The University of Nottingham
UNITED KINGDOM · CHINA · MALAYSIA

Exploring the use of Brain-Sensing Technologies for Natural Interactions

Matthew Pike

Student Number - 4193658

A thesis presented for the degree of
Doctor of Philosophy

Supervised by:
Dr Max L. Wilson
Prof. Tom Rodden

University of Nottingham, UK
June 2017

I, Matthew Pike confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

This thesis was written using a combination of Markdown, LaTeX and HTML. The mechanics of generating this document was loosely based on the template provided by Tom Pollard [182], on Github under the MIT License.

Acknowledgements

In this section I would like to express my sincere appreciation to those who have supported me during the course of this PhD.

First of all, I am extremely grateful to my supervisor, Dr Max **L** Wilson, for his guidance, advice and support throughout this process. Without Max, none of the contributions I present in this thesis nor the many personal journeys that I have been fortunate enough to experience over the last 4 years, would have been possible. Thank you Max, since meeting you at Swansea University, to submitting this thesis at Nottingham University, I have thoroughly enjoyed working with you. I wish you and your family the very best in your future travels in research and life.

I must also recognise my academic colleagues at Nottingham's Mixed Reality Lab who have provided a fantastic community over the last 4 years. In no particular order: **Wenchao Jiang, Chaoyu Ye, Richard Ramchurn, Leigh Clark, James Colley, Martin Flintham, Richard Wetzel, Neha Gupta, Khalid Bashour, Nils Jäger, Martin 'Intern' Porcheron, Lesley Fosh, Tom Lodge, Lachlan Urquhart, Dimitri Darzentas, Alexandru Ghitulescu, Joel Fischer, Tom Rodden, Steve Benford** and the many others I am surely missing here (apologies!). A special mention must go to **Horia Maior**, with whom I have spent a great deal of time over the last 4 years, in numerous locations across the world. We've shared so many experiences together, with great stories that we will continue to tell for many years to come. The spirit and sense of belonging I felt in the MRL during my PhD is unlike anything I have felt before, and is something that I will miss deeply upon my departure. Cheers to you all!

I would also like to thank the administrative staff at Nottingham University, especially **Sam Stapleford-Allen, Felicia Black and Christine Fletcher**, who have been excellent and professional throughout my time at the institution.

Thank you **Dr Julie Greensmith** for your invaluable advice and critique during my annual reviews. Thank you **Dr Holger Schnädelbach** for agreeing to be my internal examiner, and to **Professor Stephen Fairclough** for being my external examiner.

I must also say a sincere thank you to my parents, Susan and Martyn Pike, for their unconditional love, patience and timely words of encouragement. I am deeply grateful for the love and support you have always shown to me in education and in life. Thank you both.

During my time at Nottingham, I have been fortunate enough to find a fantastic partner in life - **Maizan Dianati**. Maizan, we have already travelled far and wide together, sharing many great experiences along the way and I cannot wait to share many more with you. You have been a trusted companion who has unquestionably listened to my moaning along the way. I hope we can continue on our own journey for many years to come.

Diolch.

Abstract

Recent technical innovation in the field of Brain-Computer Interfaces (BCIs) has increased the opportunity for including physical, brain-sensing devices as a part of our day-to-day lives. The potential for obtaining a time-correlated, direct, brain-based measure of a participant's mental activity is an alluring and important development for HCI researchers.

In this work, we investigate the application of BCI hardware for answering HCI centred research questions, in turn, fusing the two disciplines to form an approach we name - *Brain based Human-Computer Interaction (BHCI)*. We investigate the possibility of using BHCI to provide natural interaction - an ideal form of HCI, where communication between man-and-machine is indistinguishable from everyday forms of interactions such as Speaking and Gesturing.

We present the development, execution and output of three user studies investigating the application of BHCI. We evaluate two technologies, fNIRS and EEG, and investigate their suitability for supporting BHCI based interactions. Through our initial studies, we identify that the lightweight and portable attributes of EEG make it preferable for use in developing natural interactions. Building upon this, we develop an EEG based cinematic experience exploring natural forms of interaction through the mind of the viewer. In studying the viewers response to this experience, we were able to develop a taxonomy of control based on how viewers discovered and exerted control over the experience.

Table of Contents

Acknowledgements **4**

Abstract **6**

1 Introduction **14**

1.1 Overview 14
1.2 Brain based Human Computer Interaction 15
1.3 Research Questions 17
1.4 Roadmap 20
1.5 Publications 21

2 Literature Review **23**

2.1 Introduction 23
2.2 The Brain, it's Regions and their Relevance to BHCI 25
2.2.1 Mapping the Brain 25
2.2.1.1 Cerebral Cortex 25
2.2.1.2 Brodmann Area Map 26
2.2.2 The Pre-Frontal Cortex (PFC) 26
2.2.3 The Motor Cortex 27
2.3 Mental Workload 28
2.3.1 Model of Working Memory 30
2.3.2 Modelling Mental Workload 32
2.3.2.1 Human Information Processing 33
2.3.2.2 Limited Resource Model (LRM) 34
2.3.2.3 Cognitive Load Theory (CLT) 36
2.3.3 Measures of Mental Workload 37
2.3.3.1 Task Performance 37
2.3.3.2 Subjective 38
2.3.3.3 Objective Measures of Mental Workload 40
2.4 BHCI in the Wild 49
2.4.1 Control 49
2.4.1.1 Communication and Direct Control 49

2.4.1.2 Passive Control, Adaptive Systems and User State Monitoring . . . 50
2.4.1.3 Games and Entertainment . . . 53
2.4.2 Evaluation . . . 55
2.5 Problem Statement . . . 56

