye tracker program, and because it was lightweight (1.81kgs) and fit in a backpack. For the control participants, the laptop was placed in front of them on a dining table at a measured distance of 40cm. For the patients, it was placed on the bedside table rotated to be in front of the patient at a measured distance of 40cm.

4.2.3 Eye tracking procedure

A battery of five procedures was developed that approximates what an attending neurologist may administer to a vertigo patient (for more detail see Goebel [154]. We aimed for a contact-free exam, since this was identified as a potential issue for why the HINTS exam does not see widespread use. We attempted to replicate the procedures that are typically performed manually by the attending physician, who may sometimes use simple props such as a pen to facilitate the test. Each procedure was explained verbally to the participant just prior to starting each task. Upon confirming the participant understands, the test begins with the experimenter pressing the space bar on the laptop computer. The following descriptions first explain the task a physician would use to measure this behaviour, followed by details of our implementation.

Stare fixation

The physician will ask the patient to fixate on a stationary object (such as the physician's finger) while observing eye behaviour. The fixation period typically lasts for around 20s. The purpose of this test is twofold: 1) to determine if the patient is capable of simple fixation and 2) to assess presence of nystagmus.

Our implementation of this task is shown in Figure 4.1A, where the patient would fixate on a centerally presented white cross on a black background for approximately

20s. The cross was 10 pixels wide and 10 pixels tall with a thickness of 4 pixels.

Smooth pursuit

The physician usually performs this test by moving their finger smoothly from left to right or up to down while the patient maintains fixation on the tip. Finger movement is typically paused at the extreme peripheries of vision (i.e. far left, right, top, or bottom hemifields) to observe if nystagmus is present.

Our implementation of this task is shown in Figure 4.1B[1]. The patient will track the dot as it smoothly moves to each of the end locations on the screen, where it pauses. For the first part of the task, the dot will move to the edge and immediately move back to the centre of the screen. For the second part, the dot will move to the edge and hold its position.

The target shown on screen was a white dot on a black background with a diameter of 20 pixels. For each location, the dot would move from the centre of the screen to the extremes of the screen over the course of 2 seconds at a constant velocity. For example, for the left movement, the dot would start in the centre of the screen and move to the left side of the screen over the course of 2 seconds. During the hold portions of the task, the dot would hang at the extreme position for 3 seconds before moving back to the centre of the screen.

Target jump

Using two fingers on opposite sides of the patient's field of view, the physician will ask the patient to fixate—through a saccadic movement of the eyes—on the noted finger. This can be performed by simply audibly saying instructions (e.g. "Left, right, left") or through salient movements (i.e. wiggling fingers). Here, the ability to perform the action as well as the delay and inaccuracy (overshoot) of the targeted jump can be important clinical markers.

[1]Note that only horizontal movements are shown here, but the task also included a vertical portion as well.

Our implementation is shown in Fig. 4.1C, where a centre cross appears and shortly after a stimulus (dot) will appear on either the left or right side of the screen, which the patient must fixate on. This repeats 20 times.

The target shown on screen is a white dot on a black screen that has a diameter of 20 pixels. The dot alternates between the left and right side of the screen, with a fixation cross shown in between movements. The dot's position was determined by dividing the width of the screen in pixels by 20 and positioning the dot either at the beginning or the end of this 1/20th interval. For example, for a screen width of 1080 pixels, the dot would be placed at either pixel 54 or 1026. The dot stays on the screen for a pseudorandom amount of time sampled from a uniform distribution between 0.5s and 1.5s.

Vestibulo-ocular response (VOR)

The VOR allows for someone to move their head while still maintaining fixation on a target. Similar to the Fixation test, the physician will hold their finger in a static position, while asking the patient to fixate but simultaneously turn their head from left to right, then up and down in a cross-like pattern.

Our implementation is shown in Figure 4.1D. A white centre cross is drawn on the black screen, and the patient is instructed to maintain fixation on the cross while turning their head. The centre cross is the same size as the cross used in the Stare task.

Brightness

The eyes' ability to adjust to light is a critical function of the eye. A physician may use a flashlight or flicker the lights and see if the patient's pupils respond, or if there is a disparate response between the eyes. This test can determine if nerve damage has occurred (e.g. CN III - oculomotor) or if there are disparate responses between the eyes.

Our implementation is shown in Figure 4.1E. A black background with a white cross is shown for 5s, after which the background turns white, eliciting a pupillary constriction. The centre cross is the same size as the cross used in the Stare task.

4.2.4 Motivation

While all of the tests listed above are typically performed by either an attending emergency physician or neurologist, they can be incredibly hard to quantify; how much did the eyes move? How quickly did the pupils contract to light? How well did the patient track the moving finger tip? Additionally, it can be difficult to be systematic in stimulus presentation; did the physician always move their finger at the same velocity? Did they use the same flashlight? Did they move their finger the same distance? The purpose of standardizing these tests is that it allows for the result to be quantified. Moreover, standardizing the tests allows us to standardize and automatically extract measures from the resultant data files. This allows for two key advantages: 1) it makes the data collection much simpler and 2) results can be generated instantly at the end of the task.

We created versions of each of the above tests in software, made to be presented on a laptop computer. For a description and visualization of how these stimuli looked to the participants, see Figure 4.1[2].

4.2.5 Data collection

Data collection was performed by a 59 year old female retired physician who specialized in the care of geriatric individuals. She had extensive experience communicating with patients, but did not have any prior experience with eye tracking or data collection for the purposes of a scientific study. The data collector was given approximately 45 minutes of training for fitting the eye tracker and ensuring reasonably good data quality prior to beginning the study. Between data collection sessions, general advice

[2]The code for the stimuli generation can be found at https://github.com/scottastone/StrokeNystagmusAnalysis

and guidance was given as necessary, for example such as if technical issues arose (e.g. eye tracker not connecting to the computer, general computer maintenance, or questions about data quality). In general, the data collector worked independently and was able to collect the data with little guidance.

4.2.6 Software

A custom Python program was written to present the stimulus to the participant. Two key guiding principles were used when considering the development of this program: 1) the stimulus should be very easy to deploy—starting the program should be as simple as clicking to start and 2) the program should run ”on-rails”, meaning it should require very little input from the data collector. Adjustments to the user experience aspects of the program were made such that the data collector simply has to place the computer 40cm in front of the participant, fit the eye tracker, and start a single program to collect the data. Prior to beginning each task, verbal explanation of the task is given to the participant. The data were then automatically saved, organized, and uploaded to a secure server such that the data collector did not have to do anything post-collection.

4.3 Environment

One of the key contributions of the present study is the use of an uncontrolled environment to collect data. Typically, eye tracking studies will be deployed in controlled environments such as laboratory settings. In this case, it is critical that an uncontrolled environment is used (e.g. emergency department, hospital room, the home of an individual) to truly assess the usability of the device. It is important that the device is tested in the same way it is intended to be used in the real world. In this way, we hope to be more ecologically valid in the sense that the measures we are deriving from our data more closely match how a physician would actually use this tool and the kind of data that a patient would actually provide.

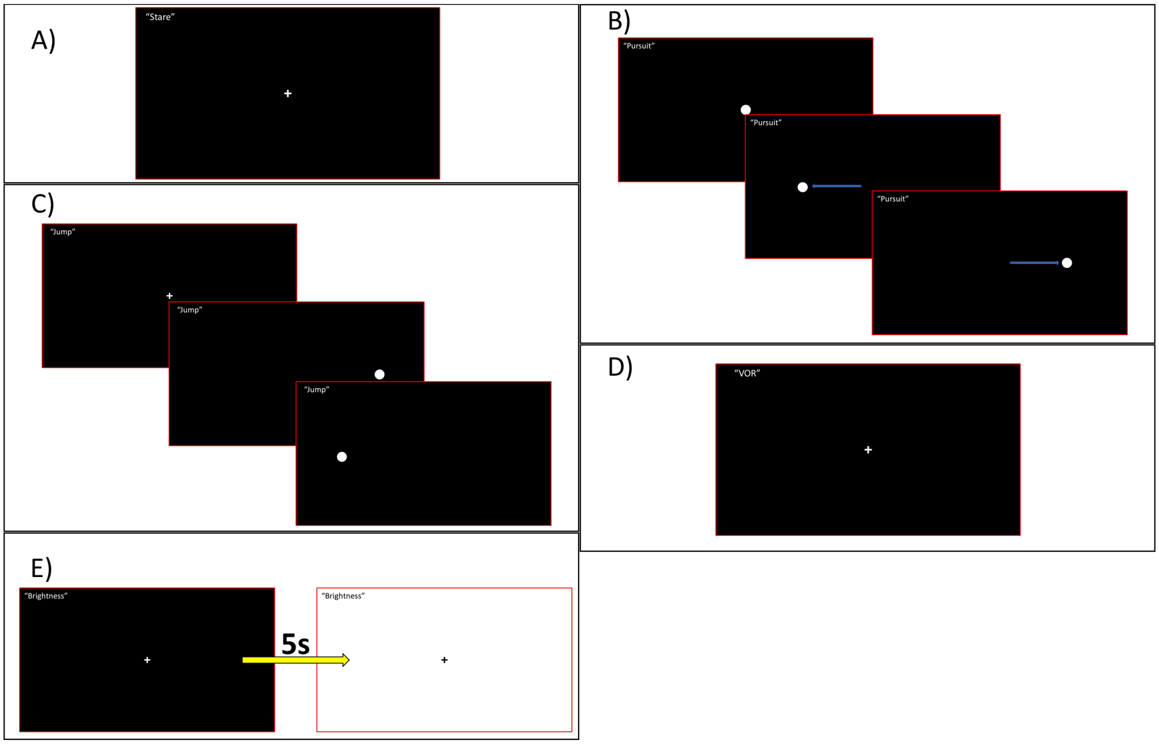


Figure 4.1: Each of the stimuli presented the patient, as described in section 4.2.3. Each stimulus represents what would be presented on the screen. A) The stare condition: a fixation cross is drawn to the centre of the screen, which the patient fixates on for 20 seconds. B) The pursuit condition: a dot appears on the screen and moves to the left, right and eventually up then down. This repeats a second time, but the dot will hold at its terminal position for approximately 1.75 seconds. Note that only left and right movements are shown here for brevity. C) The target jump condition: a fixation cross is drawn to the centre of the screen, and after a short time (approx 2s), a dot is drawn on either the left or right side of the screen, which the participant is instructed to shift their gaze towards. After the dot disappears, the participant fixates on the centre cross again, which reappears. D) The vestibulo-ocular response (VOR) condition: a centre fixation cross is drawn to the screen, which the patient is told to lock their gaze to. With their gaze locked on the cross, the patient turns their head left, then right, then tilts up then down. E) The brightness condition: the patient fixates on the centre cross and after approximately 5s, the screen turns white, testing the patient's pupillary response times.

4.4 Participants

Ethical approval was granted by the University of Alberta Human Subject Committee (PRO00066577).

4.4.1 Controls

Data from fifteen control participants was collected (7 female, mean age: 54.2 ± 21.4 years) in an uncontrolled setting—the home of the data collector. All participants gave informed consent to participate in the present study.

4.4.2 Patients

Data from seven patients was collected (2 female; mean age 56.7 ± 14.5 years) at the University of Alberta Hospital. All patients gave informed consent to participate in the present study. All patients were admitted with a chief complaint of vertigo. A summary of the patient data can be found in Table 4.1.

Patient ID	Sex	Age	Nystagmus	HINTS	Stroke	Diagnosis
P1	M	39	No	No	Yes	Left PICA stroke
P2	M	70	Yes	No	No	Peripheral vertigo
P3	F	45	Yes	Yes	No	Peripheral vertigo
P4	M	39	Yes	No	Yes	Right lateral medullary stroke
P5	M	63	No	No	Yes	Intraparenchymal hemorrhage
P6	M	63	No	No	Yes	Right PICA stroke
P7	F	78	No	No	No	Cranial nerve VI palsy

Table 4.1: A description of all admitted patients in the study. General information about sex and age were collected, as well as whether nystagmus was present. The HINTS column marks whether or not the battery was used on the patient. Stroke presence and the appropriate diagnosis is listed in the last 2 columns.

4.4.3 Data

Eye data

During each of the above procedures, the pupils of both eyes were recorded as X,Y coordinate pairs over time. All of the data were pre-processed using a custom Python script. In short, the data were recorded to the XDF file format, which contained both the eye position data (with timestamps) as well as the named stimulus labels to demarcate each of the stimuli in the methods above.

Cleaning

Raw pupil position data was cleaned by: 1) Removing any data points outside of pupil camera bounds (<0 or >1); 2) Removing any data points more than 4 standard deviations away from the mean position; 3) Removing any data points with velocities greater than 6 (meaning the pupil was travelling across the entire camera 6 or more times per second). After this removal, any gaps < 50 ms were filled using the *inpaint_nans* [111] function in MATLAB then, any remaining islands of data < 50 ms were deleted. Finally, the pupil data were filtered in MATLAB using a 4^{th} order zero-lag low-pass Butterworth filter with a cutoff frequency of 10 Hz. A 10 Hz cutoff was chosen because the demands of the tasks do not depend on eye dynamics with movements more than 10 times per second.

4.5 Analysis

The stimuli were analyzed to derive measures that would be intuitive to an attending clinician and match the features they'd typically look for in a manually administered exam. An analysis procedure will be described for each of the stimuli.

4.5.1 Stare

For the Stare task, the participant was asked to simply fixate on a cross presented in the centre of the screen for 20 seconds. During this time, the XY position of each pupil was recorded. To analyze this data, a dispersion calculation algorithm was used. The dispersion of the pupil locations over time gives a description of how still the eyes are able to stay during a fixation. Below is a general description of how the dispersion of the Stare data was calculated.

1. We calculated an error ellipse representing the uncertainty of the data at a specified level of confidence. It takes three inputs: a confidence interval (as a chisquare value), mean, and a covariance matrix.

2. The eigenvalues and eigenvectors of the covariance matrix encode the size and orientation of the error ellipse, with the angle between the largest eigenvector and the x-axis.

3. The function returns the center, size and angle of the error ellipse. The size of the ellipses are interpreted as the major (i.e. width) and minor (i.e. height) axes.

Error ellipses were calculated, and the sizes of the width and height of the ellipse were used to give an intuitive representation of how much the eyes remained stationary during fixation. The units are normalized between 0 and 1 along each dimension. Here, a high value (e.g. > 0.1) would indicate that 95% of the data can be found within 0.1 of the mean. In general, a low dispersion value along each dimension indicates that the ellipse fit tightly to the data, and the eyes were relatively stationary throughout. A clinician could interpret this value as: a lower value to zero being indicative of normative behaviour, and a higher value being indicative of a potential oculomotor problem (e.g. nystagmus). The width and height of the ellipse are rotated according to the angle of the camera that recorded the eye movements. It is important to note that due to the arbitrary nature of the camera placement and ellipse fitting routine, the labels of *width* and *height* are arbitrary, but used to be descriptive of the axes.

Clinical relevance

Stable gaze fixation is a common test used during a neurological assessment in both adults and pediatric vertigo patients [154]. When a physician asks a patient to fixate on a stationary object in space, they are observing the pupils for signs of nystagmus and other motor disorders that prevent stable fixation [154, 159]. If the eyes are not able to hold fixation, the spread of the pupil position over time will be wide. The width and height values of the ellipses are quantified values that a clinician can use to assess how stable the gaze is: the lower the value, the more stable the fixation.

Additionally, a larger error in one of the axes can be indicative of a specific neurological etiology. For example, upbeat nystagmus can be caused by damage to the brainstem (medulla and pons) or the midbrain [160]. It could potentially be useful to compare the size of the axes to one another as a way to quantify how much movement is occurring in either horizontal and vertical space. Since the data are rotated arbitrarily (by the ellipse fitting algorithm), it would be necessary to visually inspect the eyes during movement to determine which axis belongs to horizontal and vertical movement.

Examples of what the ellipse fitting algorithm looks like in practise can be found in Section 4.5.1. Here, we can see four different test cases for the dispersion calculation. In general, the control data appear to be tightly fit. We show exemplar good and poor quality patients to show the general trend of data quality. The patient (good) data are much more diffuse, with clusters of points appearing further apart, but the ellipse still fits the data reasonably well. The patient (poor) data are extremely far apart, leading to a very poor fit. Finally, uniformly sampled random noise (0 to 1) is used to demonstrate a good fit with poor quality underlying data. The spread of the data results in a very large width and height, spanning the entire capture area.

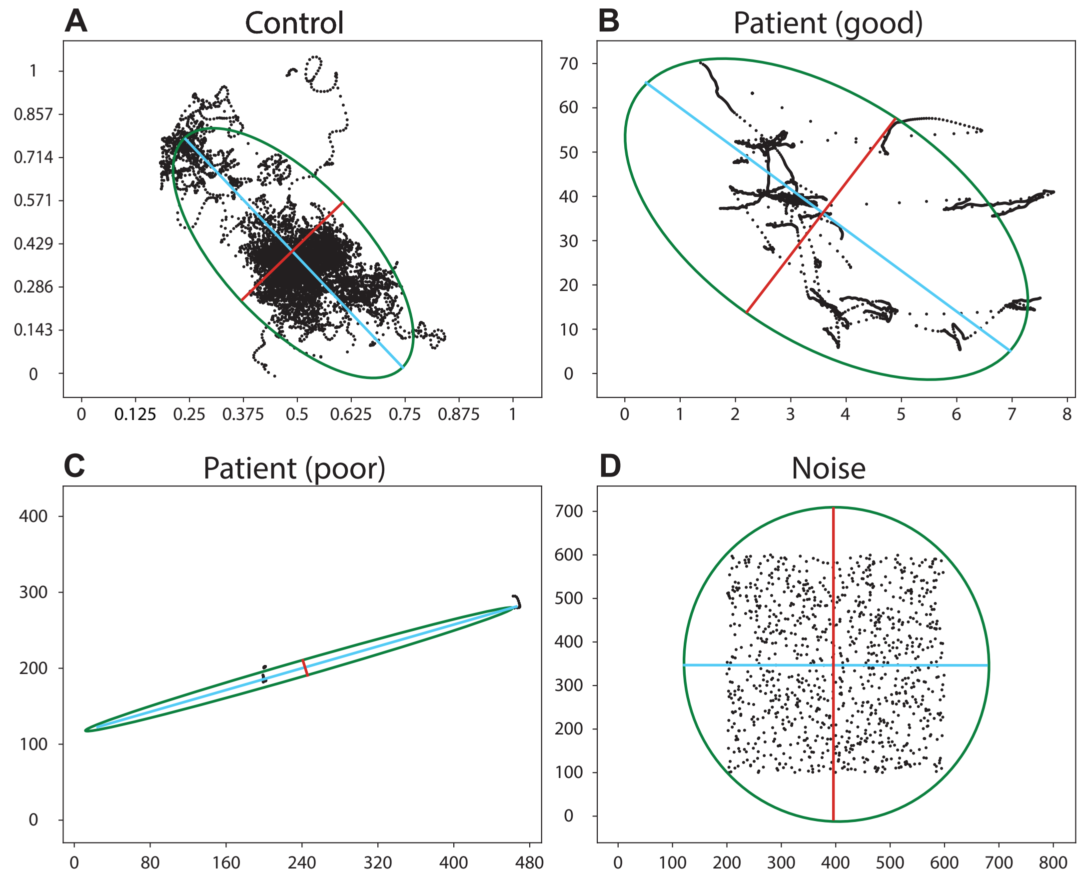


Figure 4.2: Here, four plots of eye dispersion data are shown. The X and Y axis of each plot are scaled to the control data to allow for easy reference. Individual pupil positions are plotted in black, represented in black, with a green error ellipse indicating the uncertainty at a 95% confidence level. For the top row, the blue line is the height of the ellipse, and the red line is the width. A) Control data is shown, scaled between 0 and 1 on each axis. This is used as a reference to demonstrate the scale difference between control and patient data. B) Exemplar "good" patient data is shown. Here, we can see that the X axis has to be scaled up 8x to plot the data and ellipse. The Y axis has to be scaled up over 70x to capture all of the data. C) Common patient data, categorized as "poor". Here, the scale of the X and Y axes must be scaled to 480x and 400x respectively to effectively plot the data. The control data plotted at this scale would be too small to effectively see. D) Noise reference data to get a sense of the worst possible case. Here, we see the X and Y axes must be scaled 800x and 700x respectively to plot the data. The noise data are sampled from a uniformly distributed sequence. Again, the control data would be too small to notice plotted at this scale.

4.5.2 Pursuit

The ability to smoothly pursue an object in space requires tight oculomotor control, driven primarily by cerebrocortical areas of the brain [156]. Inability or deficits in smooth pursuit are indicative of a problem with either a motor nerve or brain function, potentially giving evidence for centrally derived vertigo. Here, we assess two main components of smooth pursuit: velocity stability and the ability to start and stop smooth pursuits. The pursuit task has two parts: 1) following the smooth movement of the object and 2) fixation on the object. To capture these two aspects, we calculated the velocity during movement and fixation, as well as the dispersion during fixation. Velocity was calculated as the change in position over time. Dispersion was calculated during fixation identically to the Stare task.

In general, a low relatively constant velocity should be expected as the eyes smoothly track the object on the screen. Spikes or drops in velocity are indicative of oculomotor problems, suggesting the subject is having trouble maintaining fixation. The velocity of the cleaned XY position data per eye was computed by taking the norm of the gradient of the position data. For the dispersion, we should expect to see a tight dispersion (and elliptical fit) during the fixation portion, indicating the eyes were still.

Clinical relevance

When assessing pursuit, the eyes should be able to steadily track a moving object without having to 'catch up' or jump ahead. Assessing smooth pursuit gives insight into the brain's ability to predict where an object is going to be, and the ability to start and stop eye movements. A trained physician would look at the patient's eyes to see if their eyes are smoothly moving whilst tracking the object, and note any jumps or lags [154, 161]. Additionally, they will also assess the ability to hold their gaze at extreme positions, testing for the occurrence of nystagmus [154, 159]. Here, we quantify how smoothly the eyes were able to track the moving dot (velocity means) and the state of the eyes at the edges of the screen (dispersion at the hold positions).

A physician could interpret a low average velocity and a tight dispersion at the hold positions as normative responses. A high average velocity indicates that the eyes had to catch up or jump ahead of the target many times, suggesting a possible underlying motor problem that could be neurological in origin. An inability to consistently fixate at the hold positions will result in a large dispersion being calculated. This behaviour could be interpreted as the presence of nystagmus, which could be assessed in the same way as the Stare task.