3 Methodological Approach **59**
3.1 HCI Research Methods . . . 59
3.2 A Hybrid Approach . . . 61

4 Evaluating fNIRS in BHCI via Verbal Protocols **64**
4.1 Introduction . . . 64
4.2 Related Work . . . 67
4.2.1 Think-Aloud Protocols . . . 67
4.2.2 Working Memory (WM) . . . 67
4.3 Experiment Design . . . 68
4.4 Hypotheses . . . 70
4.5 Participants . . . 71
4.6 Procedure . . . 72
4.7 Measurements and Equipment . . . 74
4.7.1 Task Accuracy - *P* . . . 74
4.7.2 Task Time - *P* . . . 74
4.7.3 NASA-TLX questionnaire - *C* . . . 74
4.7.4 fNIRS data - *C* . . . 75
4.8 Experiment Software . . . 76
4.9 Task Results . . . 76
4.9.1 Performance . . . 77
4.10 Mental Workload: Subjective measure . . . 78
4.11 Mental Workload: fNIRS . . . 79
4.12 Mathematical Skill . . . 80
4.12.1 High Performers . . . 80
4.12.2 Low Performers . . . 81
4.13 Using fNIRS in HCI Studies . . . 81
4.14 Discussion . . . 82
4.14.1 fNIRS, Language and Mental Workload . . . 82
4.14.2 High and Low Performers . . . 83
4.14.3 Running a TAP . . . 83
4.14.4 Using fNIRS to Measure Mental Workload . . . 84
4.15 Conclusion . . . 84

4.15.1 Study Conclusion and Contributions 84
4.15.2 Thesis Conclusion and Contributions 86
4.15.3 Contribution to HCI Craft Knowledge 87
4.15.3.1 Validation of a widely used protocol 87
4.15.3.2 HCI Evaluation using fNIRS 88

5 Evaluating EEG in BHCI via Gesture based Input 89
5.1 Introduction 89
5.2 Related Work 92
5.2.1 Leap Motion 93
5.2.2 Puzzle Solving 94
5.3 Experiment Design 95
5.3.1 Experiment Task 96
5.3.2 Independent Variables 97
5.3.3 Hypothesis 98
5.4 Measurements and Equipment 99
5.4.1 Task Accuracy - P 99
5.4.2 Task Time - P 99
5.4.3 NASA-TLX questionnaire - C 99
5.4.4 EEG Data - C 100
5.4.5 Procedure 101
5.4.6 Participants 103
5.5 Results 103
5.5.1 Performance 104
5.5.1.1 Accuracy 105
5.5.1.2 Speed 105
5.5.2 Cognitive Measures 105
5.5.2.1 NASA-TLX 106
5.5.2.2 EEG 107
5.5.3 Post-hoc Analysis of Performance 108
5.6 Discussion 110
5.6.1 Possible Confounding Variables 110
5.6.1.1 Haptic feedback 110
5.6.1.2 Separation of input and output 111
5.6.1.3 Input Familiarity 111
5.6.1.4 Output Familiarity 111
5.6.2 Applying EEG to Natural Interaction 112
5.7 Conclusion 115
5.7.1 Study Conclusion and Contributions 115

5.7.2 Thesis Conclusion and Contributions 115
5.7.3 Contribution to HCI craft knowledge 116
5.7.3.1 Evaluation of an Novel Interaction Technique 116
5.7.3.2 Novel Study Task 117

6 #Scanners: The Application of BHCI for Interactive Cinematic Experiences 118
6.1 Introduction 118
6.1.1 Motivation 119
6.2 TWAL: Two Way Affect Loops 120
6.3 #Scanners : An Adaptive, Cinematic Experience 121
6.3.1 BHCI Technology Choice 122
6.3.2 The #Scanners Film 123
6.3.3 Study 127
6.3.3.1 Method and Approach 128
6.3.3.2 Overall Response to #Scanners 128
6.3.3.3 Discovering Control 132
6.3.3.4 Exerting Control 134
6.3.3.5 Releasing Control 135
6.3.3.6 Summary of Findings 137
6.4 Taxonomy of Passive Control 137
6.4.1 Extent of Voluntary Control 138
6.4.2 Extent of Self-Awareness 139
6.4.3 A Design Space for Entertaining BCIs 140
6.4.4 Journeys through Control 142
6.5 Conclusion 145
6.5.1 Contribution to HCI craft knowledge 145
6.5.1.1 Framework of Control 145
6.5.1.2 Verification of a HCI research methodology 146

7 Discussion 147
7.1 Applied BHCI for Evaluation 147
7.2 Novel Passive Interactions 149
7.2.1 Applying the Framework to Novel Interactions 150
Design of Interactive Content 150
Repeat Screening/Staging of Experiences 151
Collective Experiences 151
7.2.2 Applying the Framework to BHCI 153
7.3 BCI Technology and BHCI 155

7.4 Future Work . . . 159
7.4.1 Application of Machine Learning for Data Analysis . . . 159
7.4.2 Cognitively Queryable Video (CQV) . . . 160
7.4.3 Integrating BHCI into VR . . . 162

8 Conclusion **167**
8.1 Summary of Work . . . 167
8.2 Key Contributions . . . 169

References **171**

List of Figures

1.1 A typical BCI data processing pipeline. . . . 16

2.1 Annotated Cerebral Cortex. Image Source - McGill . . . 25
2.2 Labelled regions as specified by the Broadmann Area Map. Image Source - Wikipedia. . . . 26
2.3 Annotated regions of the components that form the Motor Cortex. Image Source - McGill. . . . 28
2.4 A framework for MWL definition and evaluation. Diagram from 'Evaluation of Human Work' - [241] . . . 30
2.5 Baddeley's Working Memory Model. . . . 31
2.6 Human Information Processing Model Lifecycle, from Wickens - [239] . . . 33
2.7 The 4-D multiple resource model, by Wickens. . . . 34
2.8 The relationship between task performance and resource supply, from Megaw - [239] . . . 35
2.9 Factors determining the level of cognitive load. Diagram from Kirschner - [111] . . . 36
2.10 A paper version of the NASA-TLX questionnaire. . . . 41
2.11 A participant laying ready to be scanned in an fMRI machine. Image Credit - UCL. . . . 44
2.12 The banana-shaped path taken by Infrared Light (700-900nm). Image Credit - [127] . . . 45
2.13 The Biopac fNIRS device being worn by a participant. . . . 46
2.14 The spectrum of EEG waveforms. Table is provided by Tatum [220]. . . . 47

2.15 The target expansion conditions with visual indications of expanded target widths presented to participants [5]. Note - The participants would not see the bounding area (indicated in red). 51
2.16 The brain-controlled, Tightrope walking game 'MindBalance' where users must control the balance of an animated character in a 3D environment [117]. 54
2.17 The interface for the software system CUES, which augments browser interactions recordings with EEG, audio and web event data [177]. 55

3.1 HCI methods: Relevance vs. Precision. Image Credit - MacKenzie [135] . 61
3.2 Overview of Performance-Led Research in the Wild. Image Credit - Benford et al. [19] 63

4.1 A summary of the recommendations provided by Solovey et al. Figure Source - [210]. 65
4.2 A Screenshot of the study task. 70
4.3 A participant, wearing fNIRS in the environment where the study was conducted. 72
4.4 Sensor layout for the Biopac fNIRS used. Figure provided by Hyosun Kwon. 76
4.5 Mean time to solve a set of tasks. 77
4.6 Mean Values for three sub-scales (Mental Effort, Mental Demands and Physical Demands) of NASA-TLX. 78
4.7 Overall HbO and Hb levels for each condition. 80

5.1 The Muse EEG headset used in this study. 90
5.2 The Leap Motion input device used in this study. Source - Amazon US . 94
5.3 The computerised version of the jigsaw puzzle task. Participants would use this version of the puzzle during the Mouse and Leap-Motion conditions of the study. 97
5.4 The environment in which the study was conducted. 101
5.5 The Experimental Procedure. 102
5.6 Overall participant accuracy for each Condition. 104
5.7 Overall Time-to-Solve in each Condition, relative to the available time. . 106
5.8 NASA-TLX weighted overall Workload Score. 107
5.9 Mental Workload in AF8 - By Condition. 108
5.10 Distribution of Participant's performance, highlighting the effect of Leap Condition Performance. 109
5.11 Distribution of Participant's performance, highlighting the effect of expertise. 112
5.12 Full-Head EEG device compared to the Muse. Full-head image provided by WUSTl 112

6.1 Neurosky Mindwave Mobile, the consumer grade EEG device used in #Scanners.Image Source - Amazon.com 122
6.2 The #Scanners film shot as four layers synchronized through 18 scenes, with scene cuts aligned between layers. 124
6.3 An original storyboard for the #Scanners film. 125
6.4 A #Scanners concept diagram, showing how a user moves between and within Dream and Reality layers. 126
6.5 A participant wearing the EEG device, experiencing #Scanners inside the caravan. 127
6.6 Visualisation of Attention (blue line), Meditation (green line) and state (Reality = purple bars, Dream = red bars). 129
6.7 A Visualisation of *P2's* journey through #Scanners. Above the X-Axis is attention, below it is meditation. P2 blinked 28 times. 131
6.8 A visualisation of *P16's* journey through #Scanners. Above the X-Axis is attention, below it is meditation. P16 blinked 158 times. 131
6.9 Attention levels of High (red) and Low Blinkers(blue) by scene. 132
6.10 Meditation levels for those that self-identified as trying to Control (Red) and those who did not (Blue). 136
6.11 Extent of Voluntary Control as a dimension. 138
6.12 Triggers affecting voluntary control. 139
6.13 Extent of Self-Awareness of control as a dimension. 140
6.14 Triggers affecting Self-Awareness of Control. 140
6.15 A design space for entertaining BCIs. 141
6.16 Major states participants travelled through whilst experiencing #Scanners. 143

7.1 The landscape of technology choices for application within BHCI. 156
7.2 The NirSport fNIRS device by NirX. Image Credit - NirX. 157
7.3 The Portalite device by Artinis. Image Credit - Artinis. 158
7.4 A frame from the video recording during the 4 hour drive. 161
7.5 Demonstration of the practicality of 2 types of BCI devices with 3 types of VR goggles 165

Chapter 1

Introduction

1.1 Overview

Human-Computer Interaction, often called HCI, is an interdisciplinary, socio-technological field which aims to enrich our day-to-day lives through developing engaging natural forms of interactions between man-and-machine. HCI centred research has provided the common interfaces that have enabled the now ubiquitous interactions between the technological devices we encounter as a part of our daily lives. Graphical User Interfaces (GUI), Speech Recognition systems (SR) and Gesture Recognition Systems are all examples of these.

In this context, Natural Interaction (NI), is a user interface that is immediately familiar to the user and is effectively invisible. Valli stated that natural interactions should allow users to interact as they are used to in their day-to-day lives, and that the basis of the interaction should be derived from the user's 'evolution and eduction' [226]. Examples of Natural User Interfaces (NUI) include: Speech driven assistants such as Siri, Amazon Echo, Google Now and Gesture based systems such as XBox Kinect, Leap Motion.

Shneiderman, a prominent figure within HCI stated:

> "*Well designed, effective computer systems generate positive feelings of success, competence, mastery, and clarity in the user community. When an interactive system is well-designed, the interface almost disappears, enabling users to concentrate on their work, exploration, or pleasure.*" - **Shneiderman** [202]

Achieving the qualities described by Shneiderman should be the goal of an interface designer. To facilitate this, HCI has a responsibility to provide the knowledge, tooling and approach necessary for enabling these forms of interaction. Technology also plays an important role in these interactions. HCI is a field that typically embraces new technology, especially those that facilitate new forms of interaction or their evaluation. One emerging area of technology of interest to HCI researchers is Brain-Computer Interfaces (BCI). BCIs are physical sensors that measure mental activity in participants. Mental activity in this context can refer broadly to the cognitive work we perform when completing a given task.

Recent commercialisation of the fundamental brain-sensing technologies that enables BCI research has begun to attract the attention of HCI researchers and may provide a new, objective form of developing natural interactions. Specifically, some HCI researchers have begun to explore the integration of BCI technology as a way of gleaning insight into how HCI occurs, how successful these interactions are and how we may use BCI technologies to augment interactions.

Through the body of work presented in this thesis, we investigate the potential of using commercialised brain-sensing technologies for natural interactions. We present the sub-research field of *Brain-based Human-Computer Interaction (BHCI)*, a direct measure of cognitive work that utilises these recent developments and availability of BCI devices to offer a novel new way of creating and studying engaging forms of natural interactions.