4.5.3 Vestibulo-ocular reflex (VOR)

The VOR is a compensatory reflex that allows gaze fixation to remain stable on the retina during head movements. The vestibular system detects movements in the head, which are transmitted to the brain stem and eventually the extraocular muscles responsible for controlling the eyes [162]. Damage to any of these areas can produce deficits in the VOR response. These deficits can be detected as an inability to stabilize gaze during head rotations, meaning the pupil position will not move congruently with the head. We can assess this by calculating how well the pupil movements match the expected head movements. For example, while fixating on a stationary target, and nodding the head up and down, then left and right, the movement of the pupils should resemble a cross (see Fig 4.3 for an example). One challenge that arises is the location of the pupil camera; it will be on an angle relative to the eye. We can generate a model that attempts to fit a cross to the generated eye movements that corrects for the angle of the pupil eye tracker camera. This is done as follows:

1. Using the cleaned XY pupil position data (per eye), an orthogonal linear least square fit is performed using a custom MATLAB function[3]. The function fits a line to the data, while attempting to minimize the residual error and remaining

[3]This function was originally written by Per Sundqvist and is available on the MATLAB File Exchange

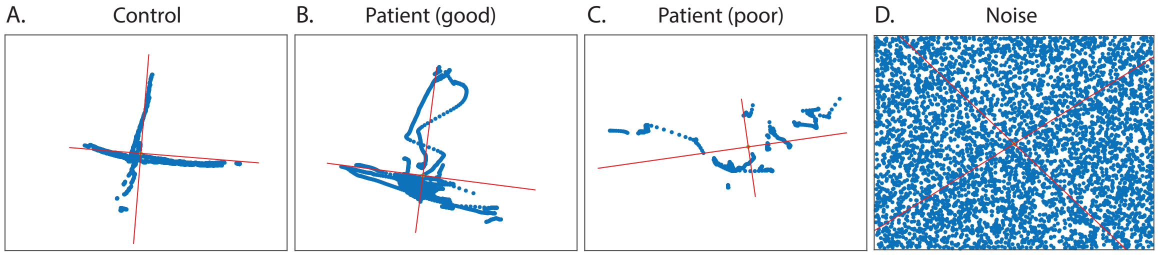


Figure 4.3: A visual example of data from a single eye collected during the VOR task, with the XY data points plotted in blue, the mean of the XY data plotted in as a green dot, and the fitted cross drawn in red. Three participants and an example of noise are shown. Cleaned data is shown for each participant. A) In the control data, the cross fits quite well, and the data are not noisy, resulting in a lower error. B) An example of good patient data, which are more noisy (even post-cleaning), which results in a cross that fits relatively well, but the data are much more spread out resulting in a higher error. C) The poor quality patient data are even worse. Further, it does not resemble a cross and cannot be easily interpreted. D) An example of uniformed distributed noise with a cross fit. This is one of the worst cases for fit, as data are randomly spread over the entire collection area.

orthogonal to the line.

2. An orthogonal line is generated using the ‘orthogonalLine‘ function [163], drawn across the mean of the XY data.

3. For each point, the minimum Euclidean distance to either of the two lines is calculated.

4. The root-mean-squared error is taken of the Euclidean distances.

A returned value of 0 would indicate that the data perfectly fit the cross, and a value of 1 would indicate the worst possible fit. To get a sense of what normative data looks like, we calculated the cross fit error and compared it to some worst case scenario data.

Clinical relevance

Deficits in the VOR can give hints for localizing neurological damage, such as to the cerebellum [164] or vestibular system [154, 165]. Typically, the VOR is assessed during the head impulse of the HINTS exam [154, 166], where saccades during the impulse are taken as a sign of vestibular imbalance. In our task, these saccades would show up during head rotations to the affected side. Additional saccades would result in the overall shape of the data being less 'cross-like', and result in a higher error. It is possible that the current analysis could benefit from saccade detection, but the data quality from patients was too low. A physician could interpret a low error (i.e. more cross-like) to be indicative of a normative VOR response, and a higher error as a signal that the patient should be investigated further for cerebellar or vestibular lesions.

4.5.4 Jump

Saccadic eye movements are controlled by many brain areas, including the frontal lobes, brain stem, and various oculomotor nuclei [154, 167]. Brain stem damage is associated with slowed and delayed saccades [154], with isolated slowed horizontal movements being diagnostically different from vertical [168]. Inaccurate saccades can be a sign of damage to the cerebellar vermis and fastigial nuclei [169]. The participant is shown a cross in the centre of the screen to fixate on and, after some amount of time, a target (a white dot) will show up on either the left or right side of the screen. It is important that the subject is not anticipating the stimulus, so the inter-trial interval is sampled from a pseudorandom uniform distribution.

For metrics, we calculate the reaction time as well as accuracy. To calculate the reaction time, pupil position data converted into velocities and then standardized (i.e Z-score). Peaks in the velocities are interpreted as saccades, and reaction time is calculated as when the velocity reaches at least 5% of the peak velocity (previously used in Stone *et al.* [124], see Figure 4.4). Accuracy is calculated by determining if the eyes went in the correct direction of the stimulus. These measures are in line with what physicians attempt to quantify manually [144, 154, 168], where high reaction times and inaccurate saccades are cause for concern.

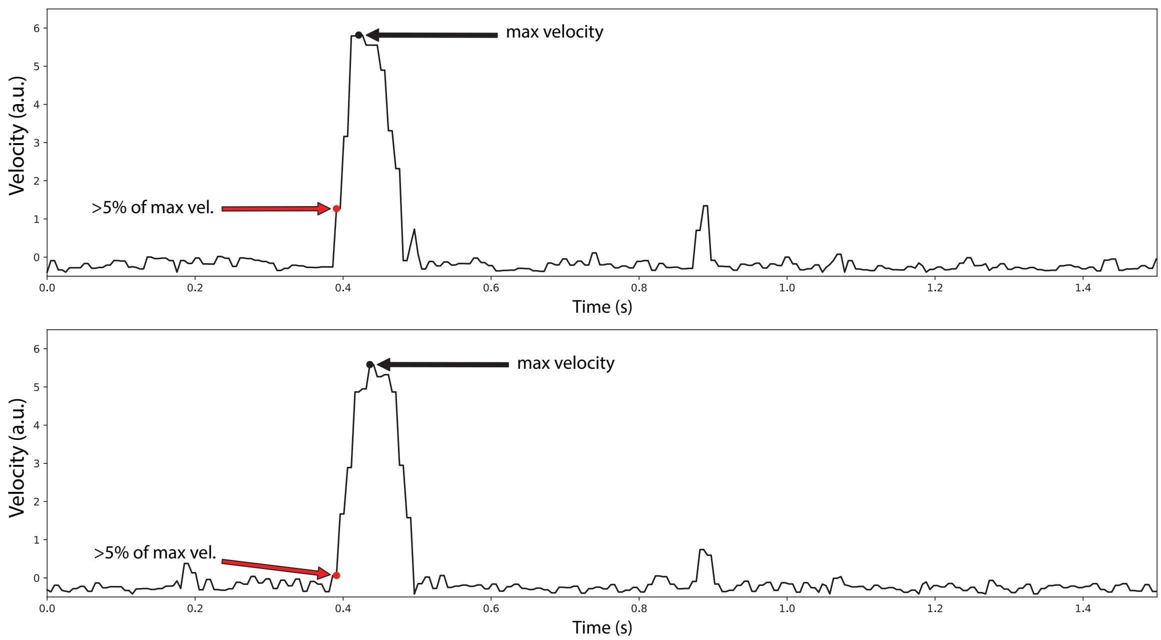


Figure 4.4: An example plot of the eye velocities during a jump in the Jump task. The x axis is time (seconds) and the y axis is velocity as a standardized unit (z-score). The top plot is the left eye data, and the bottom plot is right eye data. The maximum velocity is annotated with a black arrow, and the ¿5% threshold used to calculate reaction time is marked with a red annotated arrow. It is worth noting that this value is the first value beyond the threshold, so it may not be exactly 5% of the maximum velocity. Both eyes are shown to give a sense of the amount of variance that exists between the eyes, but also allows us to see if there are any non-consensual responses.

Clinical relevance

Deficits in the ability to make meaningful saccades give hints as to where damage may have occurred. As stated above, a physician may interpret a delayed saccade onset as a sign of brain stem damage. The current task only assesses horizontal saccadic jumps, but could easily be modified to include vertical jumps. A high average response time could be interpreted as a deficit in starting saccades, providing valuable diagnostic information. While accuracy was reduced to a simple binary response, the inability to make saccades in the correct direction (or multiple saccades being needed, interpreted as an incorrect saccade) can have diagnostic value. These metrics, when interpreted properly, can provide useful information to the attending physician.

4.5.5 Brightness

This task is split into two phases: a dark phase and a bright phase. The participant is told to fixate on a white centre cross on a black screen (i.e. identical to the stare task), and after around five seconds, the background turns white (see Fig 4.1E). This causes a pupillary constriction reflex as the eyes adjust to the light. This is meant to approximate shining a flashlight in the eyes.

The original intention of this task was to calculate parity of the pupillary constriction response between the eyes, but the data quality was too poor to conduct this analysis. Even in control participants, we did not see usable data that we speculate is because of the version of the Pupil Labs recording software used at the time[4].

Instead, we used the first five seconds of the data (i.e. the dark phase) to calculate a secondary stare dispersion measurement. For details of how this was calculated, see the Stare section.

[4] It does appear that more recent versions of the software have fixed a bug, so this may be feasible in a future iteration of this study.

Clinical relevance

It is well known that shining a light into the eyes elicits a pupillary constriction reflex. In general, normal humans will have a consensual (ipsilateral and contralateral) response regardless of which eye the light is shone in [170, 171], and non-responsiveness is indicative of neurological problems [170–173]. For example, it is possible to differentiate between an intracranial pathology (mass lesions and/or damage to the hypothalamus, midbrain, and pons) and circulatory arrest based on the pupillary response alone [171]. The ability to detect an abnormal response and report that information to a trained physician could speed up the diagnosis of a brain trauma patient.

4.6 Results

We used descriptive statistics to demonstrate the overall performance in the control and patient groups. Measures collected from the control group can be found in Table 4.2. Measures from the the patient group are shown in Table 4.3.

Following cleaning, 25.4 ± 27.2% of the data was lost in the controls. To better understand this number, below are the data loss following cleaning for each procedure. For Stare data, an average of 15.1 ± 22.9% the data was lost. For Pursuit, an average of 21.5 ± 25.4% of the data was lost. For Jump, an average of 25.6 ± 27.1% of the data was lost. For VOR, an average of 36.1 ± 20.7% of the data was lost. For Brightness, an average of 28.8 ± 36.2% of the data was lost.

The patient data had much more data loss following cleaning. Across all procedure types, an average of 73.9 ± 35.1% was lost. When split by the procedure, the data loss was relatively similar regardless of the procedure. For Stare data, an average of 74.1 ± 41.1% of the data was lost. For Pursuit, an average of 69.8 ± 38.3% of the data was lost. For Jump, an average of 68.4 ± 37.8% of the data was lost. For VOR, an average of 73.2 ± 33.1% of the data was lost. For Brightness, an average of 83.8

± 31.5.% of the data was lost. Even from just this data we can tell that the quality of patient eye data was very poor and will make subsequent analysis challenging.

4.6.1 Performance

The device was capable of collecting data in both the control and patient groups. In general, there were few issues with device dropout or technological issues, demonstrating that the device was capable of being deployed in an uncontrolled environment. None of the participants dropped out of the study, and the study was completed in typically less than 10 minutes. Data from all of the subtasks (with the exception of Brightness) are shown in Tables 4.2 and 4.3.

In the control data, it is worth noting one subject (C12) was not able to generate usable Stare dispersion data, due to a temporary dropout of pupil data.

In the patient data, the data appear to be much more variable, with much higher means and large standard deviations. Here, we believe that this data is poor quality due to the patient's general inability to complete the tasks, but not due to a failure of the device itself. Additional evidence of the poor quality data can be found in Figures 4.3 and 4.5.1C.

4.6.2 Control data

The control data, in general, were relatively homogenous. This provides a useful baseline from which subjects who deviate (i.e. have higher or lower values) could be used as a signal of disease. All referenced data can be viewed in Table 4.2.

Stare

Data was collected from each pupil separately to test the consistency of the dispersion algorithm. Testing each eye individually also gives the advantage of being able to detect non-consensual responses between the eyes. In the *Stare dispersion* column of Table 4.2, we can see that the width (x) and height (y) of the ellipses between the eyes are remarkably similar to one another, with the average width being 0.0090 $\pm$ 0.0047 and 0.0097 $\pm$ 0.0080 for P_0 and P_1. The average height was 0.0105 $\pm$ 0.0068 and 0.108 $\pm$ 0.0093 for P_0 and P_1. Given that the control subjects did not have any noted visuomotor disorders, we would expect the response from each eye to be similar to one another.

Pursuit

Pursuit data are split into three columns: *Pur. hold dispersion, Pur. hold vel., and Pur. move vel.*. *Pur. hold dispersion* describes the pupil dispersion over time while holding fixation on the extreme positions of the Pursuit task. Similarly, *Pur. hold vel.* describes the velocity of the pupils during the hold fixation portion. Finally, *Pur. move vel.* describes the average velocity of the eyes while the eyes are smoothly tracking the object.

The average hold dispersion width was 0.0075 $\pm$ 0.0049 and 0.0072 $\pm$ 0.0046 for P_0 and P_1. The average height was 0.0075 $\pm$ 0.0054 and 0.0056 $\pm$ 0.0046 P_0 and P_1.

The average hold velocity (i.e. fixation) was 0.0003 $\pm$ 0.0004 and 0.0001 $\pm$ 0.0005 for P_0 and P_1. The velocities only had a single value per eye. The average move velocity (i.e. smooth tracking) was found to be 0.0006 $\pm$ 0.0011 and 0.0003 $\pm$ 0.0001 for P_0 and P_1.

These measures are good descriptors for what a trained clinician would be looking for. A high degree of dispersion may suggest the presence of a neurological etiology. The velocity of the pupils while fixating on the stationary object can help discern if issues are arising during the smooth pursuit portion of the task or the fixation portion. Spikes in velocity during smooth movement would manifest as a higher average velocity, suggesting the eyes may be 'jumping ahead' or trying to catch up to the smoothly moving object. It is also interesting to note that the hold velocities are approximately twice as low in the *hold* state as compared to the *movement* state. This would make sense, as holding the eyes on a single target should result in lower velocities.

VOR

The VOR task tests how well the eyes are able to maintain a stable gaze while the head is moving. Each eye was assessed independently to see if the response was consensual. In our case, the movement of the eyes over time looks like a cross (see Figure 4.3 for an example), where the better the eyes are able to hold fixation on the central fixation point, the better fit the algorithm calculates (to get a better understanding of how the algorithm worked, read Section 4.5.3).

In general, we found that the control subjects were able to complete the task easily, as shown by their mean results (Table 4.2). The average VOR error was 0.0018 ± 0.009 and 0.016 ± 0.007 for P_0 and P_1. The average response from both eyes was roughly 0.017, whereas randomly sampled noise elicited a result of around 0.17—over ten times more error. Here, we think this result is quite useful for assessing someone's ability to use the VOR reflex, because deficits can be indicative of neurological damage. This measure appears to be relatively consistent between subjects, suggesting that it might be useful as a normative baseline measure.

Jump

The Jump task is designed to assess saccadic eye movements. Two measures are extracted from the data: accuracy and reaction time.

In general, most participants had a high accuracy value, with the average performance being close to 90%. Due to the way accuracy is calculated, it is possible that some participants with low accuracy values did not necessarily look in the opposite direction, but rather they either failed to respond to the stimulus, or they correctly looked at the stimulus and pre-emptively moved back towards the central fixation cross. In these cases, an incorrect response is calculated.

We calculated the eye movement reaction time for each stimulus. The average reaction time was around 250 ms, which is consistent with what we see in humans in similar tasks [174].

4.6.3 Patient data

Overall, the data quality from the patient group was poor. Average data loss following cleaning was 73.4%. In the following sections, results from each task will be discussed and compared to data generated from uniform random distributions. The data appear to be more similar to noise than to control data, suggesting that the recording quality was low. All referenced data can be viewed in Table 4.3.

Stare

Data was collected from each pupil separately to test the consistency of the dispersion algorithm. In the *Stare dispersion* column of Table 4.3, we can see that the width (x) and height (y) of the ellipses are much larger than that of the control data. Here, we have some evidence that the patient data is more similar to randomly sampled noise than it is to the control data (see Section 4.6.3A, note the logarithmic axis).

Here, the average width was 0.298 ± 0.253 and 0.164 ± 0.153 for P_0 and P_1. The average height was 0.367 ± 0.326 and 0.138 ± 0.144 for P_0 and P_1.

The patient data roughly falls into two categories: good and poor (for an example of each, see Section 4.5.1, *Patient (good) and Patient (poor)*). The good data, while much more disperse than the control data, at least appear to be roughly clustered. When using the control data as a scale, good patient data appears to have 7x more dispersion than the reference control. The poor patient data are much more diffuse—with the fitted ellipse being about 480x as large as the reference control. In this case, it is likely that the poor patient data is heavily influenced by noise.

Pursuit

Identically to the control data, Pursuit data are split into three columns: *Pur. hold dispersion, Pur. hold vel.*, and *Pur. move vel.*. For the *Pur. hold dispersion* task,

Subject	Stare dispersion				Pur. hold dispersion				Pur. hold vel.		Pur. move vel.		VOR		Jump	
ID	P_0x	P_0y	P_1x	P_1y	P_0x	P_0y	P_1x	P_1y	P_0	P_1	P_0	P_1	P_0	P_1	Acc. (%)	RT (ms)
C1	0.003	0.007	0.006	0.005	0.003	0.003	0.005	0.002	0.0002	0.0001	0.0002	0.0002	0.005	0.007	100.0	158.8
C2	0.006	0.005	0.034	0.006	0.004	0.002	0.002	0.003	0.0001	0.0001	0.0001	0.0002	0.013	0.012	100.0	311.4
C3	0.014	0.009	0.007	0.015	0.022	0.009	0.025	0.011	0.0002	0.0003	0.0002	0.0002	0.018	0.019	100.0	280.8
C4	0.004	0.014	0.005	0.013	0.009	0.005	0.002	0.019	0.0002	0.0001	0.0003	0.0002	0.036	0.020	100.0	187.7
C5	0.018	0.007	0.011	0.012	0.006	0.005	0.007	0.005	0.0001	0.0001	0.0001	0.0001	0.013	0.011	68.8	240.7
C6	0.004	0.007	0.009	0.004	0.003	0.004	0.004	0.003	0.0001	0.0001	0.0002	0.0002	0.016	0.023	93.8	206.1
C7	0.007	0.015	0.014	0.007	0.008	0.006	0.011	0.003	0.0001	0.0001	0.0002	0.0003	0.025	0.029	93.8	285.9
C8	0.008	0.004	0.005	0.005	0.008	0.007	0.007	0.005	0.0002	0.0002	0.0003	0.0002	0.020	0.016	100.0	230.0
C9	0.010	0.006	0.009	0.007	0.010	0.008	0.014	0.005	0.0001	0.0001	0.0002	0.0002	0.014	0.027	100.0	315.3
C10	0.007	0.030	0.007	0.016	0.006	0.007	0.007	0.003	0.0003	0.0002	0.0004	0.0004	0.010	0.014	100.0	248.5
C11	0.016	0.009	0.017	0.006	0.007	0.011	0.005	0.008	0.0005	0.0002	0.0006	0.0002	0.021	0.017	93.8	206.6
C12	-	-	0.002	0.012	0.005	0.005	0.004	0.003	0.0002	0.0001	0.0044	0.0003	0.036	0.009	56.3	349.2
C13	0.004	0.007	0.003	0.005	0.003	0.004	0.003	0.002	0.0002	0.0001	0.0003	0.0003	0.018	0.014	75.0	249.1
C14	0.012	0.008	0.004	0.041	0.007	0.021	0.005	0.004	0.0017	0.0001	0.0009	0.0004	0.022	0.012	37.5	160.5
C15	0.010	0.017	0.011	0.009	0.011	0.018	0.006	0.010	0.0004	0.0001	0.0005	0.0003	0.009	0.006	93.8	265.6
Mean	0.0090	0.0105	0.0097	0.0108	0.0075	0.0076	0.0072	0.0056	0.0003	0.0001	0.0006	0.0003	0.018	0.016	87.5	246.4
SD	0.0047	0.0068	0.0080	0.0093	0.0049	0.0054	0.0060	0.0046	0.0004	0.0005	0.0011	0.0001	0.009	0.007	19.3	56.5

Table 4.2: Table showing the means of all the measures collected for the control group, with between-subject means and standard deviations on the bottom two rows. Where applicable, data is show for each eye (P_0, P_1), and for dispersion measures the major (width) and minor (height) axes (x & y) are noted. All units, unless specified are in standardized (0–1) space. Stare dispersion shows the amount of dispersion calculated by the ellipse for each eye. The Pursuit dispersion at the hold locations is shown in the same form as the Stare data. The average Pursuit hold velocities (i.e. gaze should be stable) is shown. The average Pursuit movement velocity (i.e. when the eyes are smoothly tracking the object) are shown. The VOR cross-fitting root-mean-squared error is shown. The Jump accuracy (as a percentage) and average reaction times are shown.

again we see that the underlying data quality appears to be poor.

The average hold dispersion width was 0.150 ± 0.153 and 0.185 ± 0.178 for P_0 and P_1. The average height was 0.211 ± 0.282 and 0.184 ± 0.170 P_0 and P_1.

The average hold velocity (i.e. fixation) was 0.019 ± 0.022 and 0.017 ± 0.014 for P_0 and P_1. The velocities only had a single value per eye. The average move velocity (i.e. smooth tracking) was found to be 0.018 ± 0.017 and 0.017 ± 0.016 for P_0 and P_1.

When compared to control data, we can again see that the patient data are more similar to noise data than control data (see Section 4.6.3B), where control data had approximately 90 times less dispersion than the patient data. For the *Pur. move vel.* data, we can see that the patients have much higher average velocities, but it is not possible to compare this data to a noise baseline. The *Pur. hold vel.* data show a similar pattern to the *Pur. move vel*, yet it would be reasonable to expect the *hold* data to have lower overall velocities than the movement data.

VOR

As discussed in the control data, each eye was assessed independently to see if the response was consensual. The task, when done properly, will result in a 'cross-like' shape of the resulting pupil data (see Figure 4.3 Control data for an example). Using the two patient examples in Figure 4.3, we can see that at best, a patient data does not look exactly cross-like and at worst appears to be a smattering of dots somewhat randomly dispersed. It is worth noting that the quality of the fit is quite poor on most of the patients. The average VOR error is 0.063 ± 0.049 and 0.041 ± 0.018 for P_0 and P_1. For reference, the VOR cross-fitting yielded an average of 0.167 ± 0.019.[5] See Figure 4.3 for an example of noise.

On average, the cross fitting procedure produced 3 times more error when assessing patient data versus control data. In Section 4.6.3D we can see that the algorithm

[5]This was calculated from 1000 simulations of uniformly distributed noise.

assessed on uniform sampled noise results in a error much higher than the patient data, suggesting that there may be more signal remaining in the noisy data. This is evidenced in the patient data in Figure 4.3, where there are data points that do appear to fall along a cross, but much more of the data is corrupted by noise.