1.2 Brain based Human Computer Interaction

Direct control over a computer system, via a brain derived measure, has been the focus of Brain-Computer Interface (BCI) research over the last 30 years. Driven primarily to provide a communication channel to the physically disabled, BCI research has focused on understanding the meaning of the data obtained from these brain-sensing devices and is closely linked to the application of approaches and techniques derived from the field of Signal Processing. Much of the existing research into BCI has focussed on direct forms of control as an assistance technology, such as controlling a prosthetic limb.

Described as:

> "… an artificial intelligence system that can recognize a certain set of patterns in brain signals following five consecutive stages: signal acquisition, pre-processing or signal enhancement, feature extraction, classification, and the control interface." - **Khalid et al.** [109]

The description by Khalid et al. identifies the primary stages of BCI research: collection, processing, extraction, classification and resulting action/effect. The processing stages

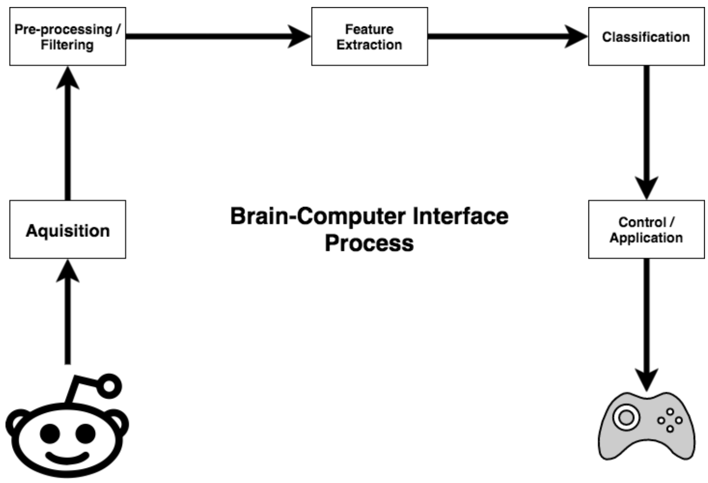


Figure 1.1: A typical BCI data processing pipeline.

described by Khalid et al. are visualised in Figure 1.1. Additionally, the description hints at the fact that the mapping between the data and interpretation is complex - requiring significant processing in order to classify the user's intentions accurately. BCI centred research will primarily focus on the first 4 stages described by Khalid et al.

BCI researchers are continuing to optimise the translation of brain data into the precise, actuated movement that the user intended. Conversely, it is the final stage - 'the control interface', that is of primary interest to HCI researchers. In the case of a prosthetic limb, the interaction is completing every day tasks, such as opening a door or navigating through a building. But this interaction present itself in nuanced forms, beyond physical movement. The same signal could, for example, provide the necessary state information to adapt a computer system to the user's cognitive state. Whereas significant work up to this point has focussed on understanding the signals we obtain from a BCI device, we believe that we are now at a sufficiently advanced level to begin the realistic exploration of HCI questions.

As HCI researchers, we have a unique opportunity in exploring the application of BCI beyond it's current context. There exists the opportunity to introduce the application of BCI approach to a wider audience, which in turn will lead to technological advancement and eventually lead to improvements for motor impaired individuals. In 2010, Tan et al. posited a similar argument, stating:

> "*As HCI researchers, we are in the unique position to think about the opportunities offered by widespread adoption of brain-computer interfaces. While it is a remarkable endeavor to use brain activity as a novel replacement for motor movement, we think that brain-computer interfaces used in this capacity will probably remain tethered to a fairly niche market.* - **Tan et al.** [219]

In the same chapter Tan et al. documents the potential applications of applying BCI within HCI. One application is the concept of a brain adapted user interface (UI), where a user's cognitive state information informs the presentation of data on an interface, adjusting the complexity or detail of the information relative to the user's current cognitive burden. Alternatively, the system could assist users who are cognitively burdened by introducing a computer based 'helper', which assumes responsibility for one aspect of the task. Knowing when to interrupt a user is another example of an classic HCI problem which can be addressed through the application of BHCI. Using brain-sensing technologies, a smart interactive system can appropriately interrupt the user according to their current mental workload, reducing the likelihood of the system interrupting the user at a moment of high concentration.

We note that both of the above applications share a common property of facilitating indirect or passive forms of natural interactions. We believe that BHCI can play a significant role in these indirect forms of interactions, and it is something we explore in the final study presented in this thesis. In the context of BHCI, this means that the system is somehow reacting to the user's current cognitive state, but that mapping or it's extent, may be unknown to the user. This unknown provides an interesting space for interaction designers in the development of novel and engaging forms of natural interactions. In this thesis we will explore the application of BHCI for developing a novel new form of natural interaction. We will evaluate the appropriate technology, techniques and approach for developing these forms of interactions and provide a robust set of contributions upon which others can develop their own future works. Also, we explore how this passive control relates to and facilitates natural interactions is perceived. Current literature documentes examples of the application of passive control in the context of NI, but little work has gone into formalising this interaction and exploring how it is perceived/experienced by users.

1.3 Research Questions

Having established the focus of applying BHCI as a tool in the development of natural interactions, we now present research questions that will provide a focus for the work presented in this thesis.

RQ1. What are the characteristics of a suitable BCI technology for supporting natural forms of interaction?

RQ2. How can BHCI be used to develop natural forms of indirect control?

RQ3. How are these natural forms of indirect control experienced by the users?

RQ4. What design considerations must we make when developing indirect natural interactions using BHCI?

To investigate and answer the questions presented above, we developed, executed and analysed 3 BHCI based user studies. In our initial two studies:

- Study 1 - “Think Aloud” (TAP)
- Study 2 - “Leap Motion” (LEAP)

we explore the application of two lightweight brain sensing technologies: fNIRS and EEG to answer **RQ1**. In our final study:

- Study 3 - #Scanners

we take the most suitable technology, as identified in answering **RQ1**, and develop an interactive cinematic experience to explore how users experience indirect brain-based interactions (**RQ2**, **RQ3**). Finally, we reflect upon this work as a whole to answer **RQ4**.