Jump

Identical to the control data, two measures were extracted from the data: accuracy and reaction time.

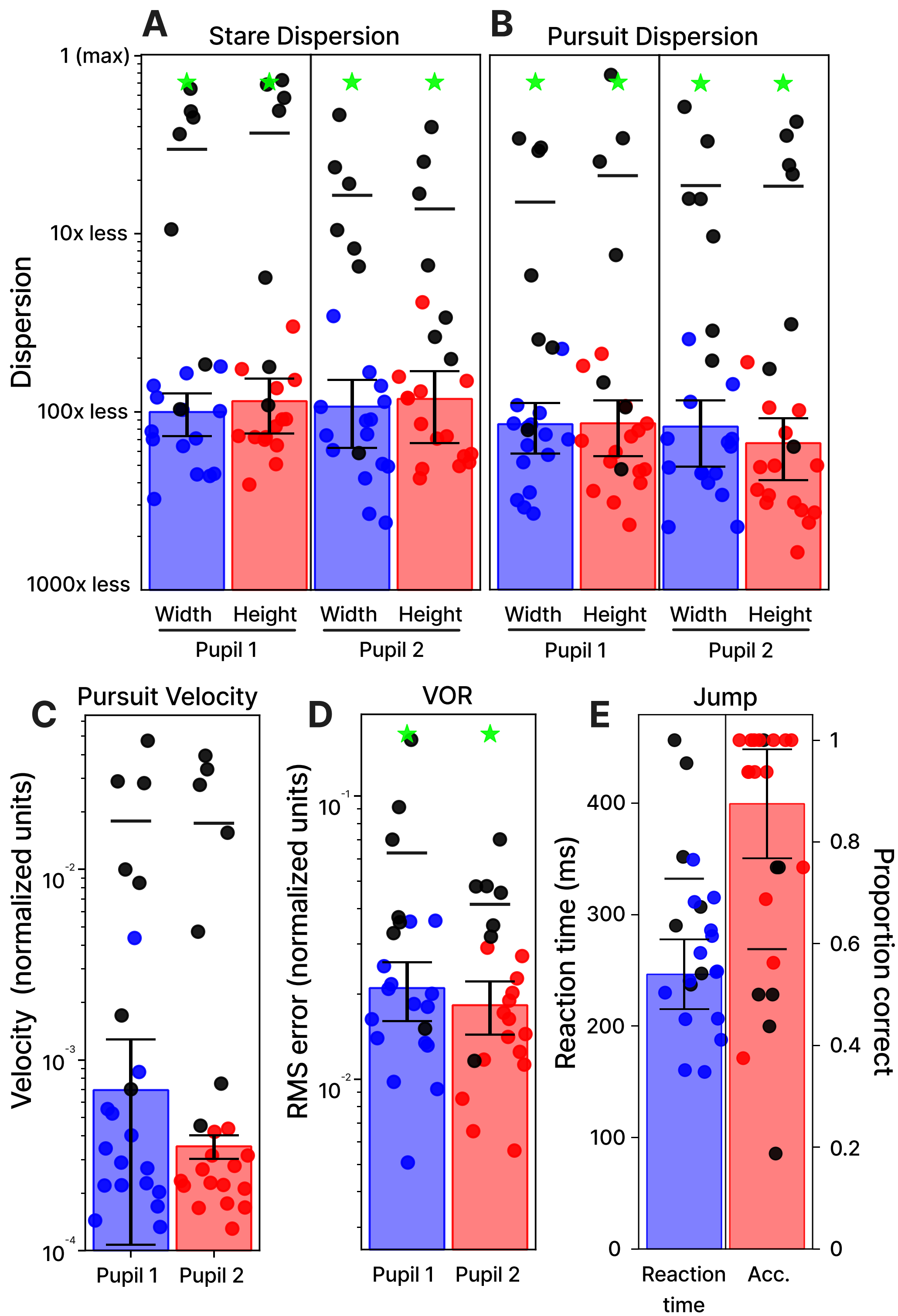
Here, the accuracy is relatively low, with an average accuracy of 58.9%—a value just above chance. It is not clear if the patients simply were not accurate in their responses or if the eye tracker was unable to reliably track the eyes during the task. When looking at the reaction time data, we can see that the mean reaction time between subjects is around 330ms. This is not unexpected, as it is slower than what we would expect for simple saccade reaction times. It is difficult to determine if these slower responses are because of the illness of the patient, or if it is due to the poor quality in the underlying data.

4.7 Discussion

We developed a test battery that approximates real-world assessments and is user-friendly for non-expert data collectors, which allowed us to evaluate the feasibility of collecting data in a clinical setting. There were three distinct goals we had when we conceived of this study. The first was using a non-expert to act as the primary data collector. This was very important because it is not reasonable to expect all end-users of a product to be experts to use it. Further, new technologies tend to be difficult to deploy in hospitals due to physician unfamiliarity [151, 175]. One way to alleviate this is by making the technology easy to use. We succeeded in this regard, as a non-expert was capable of easily collecting data with few issues. Our second goal was collecting metrics that are useful in a clinical context. Our metrics were

Subject	**Stare dispersion**				**Pur. hold dispersion**				**Pur. hold vel.**		**Pur. move vel.**		**VOR**		**Jump**	
ID	P_0x	P_0y	P_1x	P_1y	P_0x	P_0y	P_1x	P_1y	P_0	P_1	P_0	P_1	P_0	P_1	Acc. (%)	RT (ms)
P1	0.450	0.579	0.236	0.253	0.340	0.342	0.156	0.241	0.038	0.029	0.028	0.028	0.037	0.048	43.8	435.9
P2	0.010	0.018	0.065	0.020	0.008	0.005	0.019	0.006	0.000	0.001	0.010	0.000	0.070	0.035	50.0	237.1
P3	0.106	0.057	0.105	0.066	0.023	0.011	0.028	0.017	0.001	0.001	0.001	0.001	0.015	0.012	75.0	290.1
P4	0.362	0.688	0.191	0.397	0.302	0.774	0.156	0.353	0.059	0.031	0.047	0.039	0.158	0.070	18.8	306.9
P5	0.486	0.728	0.464	0.168	0.290	0.253	0.512	0.423	0.023	0.025	0.029	0.034	0.091	0.046	50.0	456.5
P6	0.652	0.490	0.006	0.026	0.058	0.075	0.096	0.031	0.011	0.008	0.008	0.005	0.036	0.032	75.0	247.1
P7	0.018	0.011	0.083	0.034	0.025	0.015	0.328	0.214	0.002	0.027	0.002	0.016	0.033	0.048	100.0	351.8
Mean	0.298	0.367	0.164	0.138	0.150	0.211	0.185	0.184	0.019	0.017	0.018	0.017	0.063	0.041	58.9	332.2
SD	0.253	0.326	0.153	0.144	0.153	0.282	0.178	0.170	0.022	0.014	0.017	0.016	0.049	0.018	26.5	86.9

Table 4.3: Table showing the means of all the measures collected for the patient group, with between-subject means and standard deviations on the bottom two rows. Where applicable, data is show for each eye (P_0, P_1), and for dispersion measures the major and minor axes (x & y) are noted. All units, unless specified are in standardized (0–1) space. Stare dispersion shows the amount of dispersion calculated by the ellipse for each eye. The Pursuit dispersion at the hold locations is shown in the same form as the Stare data. The average Pursuit hold velocities (i.e. gaze should be stable) is shown. The average Pursuit movement velocity (i.e. when the eyes are smoothly tracking the object) are shown. The VOR cross-fitting root-mean-squared error is shown. The Jump accuracy (as a percentage) and average reaction times are shown. It is worth noting that the overall quality of the data appears to be quite low.



)

Figure 4.5: (Caption is on the next page.)

Figure 4.5: General results from the control and patient populations. For all plots, he black dots are data from patients, with the horizontal black bar being the mean response. The other colored dots are data from control participants. The green star is a noise reference, where each task was assessed on uniformly distribution random values. The 95% confidence interval is shown for control data only. A) The stare dispersion values are shown for each pupil. The X axis denotes the width and height of the generated ellipse for each pupil. The Y axis is dispersion in normalized units on a logarithmic scale. The blue bar is the average width of the ellipse, with the individual participants plotted as blue dots. The red bar is the average height of the ellipse, with individual participants plotted as red dots. In general, the patients have a dispersion about 1.5 magnitudes larger than the controls, likely due to the poor quality data. B) The dispersion values for the hold periods of the pursuit task are shown for each pupil. The X axis denotes the width and height of the generated ellipse for each pupil. The Y axis is dispersion in normalized units on a logarithmic scale. The blue bar is the average width of the ellipse, with the individual participants plotted as blue dots. The red bar is the average height of the ellipse, with individual participants plotted as red dots. Here, again the average patient response appears to be about 1.5 magnitudes larger than control data, again likely due to poor quality data. C) The pupil velocity data during the pursuit task are shown for each pupil. The X axis is the pupil being assessed. The Y axis is the average velocity of the eyes while the eyes are moving during the pursuit task on a logarithmic scale. Similar to other data, the pupils of patients appeared to be moving around 1.5 magnitudes faster than controls. D) The root-mean-square error of the VOR cross fitting analysis technique is shown for each pupil. The X axis the pupil being assessed. The Y axis is the root-mean-square of the error from fitting the cross on a logarithmic scale. In general, we can see that the patient data did not fit the calculated cross as well (for an example of patient data, see Figure 4.3). E) Jump reaction time is shown. Here, the data in grey are from control participants. The Y axis is reaction time in milliseconds on a linear scale. In general, there is an average response time of around 250ms, with patients being about 330ms. F) The proportion of correct saccades is shown. Again, the data in grey are from control participants. In general, the control patients performed much better than the patients, who on average performed a correct saccade around 58% of the time.

derived from neurological assessments typically performed on patients experiencing vertigo symptoms [154]. Again, we were successful in generating quantifiable metrics that a health care professional could interpret and use to augment their diagnosis. Finally, for the device to truly be useful, we aimed to test it in the same uncontrolled environment that it would ostensibly be deployed in during actual use (for a discussion on this topic, see [176]). We used two uncontrolled environments to collect data, and found that the characteristic of being uncontrolled did not appear to negatively impact data collection. The data collector was able to successfully collect data from control participants, however, we found that the data quality was too low for proper analysis in the patient population. We found that the device performed well, but the state of the participant was much more influential on the quality of the data than we had previously expected.

Our test battery consisted of five components and was developed to approximate the real-world assessments that physicians use to evaluate the eye movements of patients with vertigo symptoms. The software was designed to be user-friendly and require minimal training to use. To test this, all data were collected by a research assistant with no prior eye tracking or research experience. This approach was chosen to evaluate the viability of collecting data in a clinical setting where an eye tracking expert will not be available. Initially, the device and software were tested on control subjects to gauge the feasibility of collecting non-patient data in an uncontrolled environment. We found that the system was effective and capable of generating measures that could potentially be useful in a clinical setting. We attempted to test this on a patient population that requires eye movement assessments for effective diagnosis. Patients with a primary complaint of dizziness were selected for the study because they often have associated eye movement disorders, which could prove to be a fruitful testing bed. Because the patients primarily complained of vertigo, we suspected it may be difficult to collect high-quality eye tracking data because many vertigo patients experience nausea when their eyes are open for extended periods of time [154, 159]. In general, we found this to be true.

Data were collected from control participants in a non-laboratory setting. The data collector was able to fit the eye tracker, and complete the stimulus presentation procedure with minimal issues. In general, the entire procedure took less than 10 minutes, including the time of fitting the eye tracker. We found that we were able to generate relatively consistent data across controls (see Section 4.6.3 and Table 4.2). The data quality, was first assessed visually to see if it matched the expected patterns. For example, the expected pattern of eye movements during the VOR task was expected to be in the shape of a cross (see Figure 4.3 for an example of control versus patient data). We found that this was true, and in general control participants tended to have more 'cross-like' data than the patients. Additionally, for the Stare task, we expected the data to be plotted closely together (during fixation) and again this was found to be true (see Section 4.5.1).

In general, the data quality from the patients was poor (see Section 4.6.3 for overall results). We speculate this is primarily because of the state of the patients when they were admitted. Because the patients were experiencing extreme vertigo, many found it too difficult to complete the tasks, yet none opted to drop out of the study. While initially, we were disappointed by this finding, it allowed us to assess one of our primary goals of the study: did the data collector influence the quality of the data? When compared to the control data, it is obvious that the data collector was not at fault for the data quality, as issues with data collection should have equally affected both controls and patients. In short, it appeared that the subjects were simply too ill to complete the task. This is useful for creating a future iteration of the device that could potentially eliminate some of the longer tasks and opt for a smaller number with a shorter duration.

Perhaps it is not surprising that the patients were unable to produce high-quality data. Despite their eagerness to participate, we were not fully cognizant of the limitations posed by vertigo, which impacted their ability to maintain focus even on a simple fixation cross. For this particular group of patients, collecting high-quality data without making substantial modifications to the software to accommodate for the difficulty in keeping their eyes open seemed unrealistic. To address this issue, it may be worth exploring tasks that are more accommodating for patients who struggle to keep their eyes open for extended periods. Additionally, there are two main strategies to enhance the data collection process and obtain high-quality data without sacrificing the user-friendliness of the software. The first would be to reduce the task space to one or two assessments such as Stare and Pursuit and reduce the amount of time needed to collect data. The second is to re-target to a patient population that would be more easily assessable such as those with smooth pursuit deficits due to strabismus, but critically without dizziness [177]. This falls beyond the scope of the present study, yet would make for a valuable area of future research.

The emergency department (ED) plays a vital role in the diagnostic process in hospitals. However, the demanding nature of EDs can pose challenges to maintaining a completely objective and systematic approach to diagnostic procedures. For example, when a patient presents with a suspected brainstem stroke, the pressure to make an accurate diagnosis can make it challenging to conduct exams such as the HINTS test. Eye trackers, by providing objective and quantifiable data, could help ED physicians make informed decisions for patients. This study represents a preliminary attempt at introducing advanced laboratory tools like eye trackers into the clinical setting and gauges the potential benefits and limitations of utilizing these tools. This of course necessitates real-world data collected in the same environment that diagnostic procedures are performed. Gathering real-world data is essential because when evaluating a system intended for clinical use, it should be tested under the same conditions and environment in which it will eventually be deployed. However, collecting data immediately after a patient's admission is challenging due to various factors. The logistics of data collection from patients can be difficult as it requires someone to be available at all times. Additionally, patients may not be in a state that allows for data collection. Furthermore, the range of symptoms among patients can be broad, making it difficult to compare data across patients. Understanding these limitations early on in the development cycle allows for us to iterate on the current design and improve it for future deployments.

This study aims to investigate the use of eye tracking technology in clinical settings. The initial objective was to collect and analyze data from patients with vertigo, in order to identify markers that suggest more severe symptoms, such as those associated with a stroke. Although this goal was not fully achieved, the data collected suggests 1) this system is technically feasible, and 2) minimal expertise is required for its deployment. The results of the present study suggest that, despite encountering logistical difficulties in data collection, it is feasible to imagine a clinical setting where the system is easily accessible and available to all staff. Future iterations of this system should be tested in more clinically useful areas, such as ambulatory or pre-hospital care. Because the system was developed with portability in mind, it can be easily deployed almost anywhere. The system, requiring only a laptop and eye tracker, is ideal for deployment in rural clinics with limited access to clinical technology [178, 179]. The laptop and eye tracker are stored in a backpack in between patients, meaning it is easy to grab to collect data. Another potential area of deployment is within lower stress environments that would benefit from quantified data, such as an optometrist's office. Optometrists regularly seek out and deploy new technology within their practises, and tend to have a favourable view on exploring newer technologies, especially if it increases the administrative workflow [180]. One study showed that minimal experience is needed to optimally detect small eye movements [181], suggesting that a well-tuned algorithm could potentially replace this component. An obvious use case for this device could be simply quantifying many of the eye movement components of a standard examination, such as smooth pursuit, fixation, and saccades. Eye trackers could be especially useful to increase the accuracy of diagnosis while simultaneously reducing the workload for the optometrist.

4.8 Conclusion

In summary, this study investigated deploying an eye tracker in an environment that is fundamentally difficult to control for the purpose of providing useful clinical information. A simple-to-deploy task was developed based on neurological examination principles. Critically, data collection was performed by physician with no prior eye tracking experience to show that data quality is not dependent on the operator's experience. Overall, the device was successful in generating useful metrics in an uncontrolled environment on control participants. However, some technological limitations prevented useful values from being generated in our preferred population: patients admitted with vertigo. Finally, we showed that it is technically possible to generate clinical metrics that an attending physician could use to augment their diagnosis of a patient.

Below are the key takeaways from this study:

1. It is important to target the end-user of the data being collected.

2. Collecting data in uncontrolled settings is possible and fruitful, depending on the context.

3. Eye tracking expertise is not necessarily needed to collect high quality eye tracking data.

4. Consider the context and environment critical when developing tools for real-world use, and ensure to validate them in similar environments to their intended use.

4.9 Acknowledgements

A huge thank you to Dr. Brenda Long for her help and enthusiasm throughout this project. Thank you to Dr. Glen Jickling and Dr. Rebecca Long for your thoughtful contributions and help. Thank you to the University Hospital Foundation for funding this work.

References

50. Kassner, M., Patera, W. & Bulling, A. *Pupil: an open source platform for pervasive eye tracking and mobile gaze-based interaction* in *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication* (Association for Computing Machinery, Seattle, Washington, Sept. 2014), 1151–1160.

111. D'Errico, J. inpaint_nans. *MATLAB Central File Exchange* (2012).

124. Stone, S. A. *et al.* Generating accurate 3D gaze vectors using synchronized eye tracking and motion capture. en. *Behav. Res. Methods.* ISSN: 1554-351X, 1554-3528. http://dx.doi.org/10.3758/s13428-022-01958-6 (Sept. 2022).

139. Baloh, R. W. Vertigo. en. *Lancet* **352,** 1841–1846. ISSN: 0140-6736, 1474-547X. http://www.thelancet.com/article/S0140673698054300/abstract (Dec. 1998).

140. Karatas, M. Central vertigo and dizziness: epidemiology, differential diagnosis, and common causes. en. *Neurologist* **14,** 355–364. ISSN: 1074-7931. http://dx.doi.org/10.1097/NRL.0b013e31817533a3 (Nov. 2008).

141. Saber Tehrani, A. S. *et al.* Rising annual costs of dizziness presentations to U.S. emergency departments. en. *Acad. Emerg. Med.* **20,** 689–696. ISSN: 1069-6563, 1553-2712. http://dx.doi.org/10.1111/acem.12168 (July 2013).

142. Newman-Toker, D. E. *et al.* Spectrum of dizziness visits to US emergency departments: cross-sectional analysis from a nationally representative sample. en. *Mayo Clin. Proc.* **83,** 765–775. ISSN: 0025-6196, 1942-5546. http://dx.doi.org/10.4065/83.7.765 (July 2008).

143. Cheung, C. S. K. *et al.* Predictors of important neurological causes of dizziness among patients presenting to the emergency department. en. *Emerg. Med. J.* **27,** 517–521. ISSN: 1472-0205, 1472-0213. http://dx.doi.org/10.1136/emj.2009.078014 (July 2010).

144. Edlow, J. A., Gurley, K. L. & Newman-Toker, D. E. A New Diagnostic Approach to the Adult Patient with Acute Dizziness. en. *J. Emerg. Med.* **54,** 469–483. ISSN: 0736-4679. http://dx.doi.org/10.1016/j.jemermed.2017.12.024 (Apr. 2018).

145. Kovacs, E., Wang, X. & Grill, E. Economic burden of vertigo: a systematic review. en. *Health Econ. Rev.* **9,** 37. ISSN: 2191-1991. http://dx.doi.org/10.1186/s13561-019-0258-2 (Dec. 2019).

146. Kattah, J. C., Talkad, A. V., Wang, D. Z., Hsieh, Y.-H. & Newman-Toker, D. E. HINTS to diagnose stroke in the acute vestibular syndrome: three-step bedside oculomotor examination more sensitive than early MRI diffusion-weighted imaging. en. *Stroke* **40,** 3504–3510. ISSN: 0039-2499, 1524-4628. http://dx.doi.org/10.1161/STROKEAHA.109.551234 (Nov. 2009).

147. Kung, N. H., Van Stavern, G. P. & Gold, D. R. HINTS in the Acute Vestibular Syndrome: Pearls and Pitfalls. en. *J. Neuroophthalmol.* **38,** 244–250. ISSN: 1070-8022, 1536-5166. http://dx.doi.org/10.1097/WNO.0000000000000608 (June 2018).

148. McDowell, T. & Moore, F. The Under-Utilization of the Head Impulse Test in the Emergency Department. en. *Can. J. Neurol. Sci.* **43,** 398–401. ISSN: 0317-1671. http://dx.doi.org/10.1017/cjn.2015.330 (May 2016).

149. Cohn, B. Can bedside oculomotor (HINTS) testing differentiate central from peripheral causes of vertigo? en. *Ann. Emerg. Med.* **64,** 265–268. ISSN: 0196-0644, 1097-6760. http://dx.doi.org/10.1016/j.annemergmed.2014.01.010 (Sept. 2014).

150. Kattah, J. C. Update on HINTS Plus, With Discussion of Pitfalls and Pearls. en. *J. Neurol. Phys. Ther.* **43 Suppl 2,** S42–S45. ISSN: 1557-0576, 1557-0584. http://dx.doi.org/10.1097/NPT.0000000000000274 (Apr. 2019).

151. Khullar, D. *et al.* Public vs physician views of liability for artificial intelligence in health care. en. *J. Am. Med. Inform. Assoc.* **28,** 1574–1577. ISSN: 1067-5027, 1527-974X. http://dx.doi.org/10.1093/jamia/ocab055 (July 2021).

152. Dai, T. & Tayur, S. R. *Designing AI-Augmented Healthcare Delivery Systems for Physician Buy-in and Patient Acceptance* http://dx.doi.org/10.2139/ssrn.4044140.

153. Shahbabu, B., Dasgupta, A., Sarkar, K. & Sahoo, S. K. Which is More Accurate in Measuring the Blood Pressure? A Digital or an Aneroid Sphygmomanometer. en. *J. Clin. Diagn. Res.* **10,** LC11–4. ISSN: 2249-782X, 0973-709X. http://dx.doi.org/10.7860/JCDR/2016/14351.7458 (Mar. 2016).

154. Goebel, J. A. The ten-minute examination of the dizzy patient. en. *Semin. Neurol.* **21,** 391–398. ISSN: 0271-8235. http://dx.doi.org/10.1055/s-2001-19410 (Dec. 2001).

155. Post, R. E. & Dickerson, L. M. Dizziness: a diagnostic approach. en. *Am. Fam. Physician* **82,** 361–8, 369. ISSN: 0002-838X, 1532-0650. https://www.ncbi.nlm.nih.gov/pubmed/20704166 (Aug. 2010).