In our initial study, we sought to answer **RQ1**, for the brain-sensing technology - fNIRS. We applied fNIRS to a HCI user study investigating the use of verbal protocols (Think-Aloud Protocol). In doing so our primary aim was to investigate the suitability of fNIRS in applied HCI contexts using a common form of natural interaction - speech. The application of fNIRS to a speech based study develops our methodological knowledge of applying this brain-sensing technology to a HCI task, extending on the existing findings presented by Solovey et al. [210]. The results of this study indicate that whilst fNIRS is an excellent indicator of Mental Workload, it is a poor enabler of natural interactions. We learn that fNIRS significantly impacts upon the ecological validity of the study with participants indicating that the device is uncomfortable to wear for extended periods, and is not suited for developing natural interactions.

Continuing from the findings of our initial TAP study, we instead explore the application of EEG in a HCI setting with the aim of answering **RQ1** and understand the suitability of applying EEG to natural forms of interaction. In our LEAP study, participants were required to solve a 3D Jigsaw puzzle using a variety of different input modalities including: physical, mouse and gesture based inputs. The aim of this study *task* is to investigate whether there are significant workload differences in introducing this novel 3D input technique to solving the Jigsaw puzzle task, again under a HCI study setting. The goal of the *overall study* however, is to interrogate question **RQ1**. Specifically we are interested in evaluating whether the effects of a smaller form factor provided by the EEG is sufficient to offset the inevitable data quality issues that will arise from using a less stable source.

Our findings indicate that EEG is indeed suited to extended periods of application and participants reported that the device did not interfere with their task completion in any way. However, our results also demonstrated that there were significant issues in using EEG to estimate Mental Workload under these task conditions using existing data processing techniques. The study demonstrates the possibility of using the technology in the context of Natural forms of control, but also highlights the possible shortcomings of applying consumer grade EEG in broader BHCI applications - **RQ1**.

In our final exploratory study, #Scanners, we build upon the results of initial studies and narrow our focus on applying BHCI for developing natural forms of interaction (**RQ2**) and understand how participants react to this type of experience in the context of BHCI (**RQ3**). To answer **RQ2**, we developed an interactive, brain-based cinematic experience where the viewer's cognitive state information was used to inform the visual and audio presentation of the experience. We deployed this cinematic experience at a prestigious arts venue in the UK and conducted a 'Performance led research in the wild' methodology in order to reveal wider issues and principles in accordance with answering **RQ3**. In doing so, we identified a relationship between how participants discovered elements of control and how they chose to exert control - knowingly or otherwise. We present these findings in the form of a taxonomy which contributes towards answering **RQ4**.

We must also acknowledge that the exploration and contributions of these studies extend far beyond the narrative of exploring Natural Interactions. In conducting the work we have described above, we also provide an applied demonstration of BHCI, something that has limited detailing in the existing literature. We also examine fundamental principles of HCI and contribute new understanding of the craftwork of conducting HCI research. Therefore, in addition to exploring the contributions described above, we will also utilise the focus and results of these studies to demonstrate the value of applied BHCI to the field of HCI.

In our TAP study, we apply and extend upon the guidelines set out by Solovey et al. in using fNIRS in a HCI setting [210]. Specifically, we apply fNIRS for the purposes of evaluating the effect of a widely used verbal protocol (TAP). We exploring the **Methods** of applying BHCI, through the use of Solovey's framework and demonstrate that BHCI can be used to perform HCI centred evaluation, a common activity within HCI. In doing so, we also add to the body of evidence validating the framework set out by Solovey et al. and contribute to it further. Also, we show the application of BHCI for evaluation, something that has not been well documented in the current literature.

Through our LEAP study, we explore and evaluate different forms of **Input Control** - a common area of study in HCI. We again demonstrate the application of BHCI for evaluation in a HCI setting, specifically exploring different forms of input control and their effect upon the amount of mental work a user performs under each form of control. Also, we demonstrate progression and refinement in conducting BHCI studies, specifically driving towards more ecologically valid settings. This means that we want to keep a study setting as close to the 'real world' and remove (as much as possible) the effect of observing somebody in a lab setting. To achieve this (in conjunction with **RQ1**), we explore the application of less invasive BCI technologies and aim to reduce the restrictions imposed by using BCI technology in a study setting.

Finally, our #Scanners study explores **Novel Interaction**. HCI is responsible, in part, for a number of innovations that have lead to how we as human beings interact with machines on a daily basis. HCI will also be responsible for developing the future forms of these interactions - something we aim to explore through #Scanners. Through developing and understanding how our passive BHCI based interaction is experienced by participants, we hope to provide the foundation for a significant body of future work that will go into exploring this and alternative forms of novel, in-direct interaction.

1.4 Roadmap

Below is a brief overview of how the contents of this thesis will be present.

Chapter 2: Literature Review HCI is an inherently multi-disciplinary field with groundings in Human Factors, Psychology, Ergonomics and Computer Science. BHCI builds upon these foundations of knowledge and requires additional detailing of works relating to the architecture and working of the human brain. The Literature Review chapter presents a broad overview of the many fields that interact when conducting BHCI research. Additionally, we document existing work which falls into the categorisation of BHCI. The aim of this chapter is to support, contextualise and formalise the contribution of the work we present in this thesis against a broad backdrop of related disciplines.

Chapter 3: Evaluating fNIRS in BHCI via Verbal Protocols Using fNIRS as the brain monitoring technology we present the application of a BHCI approach in identifying the cognitive impact a verbal evaluation protocol. Through fNIRS alone, we were able to identify that the verbal protocol had no additional impact upon users workload, except when the verbalisations were not related to the task in hand. We also identify the limitations in applying fNIRS for extended periods of time with participants reporting a significant impact upon the ecological validity of the task.

Chapter 4: Evaluating EEG in BHCI via Gesture based Input Following the results of the study presented in Chapter 3, this work aims to compare the cognitive impact of introducing a 3D input device to a Jigsaw based task, and comparing the effect of this input modality against physical and mouse based versions of a similar task. The results indicate that the sensor itself is not as sensitive as fNIRS, but critically, the form factor and experience provided by the device enables extended forms of natural interactions without affecting the ecological validity of the task, since the device is less invasive than the fNIRS used previously.

Chapter 5: #Scanners: Natural, Brain based Interactions in Film Building from the results of the previous two studies, we conclude this work by developing and evaluating the impact of indirect, brain controlled cinematic experience, controlled through a consumer grade EEG device. With the aid of an Artist/Film producer we designed, developed and studied a new form of cinematic experience controlled via the mind of the viewer. Screening the film, titled #Scanners, at the prestigious arts venue FACT (Liverpool, UK), we conducted a 'performance-led research in the wild' methodology to evaluate the impact of this novel form of "Neurocinematics". In this chapter we detail the application of BHCI in this context, explore the creative space that is afforded through this form of passive control and discuss the implications for the future of cinematic experiences.

Chapter 6: Discussion Here we discuss the application of our framework we contribute from our analysis of #Scanners. We describe how the framework might be applied in the design and development of future interaction experiences that utilise passive control in some way. Finally, we discuss the components of research that will form the future work of the author.

Chapter 7: Conclusion In this chapter we clearly communicate the contributions we can derive from the research we have conducted in this thesis.

1.5 Publications

Below are a collection of peer-reviewed publications derived from the work presented in this thesis.

Matthew Pike and Eugene Ch'ng, Evaluating Virtual Reality Experience and Performance: A Brain based Approach, ACM SIGGRAPH International Conference on Virtual-Reality Continuum and its Applications in Industry, December 2016

Matthew Pike, Richard Ramchurn, Steve Benford, Max L Wilson, #Scanners: Exploring the Control of Adaptive Films using Brain-Computer Interaction, *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, April 2016

Matthew Pike, Richard Ramchurn, Max L Wilson, #Scanners: Integrating Physiology into Cinematic Experiences, *Proceedings of the 2015 ACM SIGCHI Conference on Creativity and Cognition*, June 2015, Pages 151-152

Matthew Pike, Richard Ramchurn, Max L Wilson, Two-Way Affect Loops in Multimedia Experiences, *Proceedings of the 2015 British HCI Conference*, July 2015, Pages 117-118

Matthew Pike, Horia A Maior, Martin Porcheron, Sarah C Sharples, Max L Wilson, Measuring the Effect of Think Aloud Protocols on Workload using fNIRS, *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, April 2014, Pages 3807-3816.

Horia A Maior, Matthew Pike, Sarah Sharples, Max L Wilson, Examining the Reliability of using fNIRS in Realistic HCI Settings for Spatial and Verbal Tasks, *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, April 2015, Pages 3039-3042

Horia A Maior, Matthew Pike, Max L Wilson, Sarah Sharples,Continuous Detection of Workload Overload: An fNIRS approach, *Proceedings of the international conference on Ergonomics &Human Factors 2014*, April 2014, Pages 450 - 455

Horia A Maior, Matthew Pike, Max L Wilson, Directly Evaluating the Cognitive Impact of Search User Interfaces: a Two-Pronged Approach with fNIRs,*Proceedings of the 3rd European Workshop on Human-Computer Interaction and Information Retrieval*, August 2013, Pages 43-46

Relevant publications, prior to PhD

Matthew Pike, Max L Wilson, Anna Divoli, Alyona Medelyan, CUES: Cognitive Usability Evaluation System, *Proceedings of the 2nd European Workshop on Human-Computer Interaction and Information Retrieval*, August 2012, Volume 12, Pages 1-4.

Chapter 2

Literature Review

2.1 Introduction

As an interdisciplinary research field, BHCI benefits from the foundations of knowledge and discoveries from a diverse range of fields. Below we present the primary contributors upon which BHCI is built and detail their role in informing the design, development and execution of BHCI research.

- **Human-Computer Interaction (HCI)** - Details our current understanding of how humans interact with interfaces and the techniques for eliciting insight into these interactions. HCI details a strong set of research protocols, research methodologies, study design and data analysis.
- **Brain-Computer Interfaces (BCI) and Neuroimaging** - The measure upon which BHCI distinguishes itself from traditional forms of HCI. The fields of BCI and Neuroimaging detail the operation and application of direct brain monitoring technologies. The incorporation of these measures in a HCI setting will require researchers to have an understanding of the fundamental workings and limitations of each imaging technology.
- **Psychology** - Psychology is a very mature research field which provides BHCI with a number of relevant theoretical models and examples of applied BCI for the study of the human brain. We build upon the field's detailing of memory, specifically Short-term memory, in understanding the effects a study task may have upon the participants memory resources. This knowledge can also be used to manipulate the amount of load a participant is exposed too, allowing BHCI researchers to investigate the effect upon the interaction with a UI as a participant reaches the limits of their mental capacity.

HCI and *BCI/Neuroimaging* play important roles in the early stages of study/research development and study design. As noted above, technology choice is a very important consideration in obtaining correct and valid results.

Psychology also plays an important part in this early stage, helping in the development of study tasks and ensuring that the cognitive functions of interests are targeted and manipulated according to the goals of the study. The mature body of knowledge provided by the field of *Psychology* will enable us to ensure that study tasks targets a particular mental attribute, or that a particular amount of work for a particular modality is elicited.

The field of *Psychology* helps us understand, frame and explain results from these types of studies.

We can see from above that there is a broad range of interaction between these fields. Presenting this related work is challenging, given the potential breadth of topics and the depth of discussion. In the following sub-sections, this significant body of existing work is presented in the following format:

1. **The brain, it's regions and it's relevance to BHCI (*Neuroscience*)** - A high level discussion of the anatomy of the brain, the regions that are of interest to us and why. It is important to note that this is a *very* high level overview of the brain's anatomy, and is intended to provide the reader with the necessary 'map-reading skills' for identifying and referencing regions of the brain.

2. **Theoretical Concepts of Mental Workload (*Psychology*)** - Given the complexity of the brain and our limited knowledge of how it functions, researchers in the fields of Psychology have presented theoretical concepts that attempt to characterise the resource interaction between task types. We use these concepts in our work to explain our study findings and to ground our results in existing, validated bodies of work.

3. **Measurement technology and their qualities (*BCI/Neuroimaging, Psychology, HCI*)** - We review a selection of brain monitoring technologies that are particularly relevant to BHCI. As well as covering the fundamentals of how each technology operates we also discuss the relative strengths and weaknesses of each technique and advise 'when, where and why' a particular technology should be favoured.

4. **BHCI in the Wild (*HCI*)** - As an emerging discipline, a small body of BHCI work already exists - we document these works in this section. We also document works that have an influential role in BHCI from other fields (such as BCI and Neuroimaging) that we discuss and relate to BHCI.