156. Thier, P. & Ilg, U. J. The neural basis of smooth-pursuit eye movements. en. *Curr. Opin. Neurobiol.* **15,** 645–652. ISSN: 0959-4388. http://dx.doi.org/10.1016/j.conb.2005.10.013 (Dec. 2005).

157. Hessels, R. S., van Doorn, A. J., Benjamins, J. S., Holleman, G. A. & Hooge, I. T. C. Task-related gaze control in human crowd navigation. en. *Atten. Percept. Psychophys.* **82,** 2482–2501. ISSN: 1943-3921, 1943-393X. http://dx.doi.org/10.3758/s13414-019-01952-9 (July 2020).

158. Bianchi, L. J., Kingstone, A. & Risko, E. F. The role of cognitive load in modulating social looking: a mobile eye tracking study. en. *Cogn Res Princ Implic* **5,** 44. ISSN: 2365-7464. http://dx.doi.org/10.1186/s41235-020-00242-5 (Sept. 2020).

159. Killeen, D. E., Isaacson, B. & Walter Kutz, J. *History and Physical Examination of the Dizzy Patient* 2019. http://dx.doi.org/10.1007/978-3-319-97858-1_3.

160. Kim, J. S. *et al.* Upbeat nystagmus: clinicoanatomical correlations in 15 patients. en. *J. Clin. Neurol.* **2,** 58–65. ISSN: 1738-6586, 2005-5013. http://dx.doi.org/10.3988/jcn.2006.2.1.58 (Mar. 2006).

161. Abel, L. A. & Ziegler, A. S. Smooth pursuit eye movements in schizophrenics—What constitutes quantitative assessment? *Biol. Psychiatry* **24,** 747–761. ISSN: 0006-3223. https://www.sciencedirect.com/science/article/pii/0006322388902508 (Nov. 1988).

162. Cohen, B. & Raphan, T. in *The Vestibular System* (eds Highstein, S. M., Fay, R. R. & Popper, A. N.) 235–285 (Springer New York, New York, NY, 2004). ISBN: 9780387215679. https://doi.org/10.1007/0-387-21567-0_6.

163. Legland, D. *MatGeom: Matlab geometry toolbox for 2D/3D geometric computing* 2009.

164. Wolpert, D. M., Miall, R. C. & Kawato, M. Internal models in the cerebellum. en. *Trends Cogn. Sci.* **2,** 338–347. ISSN: 1364-6613. http://dx.doi.org/10.1016/s1364-6613(98)01221-2 (Sept. 1998).

165. Paige, G. D. Nonlinearity and asymmetry in the human vestibulo-ocular reflex. en. *Acta Otolaryngol.* **108,** 1–8. ISSN: 0001-6489. http://dx.doi.org/10.3109/00016488909107385 (1989).

166. Halmagyi, G. M. & Curthoys, I. S. A clinical sign of canal paresis. en. *Arch. Neurol.* **45,** 737–739. ISSN: 0003-9942. http://dx.doi.org/10.1001/archneur.1988.00520310043015 (July 1988).

167. Land, M. F. The coordination of rotations of the eyes, head and trunk in saccadic turns produced in natural situations. en. *Exp. Brain Res.* **159,** 151–160. ISSN: 0014-4819. http://dx.doi.org/10.1007/s00221-004-1951-9 (Nov. 2004).

168. Strupp, M., Dieterich, M. & Brandt, T. The treatment and natural course of peripheral and central vertigo. en. *Dtsch. Arztebl. Int.* **110,** 505–15, quiz 515–6. ISSN: 1866-0452. http://dx.doi.org/10.3238/arztebl.2013.0505 (July 2013).

169. Robinson, F. R., Fuchs, A. F. & Noto, C. T. Cerebellar influences on saccade plasticity. en. *Ann. N. Y. Acad. Sci.* **956,** 155–163. ISSN: 0077-8923. http://dx.doi.org/10.1111/j.1749-6632.2002.tb02816.x (Apr. 2002).

170. Choi, S. C., Narayan, R. K., Anderson, R. L. & Ward, J. D. Enhanced specificity of prognosis in severe head injury. en. *J. Neurosurg.* **69,** 381–385. ISSN: 0022-3085. http://dx.doi.org/10.3171/jns.1988.69.3.0381 (Sept. 1988).

171. Meyer, S., Gibb, T. & Jurkovich, G. J. Evaluation and significance of the pupillary light reflex in trauma patients. en. *Ann. Emerg. Med.* **22,** 1052–1057. ISSN: 0196-0644. http://dx.doi.org/10.1016/s0196-0644(05)82750-7 (June 1993).

172. Kim, T. J. *et al.* Neurological Pupil Index as an Indicator of Neurological Worsening in Large Hemispheric Strokes. en. *Neurocrit. Care* **33,** 575–581. ISSN: 1541-6933, 1556-0961. http://dx.doi.org/10.1007/s12028-020-00936-0 (Oct. 2020).

173. Hoffmann, M. *et al.* Pupil evaluation in addition to Glasgow Coma Scale components in prediction of traumatic brain injury and mortality. en. *Br. J. Surg.* **99 Suppl 1,** 122–130. ISSN: 0007-1323, 1365-2168. http://dx.doi.org/10.1002/bjs.7707 (Jan. 2012).

174. Hoffman, J. E. & Subramaniam, B. The role of visual attention in saccadic eye movements. en. *Percept. Psychophys.* **57,** 787–795. ISSN: 0031-5117. http://dx.doi.org/10.3758/bf03206794 (Aug. 1995).

175. Escarce, J. J. Externalities in hospitals and physician adoption of a new surgical technology: an exploratory analysis. en. *J. Health Econ.* **15,** 715–734. ISSN: 0167-6296. http://dx.doi.org/10.1016/s0167-6296(96)00501-2 (Dec. 1996).

176. Holleman, G. A., Hooge, I. T. C., Kemner, C. & Hessels, R. S. The 'Real-World Approach' and Its Problems: A Critique of the Term Ecological Validity. en. *Front. Psychol.* **11,** 721. ISSN: 1664-1078. http://dx.doi.org/10.3389/fpsyg.2020.00721 (Apr. 2020).

177. Lions, C., Bui-Quoc, E., Wiener-Vacher, S., Seassau, M. & Bucci, M. P. Smooth pursuit eye movements in children with strabismus and in children with vergence deficits. en. *PLoS One* **8,** e83972. ISSN: 1932-6203. http://dx.doi.org/10.1371/journal.pone.0083972 (Dec. 2013).

178. Ward, M. M., Jaana, M., Bahensky, J. A., Vartak, S. & Wakefield, D. S. Clinical information system availability and use in urban and rural hospitals. en. *J. Med. Syst.* **30,** 429–438. ISSN: 0148-5598. http://dx.doi.org/10.1007/s10916-006-9014-3 (Dec. 2006).

179. Singh, R., Lichter, M. I., Danzo, A., Taylor, J. & Rosenthal, T. The adoption and use of health information technology in rural areas: results of a national survey. en. *J. Rural Health* **28,** 16–27. ISSN: 0890-765X, 1748-0361. http://dx.doi.org/10.1111/j.1748-0361.2011.00370.x (Jan. 2012).

180. Dabasia, P. L., Edgar, D. F., Garway-Heath, D. F. & Lawrenson, J. G. A survey of current and anticipated use of standard and specialist equipment by UK optometrists. en. *Ophthalmic Physiol. Opt.* **34,** 592–613. ISSN: 0275-5408, 1475-1313. https://onlinelibrary.wiley.com/doi/10.1111/opo.12150 (Sept. 2014).

181. Fogt, N., Baughman, B. J. & Good, G. The effect of experience on the detection of small eye movements. en. *Optom. Vis. Sci.* **77,** 670–674. ISSN: 1040-5488. http://dx.doi.org/10.1097/00006324-200012000-00014 (Dec. 2000).

Chapter 5

General Discussion

Control plays a critical role in most scientific study designs. The purpose of this thesis was to investigate gaze and body movement data collection in environments that necessitate real-world behaviours at the cost of experimental control. As a consequence of removing experimental control, I hoped to demonstrate that the validity of the data would increase. As I said in the Introduction, I hope I've convinced you, the reader, that as control decreases (and noise increases), there are still useful metrics buried in the data. Of course, as we have come to find, the usefulness of a metric depends entirely on who is interpreting it. Each of the studies in this thesis reduced control in order to target an end-user in charge of data interpretation. The importance of this cannot be overstated. Investigating data collection in environments that require real-world behaviors at the cost of experimental control raises important questions about the usefulness of metrics and who is interpreting them. As control decreases and noise increases, it becomes even more crucial to identify the most relevant and informative metrics for the end-user responsible for data interpretation, whether it be psychologists studying attentional mechanisms or grocery store owners seeking to understand consumer behavior.

Recall the earlier example of the psychologist interested in understanding what cereal customers choose at a supermarket. Here, it is important to consider a) what kind of data is going to be collected, b) what kind of metrics are going to be produced from this data and, c) who is going to be interpreting the metrics. We can imagine two different groups of people who might want to interpret the resulting data metrics. The first are psychologists who are interested in studying human attention behaviour mechanisms. The second are grocery store owners who simply want to better understand what cereals people are buying. For these two groups, a different study altogether might be run. For the psychologist, you might focus on visual search mechanisms and make predictions based some attentional theory of mind, which may entail a study design that isolates the particular mechanism the theory implicates. The grocery store manager, on the other hand, might be more interested in practical recommendations for their store, such as which cereals to stock or how to better display them. Therefore, they might want a study design that provides more actionable insights, such as which cereals are the most popular or which ones are more likely to be purchased when placed at eye level on the shelf. Ultimately, the data collection, metrics produced, and interpretation of results should be tailored to the specific goals and needs of the intended audience. The critical difference between these two designs is the output, driven primarily by the ecological validity of the behaviours being performed.

The goal here is to comprehend human behavior, which necessitates understanding behavior within its contextual framework. The central theme of this thesis centers on the interplay between behavior and context, specifically, the extent to which behavior can be quantified within a given context, as well as how our attempts to measure behavior can influence it. We call this the tension between experimental control, or the ability to tightly manipulate variables in a laboratory setting, and ecological validity, or the extent to which findings can be generalized to real-world contexts. Measuring visuomotor behavioral outputs such as eye and body movements is one way to understand human behavior within its contextual framework. In this thesis, I explore the tension between experimental control and ecological validity, or the extent to which findings can be generalized to real-world contexts. By measuring visuomotor behavioral outputs, I aim to demonstrate how this tension plays out in the study of human decision-making processes.

5.1 What is the real world?

This discussion aims to situate my studies on the spectrum from high control, low validity to low control, high validity. In the following section, I will discuss the use of the term “ecological validity” and its place in modern psychology. I argue that the term is poorly defined and often misused. Below I will discuss a brief history of the term, and how it is used in contemporary psychology and neuroscience research.

5.1.1 Ecological validity

To many psychologists, studying real-world behaviours is synonymous with using an 'ecologically valid' study design. Ecological validity as a term was originally coined by Egon Brunswik, who described it as the *"correlation between a proximal sensory cue (e.g., retinal stimulation) and a distal object-variable (e.g., object in the environment)"* ([182]; paraphrased by [176]). In short, it is the relationship between how things are in the real world and how they are represented—in this case, how 'matched' the object is in the environment on the retina. This definition is very different from how we think of and use the term ecological validity today, which is generally comparing how closely laboratory studies resemble and generalize to the real world. Brunswik thought of psychology as a science of organism-environment relations, where it is important to understand the relationships between how things are perceived in a specific environment, or ecology (for a review of the history of the term, read Holleman *et al.* [176]). While this is a concrete definition, it fails to reflect how many psychologists use the term nowadays, which is not typically explicitly defined, as critiqued by Holleman *et al.* [176]. In general though, we should strive to create and conduct experiments in a manner that most closely resembles what the researchers think the natural environment would provide. My use of the terms ecological validity and real-world are meant to reflect this goal; collecting data in the contexts in which we think natural behaviours emerge is key to understanding how people actually behave.

So what does it mean to conduct research in the 'real-world'? Arguably, there is no such thing as the 'real-world', as it does not have a clear definition or any specific properties. Perhaps a good approach is to specify the behaviours and the context in which we are interested in studying them. For example, to best study where humans look when being asked to pass the salt at dinner time, it would make sense to track the eyes of a person who is actively eating dinner rather than trying to create a analogous laboratory-based task. This is peak ecological validity because we are capturing a behaviour in its natural form in the environment that it actually takes place in. Of course, this experimental design is more difficult to collect and analyze the underlying data than a laboratory-based task that may contain the same required movements. However, what is not clear is whether the tools we use in visuomotor behavioural neuroscience—eye-trackers, motion-trackers—are capable of collecting data in these naturalistic environments. The results of the three studies presented give credence to this idea, and further suggest that these tools are capable of being deployed outside the lab to generate actually useful metrics despite a clear increase in noise. Further, there is a clear trend in the technology powering these techniques becoming cheaper and more portable over time, suggesting that conducting such studies will become increasingly feasible in the near future.

5.1.2 Each study in context

Below, I will discuss each of the three studies in this thesis in the following framework. First, I will give a summary of the study and the main findings. Second, I will discuss where the study falls on the spectrum from control to validity. Third, I will discuss the limitations of the study, and how these were shaped by where it sat on the spectrum. Finally, I will discuss the utility of the study - who was the audience and what do the findings mean when factoring in the environment and behaviour being studied.

5.2 Study 1: eye and body tracking in the lab

5.2.1 Summary

In the first study, I used simultaneous eye and body tracking and had participants complete a simple everyday task: pick up and move an object around. This study was conducted in a controlled laboratory space, but critically the task was designed to be more naturalistic than typical laboratory-based tasks. One key part of the study that promoted the emergence of natural behaviours was the use of a head-mounted eye tracker as well as an open motion capture collection space. The participant was fitted with motion capture marker plates, but was otherwise free to move. As a consequence, allowing more free movement essentially permits more noise to be present in the resulting data. Arguably, this task has more real-world relevance because participants are free to choose how they move. When contrasted with many other studies investigating visuomotor reach-to-grasp movements, the participant tends to be restricted to a seat in front of a computer or tabletop [63, 69, 183, 184]. Clearly, this is not how humans naturally behave, so it unclear how the findings of many of these studies apply to the real world. Allowing free movement was crucial for this study to have any level of ecological validity. The purpose of this task was twofold: 1) participants could perform a complex behaviour with minimal restrictions and, 2) to generate a set of recommendations for scientists interested in collecting three dimensional gaze vectors from participants who were able to move freely.

For eye tracking, we used a Pupil Labs Core eye tracker, which had a low-profile form factor and granted free head movements. For motion capture, we used OptiTrack motion capture systems to track the body and environment. One of the technical hurdles overcome during this study was the simultaneous collection and synchronization of eye and body movements. To solve this, the lab streaming layer (LSL) library was used to synchronize the data from the eye tracker and motion capture systems. Ensuring the eye and motion capture data were using the same timestamps was necessary, as a key component of the calibration procedure was tracking the eyes while fixating on a moving wand. Four different calibration procedures were used, which involved the participants fixating on the tip of a wand that was either stationary or moved in a pattern. The resultant eye movements were recorded for the purpose of generating a model that was capable of predicting the 3D gaze location based on the location of the pupils in the eyes.

A simple quadratic regression model was used to predict 3D gaze vector direction based on 2D pupil position in each eye and either a Cartesian or Spherical coordinate system representing the location of the calibration wand in space. Each of the four calibration models were then assessed on a) calibration data, b) validation task data, or c) pasta box task data. The pasta box task was a real-world task in which participants picked up and moved a pasta box from point to point [14, 25, 84, 124, 185]. The validation task used the same positions the participant would move the pasta box, but critically only required fixation on these points instead of interaction. Put simply, the validation task allowed us to know exactly where a participant was looking. This framework gave us a compelling way to assess the resultant gaze vector models.

In general, we found that it was possible to generate highly accurate gaze vectors that corresponded to about 1-2cm of linear error, or around 1 visual degree (if in peripersonal space, around 60cm distance) even during the pasta box task trials. Additionally, our use of the spherical coordinate system (as opposed to the more typical Cartesian system) also reduced the amount of error in the gaze vectors. While the models were robust, head turns tended to introduce more error, likely because these kinds of eye movements were not represented in the model training input. We also found that binocular data, if given the option, should always be recorded but mostly for redundancy. There are points where a single camera may be collecting a higher quality signal than the other, and relying on both camera signals then will reduce the overall quality. We speculate on the possibility of a dynamic system that chooses which eye to collect data from using an algorithm that assesses quality over time. Finally, and perhaps the most important finding, was that the calibration routine used by the researchers should reflect the locations in space that the participant is actually going to interact with. If a researcher wishes to collect real-world data, natural behaviours should be used to inform this design.

5.2.2 Spectrum

It is worth noting that although the task completed during this study was inspired by real-world behaviours [14, 25, 84, 124, 185], it is still relatively highly-controlled compared to the other studies in this thesis. As a consequence, the real-world validity of this study is low. For this study, we took a single step away from the control paradigms typically deployed during visuomotor reach-to-grasp tasks by allowing free head and body movements. But, the study still took place in a laboratory setting, where the lighting, camera placement, eye tracker fitting, and intermediate tasks were highly controlled.

This study was an earnest attempt to try to predict gaze behaviour in a less controlled task representative of one that could take place in real life. One possible way to increase the validity of this study is to deploy the models generated in real situations, which would still require the researchers to carefully interpret the output. It is possible that such a system could exist in the future, but as of today, no system exists that can capture the combined gaze, movement and environment data necessary to generate 3D gaze vectors mapped to labelled real world objects from non-laboratory settings.

5.2.3 Limitations

The first study is arguably the least ecologically valid of the three. The task itself is understandable as a naturalistic task that almost everyone performs on a daily basis, but does not take place in the same context that we are truly interested in. Because it is simply not feasible (or possible) to deploy the eye and body tracking system used in the home of each individual participant, concessions had to be made. What this study shows is that it is feasible to collect and analyze this kind of data. Earlier studies investigating natural behaviours, such as the tea-making [67] and sandwich making [12] tasks (both are discussed together in [122]) showed that natural behaviours were vastly underrepresented in the literature. In the sandwich making task, participants were asked to make a sandwich and pour themselves a glass of cola. For this task, they were intentionally given no instructions to encourage natural behaviours to emerge. We aimed to mimic this paradigm, but the techniques and methodologies we used necessitated a laboratory setting. Additionally, although the technology was seen as quite portable and lightweight at the time, even more portable eye tracking (and motion capture) systems have become available since the time this study was conducted.

With the recent advent of markerless motion capture systems (e.g. [60, 61, 104]) and even more portable eye tracking systems, it could be possible to actually record people moving about and interacting with objects in their daily lives. This study was a first step towards this goal. The methods in this study could likely be applied to these technologies, providing a more intimate look at the natural behaviours humans use in their lives.

5.2.4 Utility

The output of this study was aimed at researchers who are interested in collecting real-world behaviours in the laboratory. To be even more specific, it is for researchers who want to conduct visuomotor behavioural studies interested in generating 3D gaze vectors. As such, the findings in this study are intended for a more technical audience who may have a more intimate understanding of the inherent limitations and capabilities of the hardware used. In the conclusion of the study, we gave six key rules of thumb to consider when collecting naturalistic visuomotor behaviour in the lab. Of these, I would argue the most important piece of advice is to make sure the task used to assess behaviour reflects the actual real-world behaviour you are interested in studying. If the task does not reflect real life, it will have considerably low utility.

5.3 Study 2: eye and body tracking in the wild

5.3.1 Summary

In the second study, I made my first attempt at moving eye and body tracking outside of the laboratory. This study was completed during an internship at BioWare in Edmonton, Alberta. Here, I explored a research field known as user experience research (UXR), which typically uses qualitative methodologies such as interview and subjective reports to study and understand user behaviours. One of the goals of this study was to explore bringing quantitative measures to UXR by exploring visuomotor behaviour. Clearly, a wealth of valuable information is embedded in the way our eyes and hands work together. The stimulus for this study was a simple menu navigation task. The menu was based on a popular video game menu system (Mass Effect 3; BioWare), where various options to modify the gameplay are embedded. For this study, I was interested in exploring the use of eye and hand movements to detect when users were not optimally navigating the menu system. This is a concept known as friction. It is worth noting that not all friction is necessarily bad friction; sometimes added friction can help ensure a user has a better understanding of the product through the facilitation of slow thinking [186, 187]. Bad friction is something that does not facilitate a successful interaction [186]. For example, imagine the last time you browsed a website with a poor layout.

I used eye and mouse tracking while users navigated the menu while given either a direct or indirect prompt. The direct prompt was meant to give step-by-step instructions (i.e. more explicit) while indirect prompts gave directions using more casual language (i.e. implicit). I predicted that the direct prompts would result in less friction overall, and the indirect prompts would more closely reflect the behaviours of a user possibly confused—i.e. more akin to a new user of the menu. As this study did not take place in a controlled laboratory setting, this was a good opportunity to assess the usability of eye and mouse tracking in the wild. Additionally, I was able to collect data from remote participants using online data collection services. This data was collected using webcam-based eye tracking algorithms [130]. The main findings of this study were 1) it is possible to collect and analyze eye and mouse tracking data in a relatively (compared to laboratory-based) uncontrolled setting, 2) the tasks can be split into natural phases to grant a deeper insight into *where* friction may be occurring, 3) metrics that calculate the dynamic relationship between the eyes and the hands (i.e. Tlead; [129]) can be co-opted to understand user intention and 4) webcam-based eye tracking is feasible, albeit with the penalty of having lower spatial and temporal resolution. This study gave credibility to the idea that not only can eye and mouse tracking data be collected outside of the lab, but it is capable of generating valuable information. These techniques, if deployed by an expert user experience researcher, can augment the UX design process to better suit user needs.

5.3.2 Spectrum

For this study, I collected data from two different sets of participants: a local and remote cohort. For the local cohort, I had a high level of control over the behaviours and contexts in which the participants acted. Here, the computer hardware, webcam, lighting, room temperature, and essentially every aspect of the environment was kept consistent between participants. For the remote cohort, I had very little control over any of these factors, providing more opportunity for noise to augment the data. While noise was certainly present in both cohorts, the variability in the hardware (webcam, computer, mouse) and environmental aspects altered the data. Still, I found that the data was of sufficient quality.

Overall, the level of control ranged from high (local cohort) to low (remote cohort). The resultant utility of the data was high. A key finding was the persistent friction point that occurred when users navigated to the Narrative sub-menu, regardless of the cohort. This suggested 1) the noise inherent in the remote cohort was not enough to overpower the signal and 2) quantitative user experience is a feasible real-world application of naturalistic data collection paradigms. In short, this study showed that natural behaviours can be quantified in non-laboratory settings with a direct application in UX research and design.