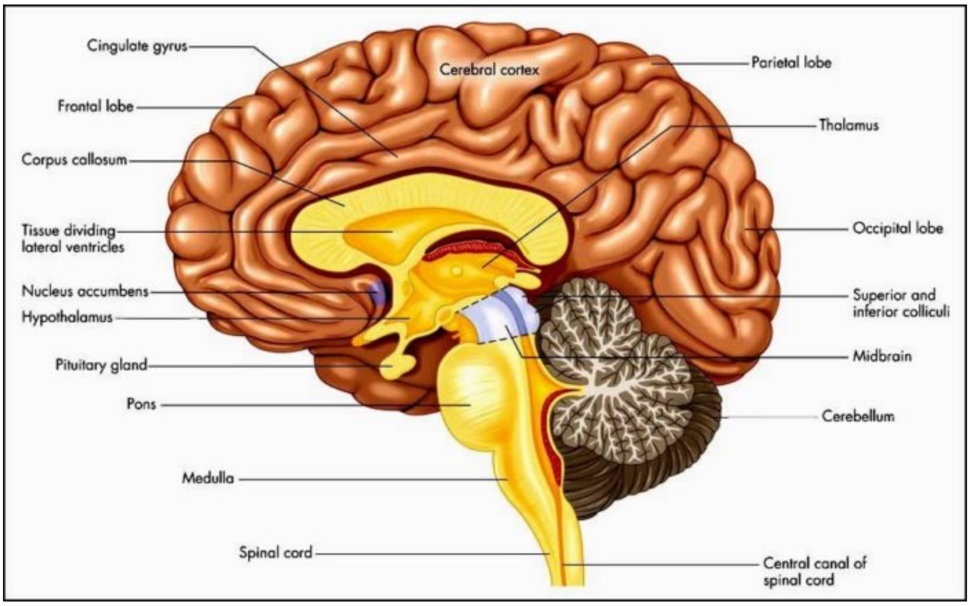


Figure 2.1: Annotated Cerebral Cortex. Image Source - McGill

2.2 The Brain, it's Regions and their Relevance to BHCI

Despite each of us possessing one, our understanding of the human brain remains limited. Our current knowledge details regions or clusters of the brain that we believe to be responsible for certain actions/reactions or forms of processing. Beginning to have a deep, neurological understanding of the brain and the interconnected workings of it's subsystems is far beyond the scope of this work. Rather, we are interested in understanding at a high level where certain processing centres of the brain reside, what we can capture from them and what this might tell us about how the brain is reacting to a particular study stimulus of interest to BHCI. In this section of related work, we introduce *Broadmann Area Map* to the reader, identify particular regions of interest (*Pre-Frontal* and *Motor* cortex) and justify our interest in these areas by relating their function to aspects of BHCI.

2.2.1 Mapping the Brain

2.2.1.1 Cerebral Cortex

As shown in Figure 2.1, the human brain consists of a number of regions, but for the purposes of the work presented in this thesis, we are going to focus our discussions on measurements and observations focussed on the Cerebral cortex - the outer layer of neural tissue closest to the skull. The cortex is a large folded structure given it a distinctive shape which conforms to the confines of the human skull that encapsulates it.

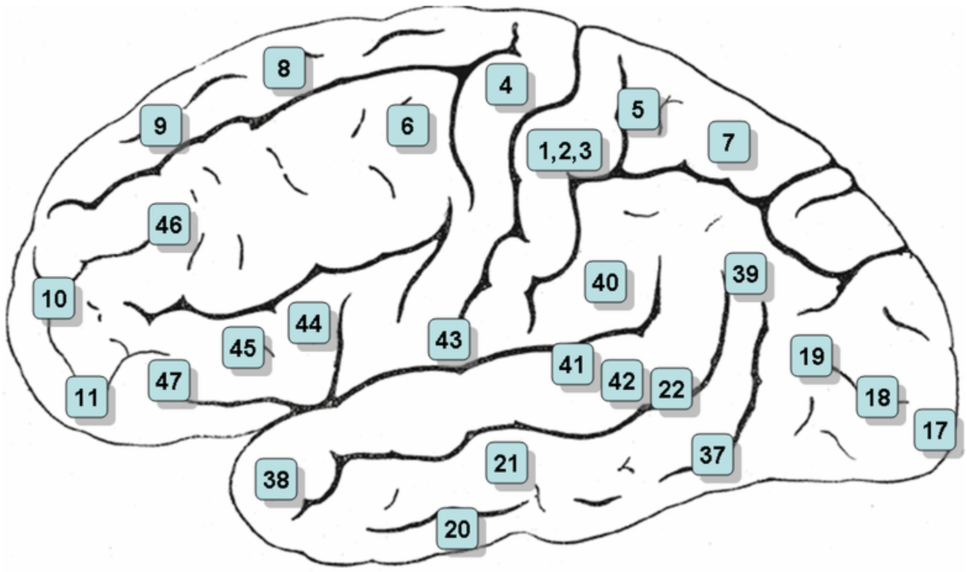


Figure 2.2: Labelled regions as specified by the Broadmann Area Map. Image Source - Wikipedia.

Due to the technical limitations of the brain-sensing technologies we use in this study (fNIRS and EEG), we are limited to observing activity originating from the cerebral cortex. A number of cognitive processes occur in this region however. The cortex is believed to play significant roles in the operation of human memory [170,62], attention [83], language [21,62], awareness and consciousness [119,183].

2.2.1.2 Brodmann Area Map

Through decades of research into the structure of the human brain, German anatomist Korbinian Brodmann identified 52 distinct regions of the Cerebral cortex, which he would, in 1909, plot in the *Broadmann Area Map* shown in Figure 2.2 [27]. Referred to as 'Broadmann Areas', these regions within the Cerebral cortex would prove invaluable in the communication of the responsibilities of the Cerebral cortex. Although the classification of these regions has developed over time, adapting to the discovery of new roles and behaviours of regions, the original name is still used, with articles commonly communicating a role of an region through the classification provided by Broadmann e.g. "Here we focus on Brodmann's area 44…" [50]. As such, we introduce the map to the reader at this early stage to aid in the understanding of future references.