5.3.3 Limitations

In this study, I collected real-world data from participants using a simple user interface with the intent to uncover points of friction. For scientific findings to be applicable to real-life, the data should be derived from naturalistic behaviours. This study, while not a perfect representation of ecological validity, took some of the necessary steps to see if our scientific control assumptions hold up when being applied to real behaviours. One obvious finding was that as experimental control decreases, analysis difficulty increases. Because the participants were given considerable freedom to complete the task, it is harder to systematically analyze the resulting data. We partially solved this issue by dynamically splitting the data into phases based on what stage of completing the task the user was on (i.e. navigating to, on, and away from the goal frame). This approach could be refined to even smaller sub-phases, such as prior to clicking on a specific button on a goal frame. Significant value can be gained from analyzing these interactions as a considerable amount of effort is invested in refining frequently used functions on websites, for example [188].

The menu system here was designed to be similar to the menu system in Mass Effect 3, rather than simply using the video game itself. Feasibility testing was performed using the actual game as a collection platform, but was quickly abandoned as it was too difficult to modify the game's source code to output the markers and variables required to do the analysis ultimately performed. For example, it was not possible for the game engine to output the location of any particular button on the screen, as the user interfacing code was performed using an internal scripting language that the game engine could not interface with. Future games (and game engines) would heavily benefit from making this kind of data accessible for external analysis, as studies like this one could be performed on the native version of the game rather than recreations. Of course, the time (and therefore financial) investment needed to ensure access to this kind of data is high, and more work should be conducted to further demonstrate value in UX fields.

Additionally, the experimenter must place a lot of trust in the participants to complete the task as designed. In especially the remote cohort, it was impossible to know if the participant truly understood the study until at least the analysis is complete. One of the main limiting factors was pre-processing of the data; it is harder to use a 'one-size-fits-all' approach; data need to be carefully analyzed individually, at least initially. Regardless of the difficulty, the result was fruitful data that were able to inform us of two candidate friction points; the Accomplishments and Narrative menu options. Qualitatively, it was relatively easy to contextualize these findings into their validity as *actual friction* the user was experiencing, or simply a byproduct of the task itself. Because the study was intended to provide probative value of collecting quantitative measures for UX research designs, the goal was rather vague—find friction in the user interface. While the experience was worthwhile, future quantitative UX studies should probably be constrained to specific questions, such as predicted problem areas, to reduce the analysis burden on the researcher.

5.3.4 Utility

This study necessitated real-world behaviour because its purpose was to develop metrics that a UX designer could use to inform their design decisions. Most UX research relies heavily on qualitative methods [125], but this study demonstrates that quantitative methods are potentially just as fruitful and can be used in conjunction with more traditional methods. Many UX researchers will rely on user reports about a product's usability, but arguably many users may not even be aware they are experiencing unwanted friction. The data here can be interpreted by trained UX researchers to uncover friction points that qualitative methods may miss, which gives very specific information about when and where the user was experiencing friction. Of course, the context of the behaviour is still the most important aspect of determining intent, such as when users navigated on the Accomplishments goal frame as opposed to the Narrative goal frame.

5.4 Study 3: eye tracking in the clinic

5.4.1 Summary

The objective of my third study was to determine the requirements for developing a device that produces valuable diagnostic data for medical professionals examining patients who suffer from dizziness. Typically, a neurologist will administer a neurological exam to assess the basic cognition and movement of the eyes. This test usually involves fixating on a stable target, smoothly pursuing a moving target, saccades between two targets, assessing the vestibulo-ocular reflex[1], and assessing pupillary light reactivity.

For this study, I created a device that records the pupils during a digitized version of the basic neurological exam described above. The primary output of the device, following analysis, are quantified metrics that a trained physician can use to augment their diagnostic process. The device used a Pupil Labs Core eye tracker, which was attached to a portable laptop computer. Critically, the device was designed to be easy to use by novice eye tracking users, with all of the data collection being performed by a non-expert. We sought to see if a non-expert was a reasonable source of noise being introduced into the data, as the main users of such a device would be non-expert health care professionals. Data were collected from two populations: a control group and patients experiencing vertigo. One of the primary motivating factors behind this study was determining if it was possible to deploy an eye tracker in an uncontrolled environment that was still capable of generating clinically useful information. Clinics and emergency departments tend to be both visually and audi-

[1]This is not always assessed, as sometimes it is not feasible to move the head of the patient.

torily busy, with patients coming and going, medical devices beeping, and patients tending to be nervous to begin with. These factors contributed to the uncontrolled nature of the environment, making it an excellent testing ground for determining feasibility of such a device.

In general, it took less than ten minutes to collect data from each patient and the role of the data collector was minimal. Prior to the beginning of each sub-task, the data collector simply had to explain the task to the patient and hit a single key to begin. To determine the influence of the patient group, the control group was collected in a similarly uncontrolled environment. In general, the data quality in the control participants was much higher than that of the patients. This was not unexpected, but ultimately the data quality of the patients was too poor to properly analyze. This provided evidence that it was not the environment alone that caused the poor quality, but rather was likely due to the patients' inability to keep their eyes open for the duration of the task. The high data quality of the control group demonstrated that not only was the device capable of generating usable diagnostic metrics, the data collector was also not the likely cause of the poor data quality in the patient group, suggesting that an expert eye tracking researcher is not required to collect data in the wild. Clinically inspired metrics were generated, and their potential value to a healthcare profession is speculated upon for the potential to improve the efficacy of a standard neurological exam.

5.4.2 Spectrum

The metrics generated here fall squarely in the "low control, high validity" category. The setting precluded the possibility of controlling many aspects of the environment, such as the auditory noise level, or the lighting. In this way, the metrics have high validity because the data were recorded in the exact kind of environment that such a device would be deployed in. Although we found that the noise overpowered the signal in patient population, the controls provided evidence that it was not due to the implementation of the software or the hardware. The patients in this study were perhaps too sick to participate, and it simply may not be possible to collect eye position data.

It is also worth noting that this may have been a point of failure where the technology is not mature enough for deployment in such uncontrolled settings.

5.4.3 Limitations

Working in a clinical setting is extremely challenging for many reasons. Because the focus of this study was collecting data from patients experiencing symptoms of vertigo—where symptoms can range from benign to completely debilitating. Unfortunately, because vertigo diagnoses tend to overlap with stroke, many of the patients in the present study were too sick to reliably collect data from. However, this was one of the proposed purposes of this research: we did not know if collecting eye movement data was a reasonable request from these kinds of patients. Because the perceived payoff was great—augmented data to increase the likelihood of a correct diagnosis, preservation of hospital imaging resources—and the relative burden on the patient was minimal, it was preferable to collect data in the environment in which the patient is naturally displaying their symptoms.

The device in its current form may be better suited for medical procedures that have more predictable patients such as an eye test at an optometrist's office. Patients here are much less likely have a debilitating illness, and it is easy to imagine how the metrics provided by the device could be used by a trained optometrist. For example, saccade measurements could be quantified to assess an oculomotor disorder, instead of relying purely on qualitative assessments. Further, controlled settings can provide additional testing grounds to find technical failings of the hardware or software.

Additionally, the sample size for the study is relatively low for both the control and patient population. For the patient population, it is remarkably difficult to collect data because of the logistical requirements. The data collector had to be essentially on call because we were interested in collecting data from patients soon after being admitted through the emergency department. This ultimately made it more difficult because extensive cooperation between the neurology and emergency departments was beyond our control for the duration of the study. Collecting more control participants is possible, and future iterations of the device and software would benefit from a large sample size from which normative measures are derived.

5.4.4 Utility

This was the most challenging environment thus far, which was necessary because the purpose of the device was to quantify aspects of an actual neurological exam. As such, the metrics generated in this study are intended for a trained health care provider for the purpose of augmenting their diagnosis of a patient. Currently, physicians assess the eye movements of patients purely qualitatively, which makes specificity difficult. Pre-determining normative behaviour (from controls) as a reference point can make it easier for a physician to categorize a potential patient as needing treatment, similar to blood test thresholds.

5.5 Retrospection and future directions

Collecting data in the real world is hard. One of the difficulties comes from the increased noise that is inherent to every day life. I would argue that while this noise is inconvenient from an analysis standpoint, it is important to consider that many of the analogs of behaviours we study in the laboratory are not representative of their natural counterparts.

The purpose of this thesis is to encourage you to think outside the box when it comes to designing experiments and to take a more liberal approach to predicted data quality. Many of the assumptions we make when designing a study for data collection or analysis are founded in reality; if we have a specific question we want to answer, we should be particular in how we go about answering it. But what about when we step away from the lens of basic science and want to know how we actually apply what we have learned? A web designer may want to gain some insights into how their users *actually* use their website, and why could they not use some of the same scientific tools and methodologies that we use in the lab?

A recent review by Snow & Culham [189] challenged the cognitive neuroscience and psychology communities to think deeply about their use of experimental proxies in visual studies. In this context, a proxy is a stimulus or task that is assumed to accurately represent a counterpart in the real world. Snow & Culham argue that while images are easy to create and manipulate in studies, they fail to capture some of the most important aspects of objects to be classified as real world. Critically, 2D images presented on a computer screen lack actability: the ability for someone to perform a genuine action. Someone cannot reach for and open a picture of a box of cereal the same way you could a real physical box. Further, there is evidence that real-object advantages exist over image analogs. Actability is a critical component of how we perceive real objects, where real tools shown to participants elicited a stronger motor preparation neural correlate than images [190]. Clearly, conducting research on real-world behaviours seems to necessitate actual genuine contexts to truly study natural behaviour. The relationship between how real an object is and its ecological validity is shown in Figure 5.1.

I contend, and have hopefully demonstrated in this thesis, that the same argument made against visual proxies by Snow & Culham extends more generally to experimental proxies for understanding behaviours in real world contexts. Coming into a lab and using a keyboard to choose what cereal you like from a picture on the screen is not the same as going to the supermarket and choosing a box off the shelf. Moving a pasta box across repetitive trials in the laboratory is not the same as making Kraft Dinner at home, but shares some key features. Measuring how you look and move a mouse navigating a video game menu, while from the comfort of your own home and using your own device, is closer to playing a real video game, but maybe not quite. But collecting data from vertigo patients as they present in the Emergency Department is exactly where you want to be to collect vital data from a real world environment where this kind of data can literally have life or death significance. Maybe it is not surprising that this is ultimately where this work hit the current technological wall, but the success collecting this kind of data across the variety of contexts on display in this thesis - combined with the rapid evolution of gaze and movement tracking technologies - means we are likely on the cusp of achieving a categorically new kind of ecological validity.

Future work investigating real-world visuomotor behaviours will likely become more prevalent as the underlying technologies improve. Tools such as DeepLabCut [104] and MediaPipe [61] are freely available and allow for the recording of body movements using consumer hardware. Recall the earlier example of the person shopping for cereal in the supermarket. Many of the technologies required to do this kind of study are available, but the most challenging aspect of collecting real behaviour in the wild is the person being recorded should not be aware. Simply being aware you are being watched is enough to alter behaviour [1, 75, 86, 136, 191]. This obviously brings about a much needed debate on the ethics of recording others without their informed consent. While it is beyond the scope of this thesis, I will admit that I do not have a good answer to this problem.

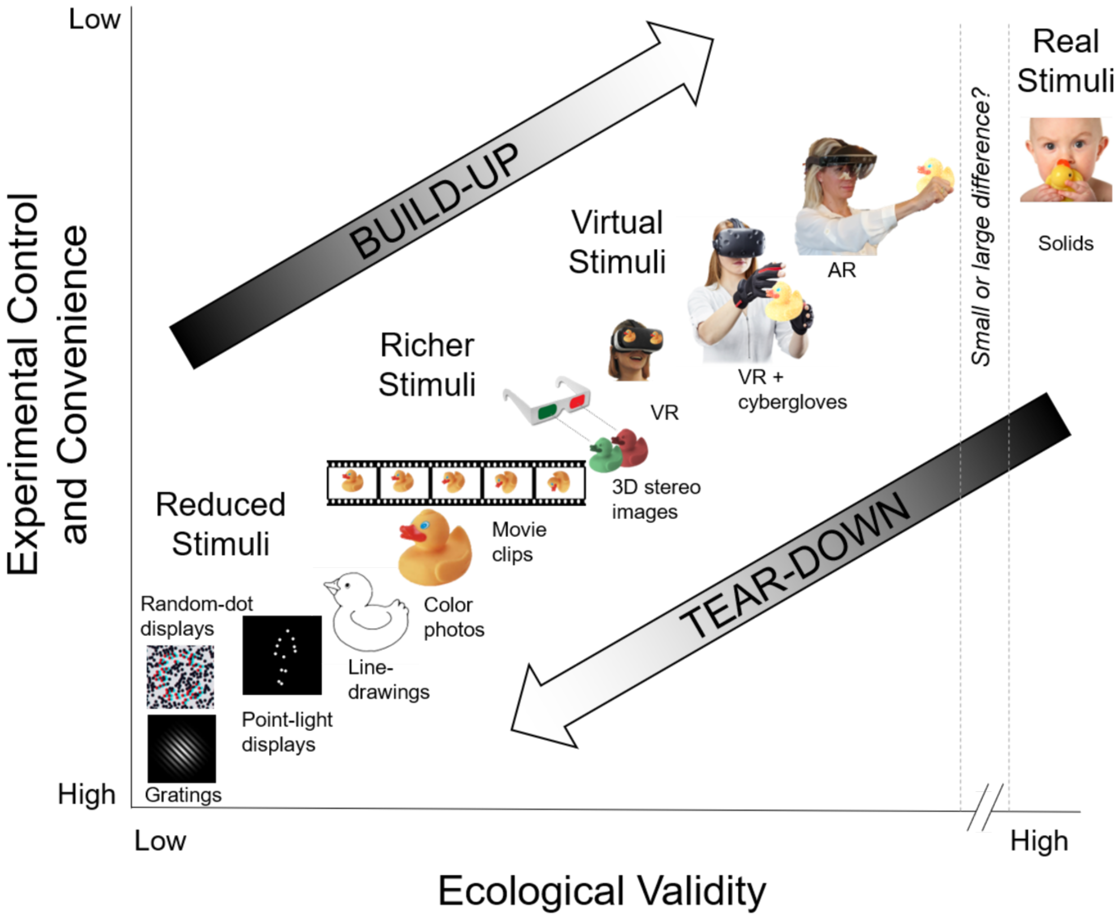


Figure 5.1: Objects can be conceptualized as falling along a continuum of realness, where artificial objects may lack ecological validity and fully real objects have high ecological validity. The actability of an object vastly affects our perception. Over time, our presentation of objects can be closer to fully real with the advent of technologies like virtual and augmented reality. However, the most ecologically valid type of stimulus is a real bona fide object. Figure taken from Snow & Culham [189].

References

1. Laidlaw, K. E. W., Foulsham, T., Kuhn, G. & Kingstone, A. Potential social interactions are important to social attention. en. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 5548–5553. ISSN: 0027-8424, 1091-6490. http://dx.doi.org/10.1073/pnas.1017022108 (Apr. 2011).

12. Hayhoe, M. Vision Using Routines: A Functional Account of Vision. *Vis. cogn.* **7,** 43–64 (Jan. 2000).

14. Lavoie, E. B. *et al.* Using synchronized eye and motion tracking to determine high-precision eye-movement patterns during object-interaction tasks. en. *J. Vis.* **18,** 18. ISSN: 1534-7362. http://dx.doi.org/10.1167/18.6.18 (June 2018).

25. Boser, Q. A. *et al.* Cluster-based upper body marker models for three-dimensional kinematic analysis: Comparison with an anatomical model and reliability analysis. en. *J. Biomech.* **72,** 228–234. ISSN: 0021-9290, 1873-2380. http://dx.doi.org/10.1016/j.jbiomech.2018.02.028 (Apr. 2018).

60. Mathis, A. *et al.* DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. en. *Nat. Neurosci.* **21,** 1281–1289. ISSN: 1097-6256, 1546-1726. http://dx.doi.org/10.1038/s41593-018-0209-y (Sept. 2018).

61. Lugaresi, C. *et al.* MediaPipe: A Framework for Building Perception Pipelines. arXiv: 1906.08172 `[cs.DC]`. http://arxiv.org/abs/1906.08172 (June 2019).

63. Voudouris, D., Smeets, J. B. J. & Brenner, E. Fixation Biases towards the Index Finger in Almost-Natural Grasping. en. *PLoS One* **11,** e0146864. ISSN: 1932-6203. http://dx.doi.org/10.1371/journal.pone.0146864 (Jan. 2016).

67. Land, M., Mennie, N. & Rusted, J. The roles of vision and eye movements in the control of activities of daily living. en. *Perception* **28,** 1311–1328. ISSN: 0301-0066. http://dx.doi.org/10.1068/p2935 (1999).

69. Johansson, R. S., Westling, G., Bäckström, A. & Flanagan, J. R. Eye-hand coordination in object manipulation. en. *J. Neurosci.* **21,** 6917–6932. ISSN: 0270-6474, 1529-2401. https://www.ncbi.nlm.nih.gov/pubmed/11517279 (Sept. 2001).

75. Freeth, M., Foulsham, T. & Kingstone, A. What affects social attention? Social presence, eye contact and autistic traits. en. *PLoS One* **8,** e53286. ISSN: 1932-6203. http://dx.doi.org/10.1371/journal.pone.0053286 (Jan. 2013).

84. Williams, H. E., Chapman, C. S., Pilarski, P. M., Vette, A. H. & Hebert, J. S. Gaze and Movement Assessment (GaMA): Inter-site validation of a visuomotor upper limb functional protocol. en. *PLoS One* **14,** e0219333. ISSN: 1932-6203. http://dx.doi.org/10.1371/journal.pone.0219333 (Dec. 2019).

86. Kingstone, A., Smilek, D. & Eastwood, J. D. Cognitive ethology: A new approach for studying human cognition. *British Journal of Psychology* **99,** 317–340 (2008).

104. Mathis, A. *et al.* DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. *Nature neuroscience* **21,** 1281–1289 (2018).

122. Land, M. F. & Hayhoe, M. In what ways do eye movements contribute to everyday activities? en. *Vision Res.* **41,** 3559–3565. ISSN: 0042-6989. http://dx.doi.org/10.1016/s0042-6989(01)00102-x (2001).

124. Stone, S. A. *et al.* Generating accurate 3D gaze vectors using synchronized eye tracking and motion capture. en. *Behav. Res. Methods.* ISSN: 1554-351X, 1554-3528. http://dx.doi.org/10.3758/s13428-022-01958-6 (Sept. 2022).

125. Nunnally, B. & Farkas, D. *UX Research: Practical Techniques for Designing Better Products* en. ISBN: 9781491951262. https://play.google.com/store/books/details?id=sfOeDQAAQBAJ ("O'Reilly Media, Inc.", Nov. 2016).

129. Deng, S., Chang, J., Kirkby, J. A. & Zhang, J. J. Gaze–mouse coordinated movements and dependency with coordination demands in tracing. *Behav. Inf. Technol.* **35,** 665–679. ISSN: 0144-929X. https://doi.org/10.1080/0144929X.2016.1181209 (Aug. 2016).

130. Finger, H., Goeke, C., Diekamp, D., Standvoß, K. & König, P. *LabVanced: a unified JavaScript framework for online studies* in *International Conference on Computational Social Science (Cologne)* (2017). https://www.researchgate.net/profile/Caspar-Goeke/publication/322273524_LabVanced_A_Unified_JavaScript_Framework_for_Online_Studies/links/5a4f7ac64585151ee284d8c2/LabVanced-A-Unified-JavaScript-Framework-for-Online-Studies.pdf.

136. Smilek, D., Birmingham, E., Cameron, D., Bischof, W. & Kingstone, A. Cognitive Ethology and exploring attention in real-world scenes. en. *Brain Res.* **1080,** 101–119. ISSN: 0006-8993. http://dx.doi.org/10.1016/j.brainres.2005.12.090 (Mar. 2006).

176. Holleman, G. A., Hooge, I. T. C., Kemner, C. & Hessels, R. S. The 'Real-World Approach' and Its Problems: A Critique of the Term Ecological Validity. en. *Front. Psychol.* **11,** 721. ISSN: 1664-1078. http://dx.doi.org/10.3389/fpsyg.2020.00721 (Apr. 2020).

182. Brunswik, E. *Representative design and probabilistic theory in a functional psychology* 1955. http://dx.doi.org/10.1037/h0047470.

183. Wispinski, N. J. *et al.* Reaching for known unknowns: Rapid reach decisions accurately reflect the future state of dynamic probabilistic information. en. *Cortex* **138,** 253–265. ISSN: 0010-9452, 1973-8102. http://dx.doi.org/10.1016/j.cortex.2021.02.010 (May 2021).

184. Diamond, J. S., Wolpert, D. M. & Flanagan, J. R. Rapid target foraging with reach or gaze: The hand looks further ahead than the eye. en. *PLoS Comput. Biol.* **13,** e1005504. ISSN: 1553-734X, 1553-7358. http://dx.doi.org/10.1371/journal.pcbi.1005504 (July 2017).

185. Stone, S. A. *et al.* in *ACM Symposium on Eye Tracking Research and Applications* 1–4 (Association for Computing Machinery, New York, NY, USA, May 2021). ISBN: 9781450383578. https://doi.org/10.1145/3450341.3458880.

186. Mejtoft, T., Hale, S. & Söderström, U. *Design Friction* in *Proceedings of the 31st European Conference on Cognitive Ergonomics* (Association for Computing Machinery, BELFAST, United Kingdom, Sept. 2019), 41–44. ISBN: 9781450371667. https://doi.org/10.1145/3335082.3335106.

187. Kahneman, D. *Thinking, Fast and Slow* en. ISBN: 9780385676526. https://play.google.com/store/books/details?id=TA7Q27RWlj0C (Doubleday Canada, Nov. 2011).

188. Craig Tomlin, W. *UX Optimization: Combining Behavioral UX and Usability Testing Data to Optimize Websites* en. ISBN: 9781484238677. https://play.google.com/store/books/details?id=7rJwDwAAQBAJ (Apress, Sept. 2018).

189. Snow, J. C. & Culham, J. C. The Treachery of Images: How Realism Influences Brain and Behavior. en. *Trends Cogn. Sci.* **25,** 506–519. ISSN: 1364-6613, 1879-307X. http://dx.doi.org/10.1016/j.tics.2021.02.008 (June 2021).

190. Marini, F., Breeding, K. A. & Snow, J. C. Distinct visuo-motor brain dynamics for real-world objects versus planar images. en. *Neuroimage* **195,** 232–242. ISSN: 1053-8119, 1095-9572. http://dx.doi.org/10.1016/j.neuroimage.2019.02.026 (July 2019).

191. Kingstone, A., Smilek, D., Ristic, J., Kelland Friesen, C. & Eastwood, J. D. Attention, Researchers! It Is Time to Take a Look at the Real World. *Curr. Dir. Psychol. Sci.* **12,** 176–180. ISSN: 0963-7214. https://doi.org/10.1111/1467-8721.01255 (Oct. 2003).

Bibliography

1. Laidlaw, K. E. W., Foulsham, T., Kuhn, G. & Kingstone, A. Potential social interactions are important to social attention. en. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 5548–5553. ISSN: 0027-8424, 1091-6490. http://dx.doi.org/10.1073/pnas.1017022108 (Apr. 2011).
2. Pickering, T. G. Clinical applications of ambulatory blood pressure monitoring: the white coat syndrome. en. *Clin. Invest. Med.* **14,** 212–217. ISSN: 0147-958X. https://www.ncbi.nlm.nih.gov/pubmed/1893653 (June 1991).
3. Pioli, M. R., Ritter, A. M., de Faria, A. P. & Modolo, R. White coat syndrome and its variations: differences and clinical impact. en. *Integr. Blood Press. Control* **11,** 73–79. ISSN: 1178-7104. http://dx.doi.org/10.2147/IBPC.S152761 (Nov. 2018).
4. Belletier, C. *et al.* Choking under monitoring pressure: being watched by the experimenter reduces executive attention. en. *Psychon. Bull. Rev.* **22,** 1410–1416. ISSN: 1069-9384, 1531-5320. http://dx.doi.org/10.3758/s13423-015-0804-9 (Oct. 2015).
5. Hayhoe, M. & Ballard, D. Eye movements in natural behavior. en. *Trends Cogn. Sci.* **9,** 188–194 (Apr. 2005).
6. Rayner, K. Eye movements and attention in reading, scene perception, and visual search. en. *Q. J. Exp. Psychol.* **62,** 1457–1506. ISSN: 1747-0218, 1747-0226. http://dx.doi.org/10.1080/17470210902816461 (Aug. 2009).
7. Eckstein, M. P. Visual search: a retrospective. en. *J. Vis.* **11.** ISSN: 1534-7362. http://dx.doi.org/10.1167/11.5.14 (Dec. 2011).
8. Hutton, S. B. Cognitive control of saccadic eye movements. en. *Brain Cogn.* **68,** 327–340. ISSN: 0278-2626, 1090-2147. http://dx.doi.org/10.1016/j.bandc.2008.08.021 (Dec. 2008).
9. Hayhoe, M. M. Advances in Relating Eye Movements and Cognition. en. *Infancy* **6,** 267–274. ISSN: 1525-0008, 1532-7078. http://dx.doi.org/10.1207/s15327078in0602_7 (Sept. 2004).
10. Hayhoe, M. M., Bensinger, D. G. & Ballard, D. H. Task constraints in visual working memory. en. *Vision Res.* **38,** 125–137. ISSN: 0042-6989. http://dx.doi.org/10.1016/s0042-6989(97)00116-8 (Jan. 1998).

11. Land, M. F. & Lee, D. N. Where we look when we steer. en. *Nature* **369,** 742–744. ISSN: 0028-0836. http://dx.doi.org/10.1038/369742a0 (June 1994).

12. Hayhoe, M. Vision Using Routines: A Functional Account of Vision. *Vis. cogn.* **7,** 43–64 (Jan. 2000).

13. Hayhoe, M. & Ballard, D. Eye movements in natural behavior. en. *Trends Cogn. Sci.* **9,** 188–194. ISSN: 1364-6613. http://dx.doi.org/10.1016/j.tics.2005.02.009 (Apr. 2005).

14. Lavoie, E. B. *et al.* Using synchronized eye and motion tracking to determine high-precision eye-movement patterns during object-interaction tasks. en. *J. Vis.* **18,** 18. ISSN: 1534-7362. http://dx.doi.org/10.1167/18.6.18 (June 2018).

15. Chapman, C. S. *et al.* Reaching for the unknown: multiple target encoding and real-time decision-making in a rapid reach task. en. *Cognition* **116,** 168–176. ISSN: 0010-0277, 1873-7838. http://dx.doi.org/10.1016/j.cognition.2010.04.008 (Aug. 2010).

16. Wispinski, N. J., Gallivan, J. P. & Chapman, C. S. Models, movements, and minds: bridging the gap between decision making and action. en. *Ann. N. Y. Acad. Sci.* ISSN: 0077-8923, 1749-6632. http://dx.doi.org/10.1111/nyas.13973 (Oct. 2018).

17. Wispinski, N. J. *et al.* Reaching for known unknowns: Rapid reach decisions accurately reflect the future state of dynamic probabilistic information. *Cortex* **138,** 253–265 (2021).

18. Gallivan, J. P., Chapman, C. S., Wolpert, D. M. & Flanagan, J. R. Decision-making in sensorimotor control. en. *Nat. Rev. Neurosci.* **19,** 519–534. ISSN: 1471-003X, 1471-0048. http://dx.doi.org/10.1038/s41583-018-0045-9 (Sept. 2018).

19. Gallivan, J. P. & Chapman, C. S. *Three-dimensional reach trajectories as a probe of real-time decision-making between multiple competing targets* 2014. http://dx.doi.org/10.3389/fnins.2014.00215.

20. Ouellette Zuk, A. A. *Going through the motions: Evaluating the impact of task, device and platform on mouse-tracking derived measures of decision difficulty* en. 2021. https://era.library.ualberta.ca/items/449f6cd6-ae12-4f9c-9e39-a0885f19b35b.

21. Freeman, J. B. & Ambady, N. MouseTracker: software for studying real-time mental processing using a computer mouse-tracking method. en. *Behav. Res. Methods* **42,** 226–241. ISSN: 1554-351X, 1554-3528. http://dx.doi.org/10.3758/BRM.42.1.226 (Feb. 2010).

22. Freeman, J. B., Dale, R. & Farmer, T. A. Hand in motion reveals mind in motion. en. *Front. Psychol.* **2,** 59. ISSN: 1664-1078. http://dx.doi.org/10.3389/fpsyg.2011.00059 (Apr. 2011).

23. Freeman, J. B. Doing psychological science by hand. en. *Curr. Dir. Psychol. Sci.* **27,** 315–323. ISSN: 0963-7214. http://dx.doi.org/10.1177/0963721417746793 (Oct. 2018).

24. Rheem, H., Verma, V. & Becker, D. V. Use of Mouse-tracking Method to Measure Cognitive Load. *Proc. Hum. Fact. Ergon. Soc. Annu. Meet.* **62,** 1982–1986. ISSN: 2169-5067. https://doi.org/10.1177/1541931218621449 (Sept. 2018).

25. Boser, Q. A. *et al.* Cluster-based upper body marker models for three-dimensional kinematic analysis: Comparison with an anatomical model and reliability analysis. en. *J. Biomech.* **72,** 228–234. ISSN: 0021-9290, 1873-2380. http://dx.doi.org/10.1016/j.jbiomech.2018.02.028 (Apr. 2018).

26. Valevicius, A. M. *et al.* Characterization of normative hand movements during two functional upper limb tasks. en. *PLoS One* **13,** e0199549 (June 2018).

27. Goodale, M. A. Transforming vision into action. en. *Vision Res.* **51,** 1567–1587. ISSN: 0042-6989, 1878-5646. http://dx.doi.org/10.1016/j.visres.2010.07.027 (July 2011).

28. Lashley, K. S. The mechanism of vision: XVIII. Effects of destroying the visual" associative areas" of the monkey. *Genet. Psychol. Monogr.* ISSN: 0016-6677. https://psycnet.apa.org/record/1949-01109-001 (1948).

29. Marr, D. & Hildreth, E. Theory of edge detection. en. *Proc. R. Soc. Lond. B Biol. Sci.* **207,** 187–217. ISSN: 0080-4649, 0950-1193. http://dx.doi.org/10.1098/rspb.1980.0020 (Feb. 1980).

30. Gregory, R. L. The intelligent eye. en. *Optom. Vis. Sci.* **48,** 851. ISSN: 1040-5488, 1538-9235. http://journals.lww.com/00006324-197110000-00016 (Oct. 1971).

31. Schneider, G. E. Two visual systems. en. *Science* **163,** 895–902. ISSN: 0036-8075. http://dx.doi.org/10.1126/science.163.3870.895 (Feb. 1969).

32. Ingle, D. Two visual systems in the frog. en. *Science* **181,** 1053–1055. ISSN: 0036-8075. http://dx.doi.org/10.1126/science.181.4104.1053 (Sept. 1973).

33. Mishkin, M. & Ungerleider, L. G. Contribution of striate inputs to the visuospatial functions of parieto-preoccipital cortex in monkeys. en. *Behav. Brain Res.* **6,** 57–77. ISSN: 0166-4328. http://dx.doi.org/10.1016/0166-4328(82)90081-x (Sept. 1982).

34. Mishkin, M., Lewis, M. E. & Ungerleider, L. G. Equivalence of parieto-preoccipital subareas for visuospatial ability in monkeys. en. *Behav. Brain Res.* **6,** 41–55. ISSN: 0166-4328. http://dx.doi.org/10.1016/0166-4328(82)90080-8 (Sept. 1982).

35. Goodale, M. A. & Milner, A. D. Separate visual pathways for perception and action. en. *Trends Neurosci.* **15,** 20–25. ISSN: 0166-2236. http://dx.doi.org/10.1016/0166-2236(92)90344-8 (Jan. 1992).

36. Milner, D. & Goodale, M. *The Visual Brain in Action* en. ISBN: 9780191545726. https://play.google.com/store/books/details?id=8CaQDwAAQBAJ (OUP Oxford, Oct. 2006).

37. Flanagan, J. R. & Johansson, R. S. Action plans used in action observation. en. *Nature* **424,** 769–771. ISSN: 0028-0836, 1476-4687. http://dx.doi.org/10.1038/nature01861 (Aug. 2003).

38. Zheng, B., Jiang, X., Bednarik, R. & Atkins, M. S. Action-related eye measures to assess surgical expertise. en. *BJS Open* **5.** ISSN: 2474-9842. http://dx.doi.org/10.1093/bjsopen/zrab068 (July 2021).

39. Gallivan, J. P. & Goodale, M. A. in *Handbook of Clinical Neurology* (eds Vallar, G. & Coslett, H. B.) 449–466 (Elsevier, Jan. 2018). https://www.sciencedirect.com/science/article/pii/B9780444636225000231.

40. Goodale, M. A., Króliczak, G. & Westwood, D. A. Dual routes to action: contributions of the dorsal and ventral streams to adaptive behavior. en. *Prog. Brain Res.* **149,** 269–283. ISSN: 0079-6123, 1875-7855. http://dx.doi.org/10.1016/S0079-6123(05)49019-6 (2005).

41. Chapman, C. S. & Goodale, M. A. Obstacle avoidance during online corrections. en. *J. Vis.* **10,** 17. ISSN: 1534-7362. http://dx.doi.org/10.1167/10.11.17 (Sept. 2010).

42. Dodge, R. & Cline, T. S. The angle velocity of eye movements. *Psychol. Rev.* **8,** 145–157. ISSN: 0033-295X, 1939-1471. https://psycnet.apa.org/fulltext/1926-02957-001.pdf (Mar. 1901).

43. Richardson, D. C. in *Encyclopedia of Biomaterials and Biomedical Engineering* 1060–1064 (CRC Press, May 2008). ISBN: 9780429154065. https://www.taylorfrancis.com/chapters/edit/10.1201/9780429154065-102/eye-tracking-characteristics-methods-daniel-richardson-michael-spivey.

44. Aserinsky, E. Effects of illumination and sleep upon amplitude of electro-oculogram. en. *AMA Arch. Ophthalmol.* **53,** 542–546. ISSN: 0096-6339. http://dx.doi.org/10.1001/archopht.1955.00930010550013 (Apr. 1954).

45. Jung, R. The registration of post-rotatory and optokinetic nystagmus and the optico-vestibular integration in humans. de. *Acta Otolaryngol.* **36,** 199–202. ISSN: 0001-6489. http://dx.doi.org/10.3109/00016484809123776 (May 1948).

46. Yarbus, A. L. Eye movements and vision Plenum. *New York.* ISSN: 0028-7369 (1967).

47. Tatler, B. W., Wade, N. J., Kwan, H., Findlay, J. M. & Velichkovsky, B. M. Yarbus, eye movements, and vision. en. *Iperception* **1,** 7–27. ISSN: 2041-6695. http://dx.doi.org/10.1068/i0382 (July 2010).

48. *Gazepoint* en. https://www.gazept.com. Accessed: 2022-7-27. Aug. 2013. https://www.gazept.com.

49. Wisiecka, K. *et al.* *Comparison of Webcam and Remote Eye Tracking* in *2022 Symposium on Eye Tracking Research and Applications* (Association for Computing Machinery, Seattle, WA, USA, June 2022), 1–7. ISBN: 9781450392525. https://doi.org/10.1145/3517031.3529615.

50. Kassner, M., Patera, W. & Bulling, A. *Pupil: an open source platform for pervasive eye tracking and mobile gaze-based interaction* in *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication* (Association for Computing Machinery, Seattle, Washington, Sept. 2014), 1151–1160.

51. Machıén, L. *et al.* The habitual nature of food purchases at the supermarket: Implications for policy making. en. *Appetite* **155,** 104844. ISSN: 0195-6663, 1095-8304. http://dx.doi.org/10.1016/j.appet.2020.104844 (Dec. 2020).

52. Muybridge, J. The Horse in Motion. en. *Nature* **25,** 605–605. ISSN: 0028-0836. https://www.nature.com/articles/025605b0 (Apr. 1882).

53. Ellard, C. G., Goodale, M. A. & Timney, B. Distance estimation in the Mongolian gerbil: the role of dynamic depth cues. en. *Behav. Brain Res.* **14,** 29–39. ISSN: 0166-4328. http://dx.doi.org/10.1016/0166-4328(84)90017-2 (Oct. 1984).

54. Smeets, J. B. & Brenner, E. A new view on grasping. en. *Motor Control* **3,** 237–271. ISSN: 1087-1640. http://dx.doi.org/10.1123/mcj.3.3.237 (July 1999).

55. Hommel, B. *et al.* No one knows what attention is. en. *Atten. Percept. Psychophys.* **81,** 2288–2303. ISSN: 1943-3921, 1943-393X. http://dx.doi.org/10.3758/s13414-019-01846-w (Oct. 2019).

56. Gallivan, J. P., Barton, K. S., Chapman, C. S., Wolpert, D. M. & Flanagan, J. R. Action plan co-optimization reveals the parallel encoding of competing reach movements. en. *Nat. Commun.* **6,** 7428. ISSN: 2041-1723. http://dx.doi.org/10.1038/ncomms8428 (July 2015).

57. Magnuson, J. S. Moving hand reveals dynamics of thought. en. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 9995–9996. ISSN: 0027-8424. http://dx.doi.org/10.1073/pnas.0504413102 (July 2005).

58. Goodale, M. A., Pelisson, D. & Prablanc, C. Large adjustments in visually guided reaching do not depend on vision of the hand or perception of target displacement. en. *Nature* **320,** 748–750. ISSN: 0028-0836. http://dx.doi.org/10.1038/320748a0 (1986).

59. Singhal, A., Monaco, S., Kaufman, L. D. & Culham, J. C. Human fMRI reveals that delayed action re-recruits visual perception. en. *PLoS One* **8,** e73629. ISSN: 1932-6203. http://dx.doi.org/10.1371/journal.pone.0073629 (Sept. 2013).

60. Mathis, A. *et al.* DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. en. *Nat. Neurosci.* **21,** 1281–1289. ISSN: 1097-6256, 1546-1726. http://dx.doi.org/10.1038/s41593-018-0209-y (Sept. 2018).

61. Lugaresi, C. *et al.* MediaPipe: A Framework for Building Perception Pipelines. arXiv: 1906.08172 [cs.DC]. http://arxiv.org/abs/1906.08172 (June 2019).

62. Song, J.-H. & Nakayama, K. Hidden cognitive states revealed in choice reaching tasks. en. *Trends Cogn. Sci.* **13,** 360–366. ISSN: 1364-6613. http://dx.doi.org/10.1016/j.tics.2009.04.009 (Aug. 2009).

63. Voudouris, D., Smeets, J. B. J. & Brenner, E. Fixation Biases towards the Index Finger in Almost-Natural Grasping. en. *PLoS One* **11,** e0146864. ISSN: 1932-6203. http://dx.doi.org/10.1371/journal.pone.0146864 (Jan. 2016).

64. Voudouris, D., Smeets, J. B. J., Fiehler, K. & Brenner, E. Gaze when reaching to grasp a glass. en. *J. Vis.* **18,** 16. ISSN: 1534-7362. http://dx.doi.org/10.1167/18.8.16 (Aug. 2018).

65. Williams, H. E., Chapman, C. S., Pilarski, P. M., Vette, A. H. & Hebert, J. S. Gaze and Movement Assessment (GaMA): Inter-site validation of a visuomotor upper limb functional protocol. en. *PLoS One* **14,** e0219333 (Dec. 2019).

66. SCCN. *Lab Streaming Layer* Mar. 2021.

67. Land, M., Mennie, N. & Rusted, J. The roles of vision and eye movements in the control of activities of daily living. en. *Perception* **28,** 1311–1328. ISSN: 0301-0066. http://dx.doi.org/10.1068/p2935 (1999).

68. Prime, S. L. & Marotta, J. J. Gaze strategies during visually-guided versus memory-guided grasping. en. *Exp. Brain Res.* **225,** 291–305. ISSN: 0014-4819, 1432-1106. http://dx.doi.org/10.1007/s00221-012-3358-3 (Mar. 2013).

69. Johansson, R. S., Westling, G., Bäckström, A. & Flanagan, J. R. Eye-hand coordination in object manipulation. en. *J. Neurosci.* **21,** 6917–6932. ISSN: 0270-6474, 1529-2401. https://www.ncbi.nlm.nih.gov/pubmed/11517279 (Sept. 2001).

70. Foulsham, T., Walker, E. & Kingstone, A. The where, what and when of gaze allocation in the lab and the natural environment. en. *Vision Res.* **51,** 1920–1931. ISSN: 0042-6989, 1878-5646. https://www.sciencedirect.com/science/article/pii/S0042698911002392 (Sept. 2011).

71. Goffman, E. *Behavior in Public Places* en. ISBN: 9781439108697. https://play.google.com/store/books/details?id=EM1NNzcR-V0C (Simon and Schuster, June 2008).

72. Kingstone, A., Smilek, D., Birmingham, E., *et al.* Cognitive ethology: Giving real life to attention research. *Measuring the mind,* 341–357. https://books.google.com/books?hl=en&lr=&id=qTwVDAAAQBAJ&oi=fnd&pg=PA341&dq=Cognitive+ethology+Giving+real+life+attention+research+Kingstone+Smilek&ots=PfAQeFSBX5&sig=jz8-sBMnA6KfFHKbkNnbE5TMx7Q (2005).

73. Kingstone, A., Smilek, D. & Eastwood, J. D. Cognitive Ethology: a new approach for studying human cognition. en. *Br. J. Psychol.* **99,** 317–340. ISSN: 0007-1269. http://dx.doi.org/10.1348/000712607X251243 (Aug. 2008).

74. Matthis, J. S., Yates, J. L. & Hayhoe, M. M. Gaze and the Control of Foot Placement When Walking in Natural Terrain. en. *Curr. Biol.* **28,** 1224–1233.e5. ISSN: 0960-9822, 1879-0445. http://dx.doi.org/10.1016/j.cub.2018.03.008 (Apr. 2018).

75. Freeth, M., Foulsham, T. & Kingstone, A. What affects social attention? Social presence, eye contact and autistic traits. en. *PLoS One* **8,** e53286. ISSN: 1932-6203. http://dx.doi.org/10.1371/journal.pone.0053286 (Jan. 2013).

76. Luck, S. J. *An Introduction to the Event-Related Potential Technique, second edition* en. ISBN: 9780262324069. https://play.google.com/store/books/details?id=y4-uAwAAQBAJ (MIT Press, June 2014).

77. Bigdely-Shamlo, N. *et al.* Hierarchical Event Descriptors (HED): Semi-Structured Tagging for Real-World Events in Large-Scale EEG. en. *Front. Neuroinform.* **10,** 42. ISSN: 1662-5196. http://dx.doi.org/10.3389/fninf.2016.00042 (Oct. 2016).

78. Gramann, K., Ferris, D. P., Gwin, J. & Makeig, S. Imaging natural cognition in action. en. *Int. J. Psychophysiol.* **91,** 22–29. ISSN: 0167-8760, 1872-7697. http://dx.doi.org/10.1016/j.ijpsycho.2013.09.003 (Jan. 2014).

79. Gramann, K., Jung, T.-P., Ferris, D. P., Lin, C.-T. & Makeig, S. Toward a new cognitive neuroscience: modeling natural brain dynamics. en. *Front. Hum. Neurosci.* **8,** 444. ISSN: 1662-5161. http://dx.doi.org/10.3389/fnhum.2014.00444 (June 2014).

80. Scanlon, J. E. M., Redman, E. X., Kuziek, J. W. P. & Mathewson, K. E. A ride in the park: Cycling in different outdoor environments modulates the auditory evoked potentials. en. *Int. J. Psychophysiol.* **151,** 59–69. ISSN: 0167-8760, 1872-7697. http://dx.doi.org/10.1016/j.ijpsycho.2020.02.016 (May 2020).

81. Robles, D. *et al. Surrounding Traffic Matters: Increases in Traffic Volume Are Related to Changes in EEG Rhythms in Urban Cyclists* en. May 2022. https://www.biorxiv.org/content/10.1101/2022.05.27.493782.

82. Jungnickel, E., Gehrke, L., Klug, M. & Gramann, K. in *Neuroergonomics* (eds Ayaz, H. & Dehais, F.) 59–63 (Academic Press, Jan. 2019). ISBN: 9780128119266. https://www.sciencedirect.com/science/article/pii/B9780128119266000105.

83. Scanlon, J. E. M., Jacobsen, N. S. J., Maack, M. C. & Debener, S. Does the electrode amplification style matter? A comparison of active and passive EEG system configurations during standing and walking. en. *Eur. J. Neurosci.* **54,** 8381–8395. ISSN: 0953-816X, 1460-9568. http://dx.doi.org/10.1111/ejn.15037 (Dec. 2021).

84. Williams, H. E., Chapman, C. S., Pilarski, P. M., Vette, A. H. & Hebert, J. S. Gaze and Movement Assessment (GaMA): Inter-site validation of a visuomotor upper limb functional protocol. en. *PLoS One* **14,** e0219333. ISSN: 1932-6203. http://dx.doi.org/10.1371/journal.pone.0219333 (Dec. 2019).

85. BioWare. *Mass Effect 3* Mar. 2012. https://www.bioware.com/games/#mass-effect-3.

86. Kingstone, A., Smilek, D. & Eastwood, J. D. Cognitive ethology: A new approach for studying human cognition. *British Journal of Psychology* **99,** 317–340 (2008).

87. Land, M. F. The coordination of rotations of the eyes, head and trunk in saccadic turns produced in natural situations. *Experimental brain research* **159,** 151–160 (2004).

88. Tomasi, M., Pundlik, S., Bowers, A. R., Peli, E. & Luo, G. Mobile gaze tracking system for outdoor walking behavioral studies. *Journal of vision* **16,** 27–27 (2016).

89. Lappi, O. Eye movements in the wild: Oculomotor control, gaze behavior & frames of reference. *Neuroscience & Biobehavioral Reviews* **69,** 49–68 (2016).

90. Sperandio, I., Kaderali, S., Chouinard, P. A., Frey, J. & Goodale, M. A. Perceived size change induced by nonvisual signals in darkness: the relative contribution of vergence and proprioception. en. *J. Neurosci.* **33,** 16915–16923 (Oct. 2013).

91. Neggers, S. F. & Bekkering, H. Ocular gaze is anchored to the target of an ongoing pointing movement. en. *J. Neurophysiol.* **83,** 639–651 (Feb. 2000).

92. Land, M. F. & Hayhoe, M. In what ways do eye movements contribute to everyday activities? en. *Vision Res.* **41,** 3559–3565 (2001).

93. Land, M., Mennie, N. & Rusted, J. The roles of vision and eye movements in the control of activities of daily living. en. *Perception* **28,** 1311–1328 (1999).

94. Lavoie, E. B. *et al.* Using synchronized eye and motion tracking to determine high-precision eye-movement patterns during object-interaction tasks. en. *J. Vis.* **18,** 18 (June 2018).

95. Parr, J., Vine, S. J., Harrison, N. & Wood, G. Examining the spatiotemporal disruption to gaze when using a myoelectric prosthetic hand. *Journal of motor behavior* **50,** 416–425 (2018).

96. Wilson, M. R., Vine, S. J. & Wood, G. The influence of anxiety on visual attentional control in basketball free throw shooting. *Journal of Sport and Exercise Psychology* **31,** 152–168 (2009).

97. Vickers, J. N. & Williams, A. M. Performing under pressure: the effects of physiological arousal, cognitive anxiety, and gaze control in biathlon. en. *J. Mot. Behav.* **39,** 381–394. ISSN: 0022-2895. http://dx.doi.org/10.3200/JMBR.39.5.381-394 (Sept. 2007).

98. Parr, J. V. V., Vine, S. J., Harrison, N. R. & Wood, G. Examining the Spatiotemporal Disruption to Gaze When Using a Myoelectric Prosthetic Hand. en. *J. Mot. Behav.* **50,** 416–425. ISSN: 0022-2895, 1940-1027. http://dx.doi.org/10.1080/00222895.2017.1363703 (July 2018).

99. Van der Kruk, E. & Reijne, M. M. Accuracy of human motion capture systems for sport applications; state-of-the-art review. en. *EJSS* **18,** 806–819 (July 2018).

100. Chapman, H. *et al.* Posterior parietal cortex control of reach-to-grasp movements in humans. *European Journal of Neuroscience* **15,** 2037–2042 (2002).

101. Rossit, S. *et al.* No neglect-specific deficits in reaching tasks. *Cerebral Cortex* **19,** 2616–2624 (2009).

102. Rossit, S. *et al.* No Neglect-Specific Deficits in Reaching Tasks. *Cerebral Cortex* **19,** 2616–2624. ISSN: 1047-3211. eprint: https://academic.oup.com/cercor/article-pdf/19/11/2616/1073442/bhp016.pdf. https://doi.org/10.1093/cercor/bhp016 (Mar. 2009).

103. Whitney, D., Westwood, D. A. & Goodale, M. A. The influence of visual motion on fast reaching movements to a stationary object. *Nature* **423,** 869–873 (2003).

104. Mathis, A. *et al.* DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. *Nature neuroscience* **21,** 1281–1289 (2018).

105. Hebert, J. S. *et al.* Quantitative eye gaze and movement differences in visuomotor adaptations to varying task demands among upper-extremity prosthesis users. *JAMA network open* **2,** e1911197–e1911197 (2019).

106. SCCN. *Lab Streaming Layer* Mar. 2021.

107. Scheel, C. & Staadt, O. *Mobile 3D gaze tracking calibration* in *2015 12th Conference on Computer and Robot Vision* (2015), 176–183.

108. Browatzki, B., Bülthoff, H. H. & Chuang, L. L. A comparison of geometric-and regression-based mobile gaze-tracking. *Frontiers in human neuroscience* **8,** 200 (2014).

109. Ohno, T. & Mukawa, N. *A free-head, simple calibration, gaze tracking system that enables gaze-based interaction* in *Proceedings of the 2004 symposium on Eye tracking research & applications* (2004), 115–122.

110. Boser, Q. A. *et al.* Cluster-based upper body marker models for three-dimensional kinematic analysis: Comparison with an anatomical model and reliability analysis. *Journal of biomechanics* **72,** 228–234 (2018).

111. D'Errico, J. inpaint_nans. *MATLAB Central File Exchange* (2012).

112. JASP Team. *JASP (Version)[Computer software]* 2021. https://jasp-stats.org/.

113. Cramer, A. O. *et al.* Hidden multiplicity in exploratory multiway ANOVA: Prevalence and remedies. *Psychonomic bulletin & review* **23,** 640–647 (2016).

114. Zangemeister, W. H. & Stark, L. Active head rotations and eye-head coordination. *Annals of the New York Academy of Sciences* **374,** 540–559 (1981).

115. Prime, S. L. & Marotta, J. J. Gaze strategies during visually-guided versus memory-guided grasping. *Experimental brain research* **225,** 291–305 (2013).

116. Law, E. L.-C., van Schaik, P. & Roto, V. Attitudes towards user experience (UX) measurement. *Int. J. Hum. Comput. Stud.* **72,** 526–541. ISSN: 1071-5819. https://www.sciencedirect.com/science/article/pii/S1071581913001304 (June 2014).

117. Hinderks, A., Schrepp, M., Domıénguez Mayo, F. J., Escalona, M. J. & Thomaschewski, J. Developing a UX KPI based on the user experience questionnaire. *Comput. Stand. Interfaces* **65,** 38–44. ISSN: 0920-5489. https://www.sciencedirect.com/science/article/pii/S0920548918301247 (July 2019).

118. Huang, J., White, R. & Buscher, G. *User see, user point: gaze and cursor alignment in web search* in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Association for Computing Machinery, Austin, Texas, USA, May 2012), 1341–1350. ISBN: 9781450310154. https://doi.org/10.1145/2207676.2208591.

119. Papoutsaki, A., Laskey, J. & Huang, J. *SearchGazer: Webcam Eye Tracking for Remote Studies of Web Search* in *Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval* (Association for Computing Machinery, Oslo, Norway, Mar. 2017), 17–26. ISBN: 9781450346771. https://doi.org/10.1145/3020165.3020170.

120. Papoutsaki, A. *Scalable Webcam Eye Tracking by Learning from User Interactions* in *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems* (Association for Computing Machinery, Seoul, Republic of Korea, Apr. 2015), 219–222. ISBN: 9781450331463. https://doi.org/10.1145/2702613.2702627.

121. Hauger, D., Paramythis, A. & Weibelzahl, S. *Using Browser Interaction Data to Determine Page Reading Behavior* in *User Modeling, Adaption and Personalization* (Springer Berlin Heidelberg, 2011), 147–158. http://dx.doi.org/10.1007/978-3-642-22362-4_13.

122. Land, M. F. & Hayhoe, M. In what ways do eye movements contribute to everyday activities? en. *Vision Res.* **41,** 3559–3565. ISSN: 0042-6989. http://dx.doi.org/10.1016/s0042-6989(01)00102-x (2001).

123. Gallivan, J. P., Chapman, C. S., Wolpert, D. M. & Flanagan, J. R. Decision-making in sensorimotor control. *Nature Reviews Neuroscience* **19,** 519–534 (2018).

124. Stone, S. A. *et al.* Generating accurate 3D gaze vectors using synchronized eye tracking and motion capture. en. *Behav. Res. Methods.* ISSN: 1554-351X, 1554-3528. http://dx.doi.org/10.3758/s13428-022-01958-6 (Sept. 2022).

125. Nunnally, B. & Farkas, D. *UX Research: Practical Techniques for Designing Better Products* en. ISBN: 9781491951262. https://play.google.com/store/books/details?id=sfOeDQAAQBAJ ("O'Reilly Media, Inc.", Nov. 2016).

126. Guo, Q., Jin, H., Lagun, D., Yuan, S. & Agichtein, E. *Mining touch interaction data on mobile devices to predict web search result relevance* in *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval* (Association for Computing Machinery, Dublin, Ireland, July 2013), 153–162. ISBN: 9781450320344. https://doi.org/10.1145/2484028.2484100.

127. Westwood, D. A. & Goodale, M. A. Perceptual illusion and the real-time control of action. en. *Spat. Vis.* **16,** 243–254. ISSN: 0169-1015. http://dx.doi.org/10.1163/156856803322467518 (2003).

128. Ballard, D. H., Hayhoe, M. M., Li, F. & Whitehead, S. D. Hand-eye coordination during sequential tasks. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **337,** 331–339 (1992).

129. Deng, S., Chang, J., Kirkby, J. A. & Zhang, J. J. Gaze–mouse coordinated movements and dependency with coordination demands in tracing. *Behav. Inf. Technol.* **35,** 665–679. ISSN: 0144-929X. https://doi.org/10.1080/0144929X.2016.1181209 (Aug. 2016).

130. Finger, H., Goeke, C., Diekamp, D., Standvoß, K. & König, P. *LabVanced: a unified JavaScript framework for online studies* in *International Conference on Computational Social Science (Cologne)* (2017). https://www.researchgate.net/profile/Caspar-Goeke/publication/322273524_LabVanced_A_Unified_JavaScript_Framework_for_Online_Studies/links/5a4f7ac64585151ee284d8c2/LabVanced-A-Unified-JavaScript-Framework-for-Online-Studies.pdf.

131. jamovi. *jamovi (version 2.2.5)[Computer software]* 2022. https://www.jamovi.org/.

132. Mejtoft, T., Hale, S. & Söderström, U. *Design friction* in *Proceedings of the 31st European Conference on Cognitive Ergonomics* (2019), 41–44.

133. Bridgeman, B. Conscious vs Unconscious Processes: The Case of Vision. *Theory Psychol.* **2,** 73–88. ISSN: 0959-3543. https://doi.org/10.1177/0959354392021004 (Feb. 1992).

134. Wisiecka, K. *et al. Comparison of Webcam and Remote Eye Tracking* in *2022 Symposium on Eye Tracking Research and Applications* (2022), 1–7.

135. Chisholm, J. D. *et al. A Cognitive Ethology Study of First- and Third-Person Perspectives* 2014. http://dx.doi.org/10.1371/journal.pone.0092696.

136. Smilek, D., Birmingham, E., Cameron, D., Bischof, W. & Kingstone, A. Cognitive Ethology and exploring attention in real-world scenes. en. *Brain Res.* **1080,** 101–119. ISSN: 0006-8993. http://dx.doi.org/10.1016/j.brainres.2005.12.090 (Mar. 2006).

137. Yang, X. & Krajbich, I. *Webcam-based online eye-tracking for behavioral research* Dec. 2020. http://dx.doi.org/10.31234/osf.io/qhme6.

138. Holland, C. & Komogortsev, O. V. *Biometric identification via eye movement scanpaths in reading* in *2011 International Joint Conference on Biometrics (IJCB)* (Oct. 2011), 1–8. http://dx.doi.org/10.1109/IJCB.2011.6117536.

139. Baloh, R. W. Vertigo. en. *Lancet* **352,** 1841–1846. ISSN: 0140-6736, 1474-547X. http://www.thelancet.com/article/S0140673698054300/abstract (Dec. 1998).

140. Karatas, M. Central vertigo and dizziness: epidemiology, differential diagnosis, and common causes. en. *Neurologist* **14,** 355–364. ISSN: 1074-7931. http://dx.doi.org/10.1097/NRL.0b013e31817533a3 (Nov. 2008).

141. Saber Tehrani, A. S. *et al.* Rising annual costs of dizziness presentations to U.S. emergency departments. en. *Acad. Emerg. Med.* **20,** 689–696. ISSN: 1069-6563, 1553-2712. http://dx.doi.org/10.1111/acem.12168 (July 2013).

142. Newman-Toker, D. E. *et al.* Spectrum of dizziness visits to US emergency departments: cross-sectional analysis from a nationally representative sample. en. *Mayo Clin. Proc.* **83,** 765–775. ISSN: 0025-6196, 1942-5546. http://dx.doi.org/10.4065/83.7.765 (July 2008).

143. Cheung, C. S. K. *et al.* Predictors of important neurological causes of dizziness among patients presenting to the emergency department. en. *Emerg. Med. J.* **27,** 517–521. ISSN: 1472-0205, 1472-0213. http://dx.doi.org/10.1136/emj.2009.078014 (July 2010).

144. Edlow, J. A., Gurley, K. L. & Newman-Toker, D. E. A New Diagnostic Approach to the Adult Patient with Acute Dizziness. en. *J. Emerg. Med.* **54,** 469–483. ISSN: 0736-4679. http://dx.doi.org/10.1016/j.jemermed.2017.12.024 (Apr. 2018).

145. Kovacs, E., Wang, X. & Grill, E. Economic burden of vertigo: a systematic review. en. *Health Econ. Rev.* **9,** 37. ISSN: 2191-1991. http://dx.doi.org/10.1186/s13561-019-0258-2 (Dec. 2019).

146. Kattah, J. C., Talkad, A. V., Wang, D. Z., Hsieh, Y.-H. & Newman-Toker, D. E. HINTS to diagnose stroke in the acute vestibular syndrome: three-step bedside oculomotor examination more sensitive than early MRI diffusion-weighted imaging. en. *Stroke* **40,** 3504–3510. ISSN: 0039-2499, 1524-4628. http://dx.doi.org/10.1161/STROKEAHA.109.551234 (Nov. 2009).

147. Kung, N. H., Van Stavern, G. P. & Gold, D. R. HINTS in the Acute Vestibular Syndrome: Pearls and Pitfalls. en. *J. Neuroophthalmol.* **38,** 244–250. ISSN: 1070-8022, 1536-5166. http://dx.doi.org/10.1097/WNO.0000000000000608 (June 2018).

148. McDowell, T. & Moore, F. The Under-Utilization of the Head Impulse Test in the Emergency Department. en. *Can. J. Neurol. Sci.* **43,** 398–401. ISSN: 0317-1671. http://dx.doi.org/10.1017/cjn.2015.330 (May 2016).

149. Cohn, B. Can bedside oculomotor (HINTS) testing differentiate central from peripheral causes of vertigo? en. *Ann. Emerg. Med.* **64,** 265–268. ISSN: 0196-0644, 1097-6760. http://dx.doi.org/10.1016/j.annemergmed.2014.01.010 (Sept. 2014).

150. Kattah, J. C. Update on HINTS Plus, With Discussion of Pitfalls and Pearls. en. *J. Neurol. Phys. Ther.* **43 Suppl 2,** S42–S45. ISSN: 1557-0576, 1557-0584. http://dx.doi.org/10.1097/NPT.0000000000000274 (Apr. 2019).

151. Khullar, D. *et al.* Public vs physician views of liability for artificial intelligence in health care. en. *J. Am. Med. Inform. Assoc.* **28,** 1574–1577. ISSN: 1067-5027, 1527-974X. http://dx.doi.org/10.1093/jamia/ocab055 (July 2021).

152. Dai, T. & Tayur, S. R. *Designing AI-Augmented Healthcare Delivery Systems for Physician Buy-in and Patient Acceptance* http://dx.doi.org/10.2139/ssrn.4044140.

153. Shahbabu, B., Dasgupta, A., Sarkar, K. & Sahoo, S. K. Which is More Accurate in Measuring the Blood Pressure? A Digital or an Aneroid Sphygmomanometer. en. *J. Clin. Diagn. Res.* **10,** LC11–4. ISSN: 2249-782X, 0973-709X. http://dx.doi.org/10.7860/JCDR/2016/14351.7458 (Mar. 2016).

154. Goebel, J. A. The ten-minute examination of the dizzy patient. en. *Semin. Neurol.* **21,** 391–398. ISSN: 0271-8235. http://dx.doi.org/10.1055/s-2001-19410 (Dec. 2001).

155. Post, R. E. & Dickerson, L. M. Dizziness: a diagnostic approach. en. *Am. Fam. Physician* **82,** 361–8, 369. ISSN: 0002-838X, 1532-0650. https://www.ncbi.nlm.nih.gov/pubmed/20704166 (Aug. 2010).

156. Thier, P. & Ilg, U. J. The neural basis of smooth-pursuit eye movements. en. *Curr. Opin. Neurobiol.* **15,** 645–652. ISSN: 0959-4388. http://dx.doi.org/10.1016/j.conb.2005.10.013 (Dec. 2005).

157. Hessels, R. S., van Doorn, A. J., Benjamins, J. S., Holleman, G. A. & Hooge, I. T. C. Task-related gaze control in human crowd navigation. en. *Atten. Percept. Psychophys.* **82,** 2482–2501. ISSN: 1943-3921, 1943-393X. http://dx.doi.org/10.3758/s13414-019-01952-9 (July 2020).

158. Bianchi, L. J., Kingstone, A. & Risko, E. F. The role of cognitive load in modulating social looking: a mobile eye tracking study. en. *Cogn Res Princ Implic* **5,** 44. ISSN: 2365-7464. http://dx.doi.org/10.1186/s41235-020-00242-5 (Sept. 2020).

159. Killeen, D. E., Isaacson, B. & Walter Kutz, J. *History and Physical Examination of the Dizzy Patient* 2019. http://dx.doi.org/10.1007/978-3-319-97858-1_3.

160. Kim, J. S. *et al.* Upbeat nystagmus: clinicoanatomical correlations in 15 patients. en. *J. Clin. Neurol.* **2,** 58–65. ISSN: 1738-6586, 2005-5013. http://dx.doi.org/10.3988/jcn.2006.2.1.58 (Mar. 2006).

161. Abel, L. A. & Ziegler, A. S. Smooth pursuit eye movements in schizophrenics—What constitutes quantitative assessment? *Biol. Psychiatry* **24,** 747–761. ISSN: 0006-3223. https://www.sciencedirect.com/science/article/pii/0006322388902508 (Nov. 1988).

162. Cohen, B. & Raphan, T. in *The Vestibular System* (eds Highstein, S. M., Fay, R. R. & Popper, A. N.) 235–285 (Springer New York, New York, NY, 2004). ISBN: 9780387215679. https://doi.org/10.1007/0-387-21567-0_6.

163. Legland, D. *MatGeom: Matlab geometry toolbox for 2D/3D geometric computing* 2009.

164. Wolpert, D. M., Miall, R. C. & Kawato, M. Internal models in the cerebellum. en. *Trends Cogn. Sci.* **2,** 338–347. ISSN: 1364-6613. http://dx.doi.org/10.1016/s1364-6613(98)01221-2 (Sept. 1998).

165. Paige, G. D. Nonlinearity and asymmetry in the human vestibulo-ocular reflex. en. *Acta Otolaryngol.* **108,** 1–8. ISSN: 0001-6489. http://dx.doi.org/10.3109/00016488909107385 (1989).

166. Halmagyi, G. M. & Curthoys, I. S. A clinical sign of canal paresis. en. *Arch. Neurol.* **45,** 737–739. ISSN: 0003-9942. http://dx.doi.org/10.1001/archneur.1988.00520310043015 (July 1988).

167. Land, M. F. The coordination of rotations of the eyes, head and trunk in saccadic turns produced in natural situations. en. *Exp. Brain Res.* **159,** 151–160. ISSN: 0014-4819. http://dx.doi.org/10.1007/s00221-004-1951-9 (Nov. 2004).

168. Strupp, M., Dieterich, M. & Brandt, T. The treatment and natural course of peripheral and central vertigo. en. *Dtsch. Arztebl. Int.* **110,** 505–15, quiz 515–6. ISSN: 1866-0452. http://dx.doi.org/10.3238/arztebl.2013.0505 (July 2013).

169. Robinson, F. R., Fuchs, A. F. & Noto, C. T. Cerebellar influences on saccade plasticity. en. *Ann. N. Y. Acad. Sci.* **956,** 155–163. ISSN: 0077-8923. http://dx.doi.org/10.1111/j.1749-6632.2002.tb02816.x (Apr. 2002).

170. Choi, S. C., Narayan, R. K., Anderson, R. L. & Ward, J. D. Enhanced specificity of prognosis in severe head injury. en. *J. Neurosurg.* **69,** 381–385. ISSN: 0022-3085. http://dx.doi.org/10.3171/jns.1988.69.3.0381 (Sept. 1988).

171. Meyer, S., Gibb, T. & Jurkovich, G. J. Evaluation and significance of the pupillary light reflex in trauma patients. en. *Ann. Emerg. Med.* **22,** 1052–1057. ISSN: 0196-0644. http://dx.doi.org/10.1016/s0196-0644(05)82750-7 (June 1993).

172. Kim, T. J. *et al.* Neurological Pupil Index as an Indicator of Neurological Worsening in Large Hemispheric Strokes. en. *Neurocrit. Care* **33,** 575–581. ISSN: 1541-6933, 1556-0961. http://dx.doi.org/10.1007/s12028-020-00936-0 (Oct. 2020).

173. Hoffmann, M. *et al.* Pupil evaluation in addition to Glasgow Coma Scale components in prediction of traumatic brain injury and mortality. en. *Br. J. Surg.* **99 Suppl 1,** 122–130. ISSN: 0007-1323, 1365-2168. http://dx.doi.org/10.1002/bjs.7707 (Jan. 2012).

174. Hoffman, J. E. & Subramaniam, B. The role of visual attention in saccadic eye movements. en. *Percept. Psychophys.* **57,** 787–795. ISSN: 0031-5117. http://dx.doi.org/10.3758/bf03206794 (Aug. 1995).

175. Escarce, J. J. Externalities in hospitals and physician adoption of a new surgical technology: an exploratory analysis. en. *J. Health Econ.* **15,** 715–734. ISSN: 0167-6296. http://dx.doi.org/10.1016/s0167-6296(96)00501-2 (Dec. 1996).

176. Holleman, G. A., Hooge, I. T. C., Kemner, C. & Hessels, R. S. The 'Real-World Approach' and Its Problems: A Critique of the Term Ecological Validity. en. *Front. Psychol.* **11,** 721. ISSN: 1664-1078. http://dx.doi.org/10.3389/fpsyg.2020.00721 (Apr. 2020).

177. Lions, C., Bui-Quoc, E., Wiener-Vacher, S., Seassau, M. & Bucci, M. P. Smooth pursuit eye movements in children with strabismus and in children with vergence deficits. en. *PLoS One* **8,** e83972. ISSN: 1932-6203. http://dx.doi.org/10.1371/journal.pone.0083972 (Dec. 2013).

178. Ward, M. M., Jaana, M., Bahensky, J. A., Vartak, S. & Wakefield, D. S. Clinical information system availability and use in urban and rural hospitals. en. *J. Med. Syst.* **30,** 429–438. ISSN: 0148-5598. http://dx.doi.org/10.1007/s10916-006-9014-3 (Dec. 2006).

179. Singh, R., Lichter, M. I., Danzo, A., Taylor, J. & Rosenthal, T. The adoption and use of health information technology in rural areas: results of a national survey. en. *J. Rural Health* **28,** 16–27. ISSN: 0890-765X, 1748-0361. http://dx.doi.org/10.1111/j.1748-0361.2011.00370.x (Jan. 2012).

180. Dabasia, P. L., Edgar, D. F., Garway-Heath, D. F. & Lawrenson, J. G. A survey of current and anticipated use of standard and specialist equipment by UK optometrists. en. *Ophthalmic Physiol. Opt.* **34,** 592–613. ISSN: 0275-5408, 1475-1313. https://onlinelibrary.wiley.com/doi/10.1111/opo.12150 (Sept. 2014).

181. Fogt, N., Baughman, B. J. & Good, G. The effect of experience on the detection of small eye movements. en. *Optom. Vis. Sci.* **77,** 670–674. ISSN: 1040-5488. http://dx.doi.org/10.1097/00006324-200012000-00014 (Dec. 2000).

182. Brunswik, E. *Representative design and probabilistic theory in a functional psychology* 1955. http://dx.doi.org/10.1037/h0047470.

183. Wispinski, N. J. *et al.* Reaching for known unknowns: Rapid reach decisions accurately reflect the future state of dynamic probabilistic information. en. *Cortex* **138,** 253–265. ISSN: 0010-9452, 1973-8102. http://dx.doi.org/10.1016/j.cortex.2021.02.010 (May 2021).

184. Diamond, J. S., Wolpert, D. M. & Flanagan, J. R. Rapid target foraging with reach or gaze: The hand looks further ahead than the eye. en. *PLoS Comput. Biol.* **13,** e1005504. ISSN: 1553-734X, 1553-7358. http://dx.doi.org/10.1371/journal.pcbi.1005504 (July 2017).

185. Stone, S. A. *et al.* in *ACM Symposium on Eye Tracking Research and Applications* 1–4 (Association for Computing Machinery, New York, NY, USA, May 2021). ISBN: 9781450383578. https://doi.org/10.1145/3450341.3458880.

186. Mejtoft, T., Hale, S. & Söderström, U. *Design Friction* in *Proceedings of the 31st European Conference on Cognitive Ergonomics* (Association for Computing Machinery, BELFAST, United Kingdom, Sept. 2019), 41–44. ISBN: 9781450371667. https://doi.org/10.1145/3335082.3335106.

187. Kahneman, D. *Thinking, Fast and Slow* en. ISBN: 9780385676526. https://play.google.com/store/books/details?id=TA7Q27RWlj0C (Doubleday Canada, Nov. 2011).

188. Craig Tomlin, W. *UX Optimization: Combining Behavioral UX and Usability Testing Data to Optimize Websites* en. ISBN: 9781484238677. https://play.google.com/store/books/details?id=7rJwDwAAQBAJ (Apress, Sept. 2018).

189. Snow, J. C. & Culham, J. C. The Treachery of Images: How Realism Influences Brain and Behavior. en. *Trends Cogn. Sci.* **25,** 506–519. ISSN: 1364-6613, 1879-307X. http://dx.doi.org/10.1016/j.tics.2021.02.008 (June 2021).

190. Marini, F., Breeding, K. A. & Snow, J. C. Distinct visuo-motor brain dynamics for real-world objects versus planar images. en. *Neuroimage* **195,** 232–242. ISSN: 1053-8119, 1095-9572. http://dx.doi.org/10.1016/j.neuroimage.2019.02.026 (July 2019).

191. Kingstone, A., Smilek, D., Ristic, J., Kelland Friesen, C. & Eastwood, J. D. Attention, Researchers! It Is Time to Take a Look at the Real World. *Curr. Dir. Psychol. Sci.* **12,** 176–180. ISSN: 0963-7214. https://doi.org/10.1111/1467-8721.01255 (Oct. 2003).

Appendix A: ETRA 3D gaze vector publication

Sub-centimeter 3D gaze vector accuracy on real-world tasks: an investigation of eye and motion capture calibration routines

Scott A. Stone
Department of Psychology
University of Alberta
Edmonton, Alberta, Canada
sastone@ualberta.ca

Quinn A. Boser
Division of Physical Medicine and Rehabilitation, Department of Medicine, Faculty of Medicine and Dentistry
University of Alberta
Edmonton, Alberta, Canada
boser@ualberta.ca

T. Riley Dawson
Division of Physical Medicine and Rehabilitation, Department of Medicine, Faculty of Medicine and Dentistry
University of Alberta
Edmonton, Alberta, Canada
rileydawson@ualberta.ca

Albert H. Vette
Department of Mechanical Engineering
University of Alberta
Edmonton, Alberta, Canada
Glenrose Rehabilitation Hospital
Alberta Health Services
Edmonton, Alberta, Canada
vette@ualberta.ca

Jacqueline S. Hebert
Division of Physical Medicine and Rehabilitation, Department of Medicine, Faculty of Medicine and Dentistry
University of Alberta
Edmonton, Alberta, Canada
jhebert@ualberta.ca

Patrick M. Pilarski
Division of Physical Medicine and Rehabilitation, Department of Medicine, Faculty of Medicine and Dentistry
University of Alberta
Edmonton, Alberta, Canada
Alberta Machine Intelligence Institute
Edmonton, Alberta, Canada
pilarski@ualberta.ca

Craig S. Chapman
Faculty of Kinesiology, Sport, and Recreation
University of Alberta
Edmonton, Alberta, Canada
Neuroscience and Mental Health Institute
University of Alberta
Edmonton, Alberta, Canada
c.s.chapman@ualberta.ca

ABSTRACT

Measuring where people look in real-world tasks has never been easier but analyzing the resulting data remains laborious. One solution integrates head-mounted eye tracking with motion capture but no best practice exists regarding what calibration data to collect. Here, we compared four ~1 min calibration routines used to train linear regression gaze vector models and examined how the coordinate system, eye data used and location of fixation changed gaze vector accuracy on three trial types: calibration, validation (static fixation to task relevant locations), and task (naturally occurring fixations during object interaction). Impressively, predicted gaze vectors show ~1 cm of error when looking straight ahead toward objects during natural arms-length interaction. This result was achieved predicting fixations in a Spherical coordinate frame, from the best monocular data, and, surprisingly, depends little on the calibration routine.

CCS CONCEPTS

• **Human-centered computing → Interaction devices**.

KEYWORDS

eye tracking, motion capture, gaze vector, calibration, linear regression

ETRA '21 Adjunct, May 25–27, 2021, Virtual Event, Germany

ACM ISBN 978-1-4503-8357-8/21/05…$15.00
https://doi.org/10.1145/3450341.3458880

ACM Reference Format:
Scott A. Stone, Quinn A. Boser, T. Riley Dawson, Albert H. Vette, Jacqueline S. Hebert, Patrick M. Pilarski, and Craig S. Chapman. 2021. Sub-centimeter 3D gaze vector accuracy on real-world tasks: an investigation of eye and motion capture calibration routines. In *2021 Symposium on Eye Tracking Research and Applications (ETRA '21 Adjunct), May 25–27, 2021, Virtual*

Event, Germany. ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3450341.3458880

1 INTRODUCTION

Eye tracking has become cheaper and easier to use. However, video-based analysis requires time-consuming manual labour. This analysis can be simplified by integrating eye tracking with three-dimensional (3D) motion capture (mo-cap) to generate 3D gaze vectors (GVs) in the mo-cap space. Recently, open source solutions have made it easier to merge multiple data streams. No standard method exists for collecting eye/mo-cap calibration data or using it to generate GVs [Nyström et al. 2013]. Here, we investigate four calibration routines to address this gap. We also explore using monocular (only left or right eye) versus binocular (both eyes) data and the effect the coordinate system (CS) has on accuracy when used to predict gaze fixation. We ask two main questions: 1) which combination of calibration, eye data, and CS type produces GVs with the highest accuracy? and 2) how accurate are GVs that are generated from 2D eye and 3D mo-cap data?

2 METHODS

2.1 Equipment

Eye tracking data were collected using a Pupil Labs Core USB headset (Pupil Core v1.8;[Kassner et al. 2014]) and synchronized with mo-cap data using Lab Streaming Layer (LSL; [SCCN 2021]). Mo-cap data were collected using an OptiTrack mo-cap system (two systems: 12-camera Prime 13W system, 200Hz; and 12-camera Flex 13 system, 120Hz). Reflective markers tracked position and orientation of the Head, Right Hand, Task Cart, Side Cart, Pasta box (task), and Calibration Wand (calibration). Any combination of eye and motion trackers could be used, provided they collect time series data as 3D marker position and 2D pupil positions.

2.2 Participants

Twenty-one undergraduate and graduate University of Alberta Department of Psychology students participated. Eight of these participants were collected at 120 Hz and 13 were collected at 200 Hz. One 200 Hz participant was removed due to unusable data, for a total of twenty subjects. This study was approved by the University of Alberta Health Research Ethics Board under protocol Pro00087329.

2.3 Procedure

Each experiment consisted of 3 sets of Calibration/Validation trials (cal. set) and 2 sets of Task trials, proceeding in the order: cal. set > Task trials > cal. set > Task trials > cal. set. Each calibration/validation set included four calibrations (one of each type) and one validation, in a random order. Each set of task trials included 10 repetitions of the previously published Pasta Box task ([Lavoie et al. 2018; Valevicius et al. 2018]; see Figure 1). In total, participants performed 12 Calibration trials, 3 Validation trials and 20 Task trials.

2.3.1 Calibration Trials. Participants fixated on a mo-cap marker (14 mm) for ~1 min per trial. The marker was placed at the tip of a 40 cm wand which moved through the task space in one of four calibration routines (CR; see Figure 2 for example pupil data):

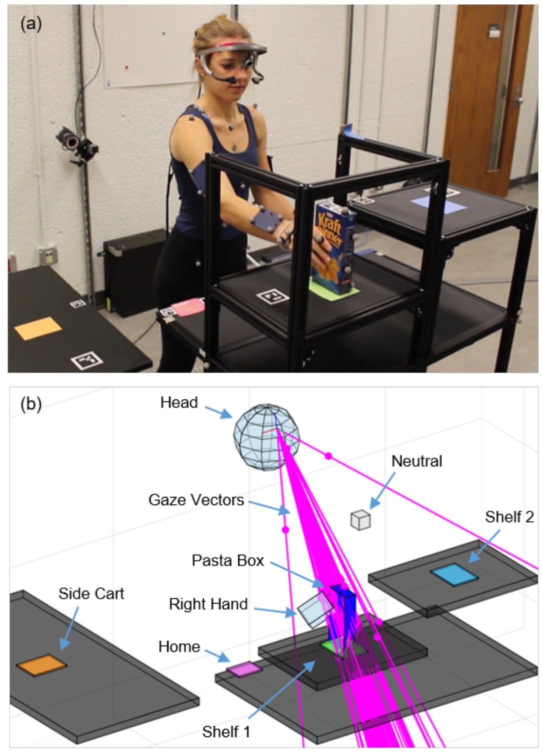


Figure 1: The Pasta Box Task. (a) A real-world photo the Pasta Box Task. (b) A labeled screenshot of GaMA analysis software, the task reconstruction and the 72 Gaze Vectors generated for this analysis.

(1) Experimenter Sweep (ES): The experimenter moved the wand in slow S-shaped curves along each of the room-coordinate axes (parallel to floor left/right, parallel to floor in/out, parallel to wall up/down).
(2) Self Sweep (SS): Replicating ES but with the participant holding and moving the wand.
(3) Experimenter Paint (EP): The experimenter moved the wand to each of the task relevant locations (minus Neutral, see below) and explored small (10-20 cm in each dimension) volumes at these locations.
(4) Stationary Target (ST): The wand was fixed in front of the participant (~60 cm), who maintained fixation while nodding their head, turning it, then rotating it in a clockwise then counterclockwise spiral.

2.3.2 Validation Trials. Participants were asked to fixate on five stationary targets presented at task relevant locations (see Figure 1) for ~5 s. A beep cued the first fixation and a beep every 5 s cued switches to each location: Neutral ⟶ Side Cart ⟶ Shelf 1 ⟶ Home ⟶ Shelf 1 ⟶ Shelf 2 ⟶ Home ⟶ Shelf 2 ⟶ Side Cart ⟶ Home ⟶ Neutral.

2.3.3 Task Trials. The Pasta Box task is shown in Figure 1. Participants began with their hand on Home and eyes fixating on Neutral. A beep cued them to initiate a sequence of 3 movements: 1) Grasp the Pasta Box at the Side Cart, move it to Shelf 1 then return hand to Home; 2) Grasp the Pasta Box at Shelf 1, move it to Shelf 2 then return the hand to Home; 3) Grasp the Pasta Box at Shelf 2, move it to the Side Cart then return the hand to Home.

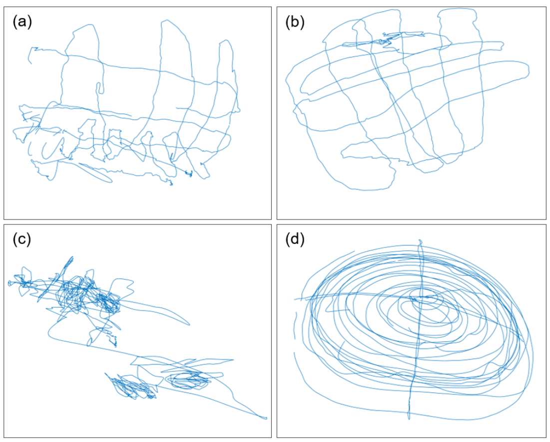


Figure 2: Sample eye tracking data from one pupil (pupil y-coordinate vs. x-coordinate) for each of the four calibration routines: (a) Experimenter Sweep (ES), (b) Self Sweep (SS), (c) Experimenter Paint (EP), (d) Stationary Target (ST).

2.4 Data Processing

2.4.1 Pre-processing. Mo-cap and eye tracking data were filtered (Butterworth low pass: mo-cap: 6 Hz; eye: 10 Hz) and cleaned (mo-cap: gap filling, marker swaps; eye data: gap filling, outlier removal) using automated scripts.

2.4.2 Gaze Vector Modelling. Creating a GV consists of 1) generating three eye gaze models using data from a specific Calibration trial, then 2) using the models to predict the GV direction at each frame in a given trial.

In step 1, calibration data is used to fit three eye gaze models. Each model takes eye data as input and predicts a single coordinate of the 3D gaze fixation point relative to the Head rigid body CS. Each model was generated using the built-in MATLAB function fitlm with the 'quadratic' model specification and robust fitting using the 'bisquare' weight function. In step 2, the predicted fixation point is transformed into the global mo-cap CS, and the GV is represented by a line originating at the Head and extending infinitely through the fixation point (i.e. the depth of fixation or radial distance from the head to fixation point was not relevant in subsequent analyses).

We explored three options for model input (eye data from right eye only, left eye only, or binocular data), as well as two options for expressing the fixation point relative to the Head CS (Cartesian $[x,y,z]$ coordinates, or Spherical $[r, \theta, \phi]$ coordinates). We anticipated that using the Spherical CS would increase accuracy of the GV direction because it isolates depth of fixation to the 'r' model, whereas in Cartesian, all three models are influenced by depth of fixation.

3 RESULTS

We generated 72 possible GVs (12 CR x 3 eye data sets x 2 Head CS) on each trial of the following types. Analysis was performed using the "best" (lowest average distance to targets) GVs of each CR. Statistical analysis was performed in JASP 0.14.1.0 [JASP Team 2020]. Repeated measures ANOVAs (α = 0.05) were used to analyze the 3 trial types, which used the same participant pool but were statistically independent of one another. GVs whose average distance from the target was more than 30 cm were removed. After removal, Calibration n=19, Validation n=17, and Task n=18 participants.

3.1 Calibration Trials

We measured mean absolute distance between each GV and the Wand Tip over each Calibration trial (excluding the trial a GV was trained on).

Collapsing across all variables except Head CS, error was lower when using a Spherical CS (Cart: 62.7 mm; Sph: 50.6 mm). When collapsing across all variables except CR used to train the models, ST performed the best overall (ES: 56.2 mm; SS: 55.2 mm; EP: 69.8 mm; ST: 45.4 mm). However, looking at how well data from each CR can be predicted by other GVs, ST is hardest to predict (ES: 48.6 mm; SS: 48.8 mm; EP: 48.5 mm; ST: 80.7 mm).

3.2 Validation Trials

We measured the mean 3D distance between each GV and markers at task relevant locations. Average 3D distances were calculated over stable gaze periods of 1 s (out of the 5 s windows).

Average error was lowest using a Spherical CS (Cart: 42.9 mm, Sph: 38.7 mm). Notably, error increased when participants were looking at the Side Cart (Cart: 64.9 mm; Sph: 57.2 mm). Further, using data from only the left eye (53.2 mm) generated lower error than binocular (63.1 mm) or right eye (67.0 mm) at the Side Cart.

3.3 Task Trials

We measured the mean 3D distance between each GV and the nearest face of the Neutral (4 cm cube, at trial start) and Pasta Box (9 x 4 x 18 cm, at the start of each grasp and release) object (distance was 0 mm if GV intersected object). Previous work with this and other grasping tasks has shown that people look at objects they are interacting with at these times [Land and Hayhoe 2001; Lavoie et al. 2018; Williams et al. 2019].

Average error was sub-centimeter towards front-facing targets when using the best monocular data and a Spherical CS (8.7 mm), while a Cartesian CS (10.6 mm) produced ~1 cm accuracy. The average error when gazing at the Side Cart was relatively worse overall (Sph: 28.5 mm; Cart: 40.5 mm).

4 DISCUSSION

We generated GVs using combined eye tracking and mo-cap data. Calibration data (~1 min of eyes continuously fixating a tracked marker) were used to train linear models to predict 3D gaze direction. We tested four different routines, three options for model input, and two options for model CS. The resulting GV accuracy was assessed on Calibration trials, Validation trials, and Task trials to produce best practice recommendations.

No one calibration routine consistently outperformed the others. When assessed on other calibration data, GVs generated from the ST routine performed best, but this was driven by ST data being hard for other models to predict. Importantly, the consistently extreme pupil positions observed in ST data (see Figure 2d) are not typical in real-world behaviour (see Figure 2c). Instead, on the more natural Validation and Task trials there was no main effect of calibration routine. Thus, for natural, real-world tasks, any of the calibration routines are likely to produce similarly accurate results, although the ES routine produced the lowest nominal mean errors on Task trials (< 4 mm for looks to Neutral).

One surprising result was that GVs generated using one eye's data were often more accurate than using both. We still recommend collecting binocular eye data, but suggest generating GVs for each eye individually, and determining which eye or combination is most accurate for a given participant and task.

Using a Spherical CS produces GVs with a more accurate direction than a Cartesian CS, aligning with our prediction. This indicates that when only the direction of gaze is of interest, a Spherical CS should be used to generate GVs. Future work will explore the accuracy of the depth of the fixation point.

The accuracy of the GVs generated with this method was impressive. In the best case, sub-centimeter accuracy was achieved for functional task data using an ES calibration routine and Spherical CS model. In all results (except predicting ST calibration routines), error was below 10 cm, and most errors were below 5 cm. Of note, within the 72 GVs generated for each participant, most participants had at least one outlier GV that produced clearly inaccurate results. Thus, we recommend collecting multiple calibration routines per participant.

ACKNOWLEDGMENTS

This work was supported by an NSERC Discovery Grant (Grant no. RGPIN-2020-05396), as well as a CFI John R. Evans Leaders Fund Grant (Grant no. 35806).

REFERENCES

JASP Team. 2020. JASP (Version 0.14.1)[Computer software]. https://jasp-stats.org/

Moritz Kassner, William Patera, and Andreas Bulling. 2014. Pupil: an open source platform for pervasive eye tracking and mobile gaze-based interaction. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication* (Seattle, Washington) *(UbiComp '14 Adjunct)*. Association for Computing Machinery, New York, NY, USA, 1151–1160.

M F Land and M Hayhoe. 2001. In what ways do eye movements contribute to everyday activities? *Vision Res.* 41, 25-26 (2001), 3559–3565.

Ewen B Lavoie, Aïda M Valevicius, Quinn A Boser, Ognjen Kovic, Albert H Vette, Patrick M Pilarski, Jacqueline S Hebert, and Craig S Chapman. 2018. Using synchronized eye and motion tracking to determine high-precision eye-movement patterns during object-interaction tasks. *J. Vis.* 18, 6 (June 2018), 18.

Marcus Nyström, Richard Andersson, Kenneth Holmqvist, and Joost van de Weijer. 2013. The influence of calibration method and eye physiology on eyetracking data quality. *Behav. Res. Methods* 45, 1 (March 2013), 272–288.

SCCN. 2021. Lab Streaming Layer.

Aïda M Valevicius, Quinn A Boser, Ewen B Lavoie, Glyn S Murgatroyd, Patrick M Pilarski, Craig S Chapman, Albert H Vette, and Jacqueline S Hebert. 2018. Characterization of normative hand movements during two functional upper limb tasks. *PLoS One* 13, 6 (June 2018), e0199549.

Heather E Williams, Craig S Chapman, Patrick M Pilarski, Albert H Vette, and Jacqueline S Hebert. 2019. Gaze and Movement Assessment (GaMA): Inter-site validation of a visuomotor upper limb functional protocol. *PLoS One* 14, 12 (Dec. 2019), e0219333.

Appendix B: Supplemental mouse movement statistics

1 SUPPLEMENTAL: MOUSE MOVEMENT STATISTICS

Below are statistics from the mouse movements omitted from the main text of the manuscript. Data here are to help provide context to the eye and coordination measures collected.

Mouse: distance to navigate to the goal frame

This is the average distance the mouse traveled in pixels when the participant was navigating to the goal frame. In the Local cohort, a significant main effect of GoalFrame was detected (F(1,1.306) = 15.850, p = 0.003, η^2 = 0.393), where a Narrative goal frame required the most mouse movement for users to find. The Remote cohort showed the same pattern and the cohort comparison showed no significant main effects or interactions involving Cohort.

Mouse: distance on the goal frame

This is the average total distance the mouse traveled in pixels when the participant was navigating on the goal frame to complete the task outlined by the prompt. In the Local cohort, a significant main effect of GoalFrame was detected (F(1,2.623) = 6.898, p = 0.004, η^2 = 0.315), where a Sound goal frame had the most mouse movement. For the cohort comparison, a significant Condition $\times$ Cohort interaction was detected (F(1,1) = 5.274, p = 0.036, η^2 = 0.004), where Remote participants spent longer on the goal frame when given an Indirect prompt.

Mouse: distance to end trial

This is the average total distance the mouse traveled in pixels when the participant had completed the task on the intended goal frame and went to end the current trial. In the Local cohort, no significant main effects or interactions were detected. The Remote cohort showed the same pattern and the cohort comparison showed no significant main effects or interactions involving Cohort.