2.2.2 The Pre-Frontal Cortex (PFC)

The Pre-Frontal Cortex (PFC) is the anterior (front, towards the eyes) part of the Cerebral cortex of the brain and is considered central to the function of Working Memory

(WM) (discussed below), dealing with executive and attention processes [107]. Brodmann areas 9, 10, 11, 46, and 47 are located within the PFC. As detailed by Miller and Cohen [149], our understanding of the PFC, as with much of our understanding of the human brain, is identified from those who are unfortunate enough to have sustained damage to that region.

> "On initial examination, PFC damage has remarkably little overt effect; patients can perceive and move, there is little impairment in their memory and they can appear remarkably normal in casual conversation. However, despite the superficial appearance of normality, PFC damage seems to devastate a person's life. They have difficulty in sustaining attention, in keeping 'on task', and seem to act on whims and impulses without regard to future consequences. This pattern of high-level deficits coupled with a sparing of lower-level basic functions has been called a 'dysexecutive syndrome' (Baddeley & Della Sala 1996) and 'goal neglect'(Duncan et al. 1996)." - **Miller and Cohen** [149]

From their analysis, Miller and Cohen theorise that cognitive control stems from the active maintenance of patterns of activity in the PFC that represent goals and means to solve them [149]. To provide an abstraction of this idea, one can visualise Miller and Cohen's proposed influence of the PFC as being akin to the switchboard operator of old who would connect incoming calls (sensory input) to their desired contact (cognitive actions). This description of role of the PFC's function aligns with that proposed by Baddeley's notion of an executive control (a component within Baddeley's model of Working Memory - discussed below) [14].

The PFC has been shown to play a role in the encoding and retrieval of **memory**. Fletcher et al. identified lateral localisation of encoding (right-PFC) and retrieval (left-PFC) through Neuroimaging studies [60]. Studies in patients with frontal lobe damage identified the role of the PFC in the storage and retrieval of recent (short-term) memories [132]. It has also been shown that the PFC is involved with learning and the encoding and representation of temporal information [60]. The PFC is a key area for studying workload (encoding and storage of information used in decision making) and control (decision making) both key interests of a BHCI researcher.

2.2.3 The Motor Cortex

The motor cortex is located towards the rear portion of the frontal lobe in the Cerebral cortex and is responsible for regulating the planning, control, and execution of voluntary

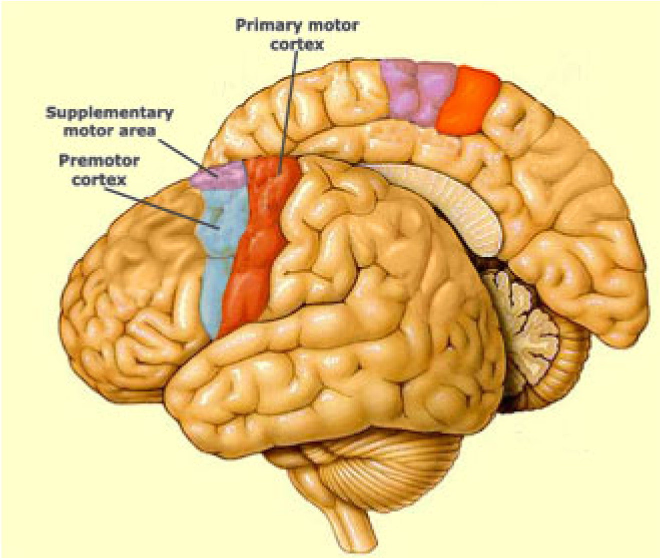


Figure 2.3: Annotated regions of the components that form the Motor Cortex. Image Source - McGill.

movements. The motor cortex is classically divided into two constituent regions: 1)The Pre-motor Cortex and 2)Primary Motor Cortex.

The primary motor cortex, located in Broadmann Area 4, is the main contributor to generating neural impulses that control the execution of an individual's movement via the nervous system [106,79]. Identified by Dr. Wilder Penfield during the mid-20th century, whilst treating a patient with epilepsy, Dr Penfield would later identify the vital area of the primary motor cortex and it's interaction with the pre-motor cortex [174]. The pre-motor cortex, located in Broadmann Area 6 (immediately forward of area 4), is responsible for aspects of preparatory control processes including: sensory guidance, spatial guidance, guidance of reaching (depth perception) as well as direct control of some movements [191,63,74].

The utilisation of the Motor cortex has been heavily explored by BCI researchers interested in providing direct control to users. These applications are especially focussed on providing disabled individuals with a new form of control over artificial limbs or assistive devices e.g. Robots [22,24,90]. In the context of HCI, the ability to control a hardware device or it's software analogue, provides a novel new form of interaction. This rich form of interaction is particularly of interest to researchers in the emerging field of virtual reality and gaming.

2.3 Mental Workload

Driven by the transition from physical to mental work observed by the labour force in the twenty-first century, researchers in the field of Psychology, Human Factors and

Ergonomics needed a theoretical construct for discussing the amount of ‘work’ on a mental level that was performed by an operator. The concept of Mental Workload (MWL) emerged in an attempt to model this mental work using a vocabulary that is consistent amongst researchers. Despite the attempt at consistency, the precise definition of what constitutes Mental Workload (MWL) is something that is not agreed upon in the current literature. Examples of definitions include:

> “The relative capacity to respond” - **Lysaght et al.** [133]
>
> “A construct that is used to describe the extent to which an operator has engaged the cognitive and physical resources required for a task performance” - **Backs et al.** [10]
>
> “Workload is a multidimensional and complex construct, that is affected by external task demands, environmental, organizational and psychological factors, and perceptive and cognitive abilities” - **Weinger et al.** [236]

The lack of a precise definition, does not hinder the work conducted in this area argues Sharples and Megaw, who state that a precise definition, is in fact less “profitable” (in terms of contribution), than providing a framework of the components that comprise the elements of MWL [241]. By analysing the various definitions of MWL available in the literature, we can de-construct the following components that appear consistent amongst definitions:

- There is an **Operator** under observation, who is performing work (a given task) but has **finite** mental resources
- These resources are being **utilised**, to **varying** degrees, in order to complete a given task.
- This task **elicits** some form of external physical/cognitive demands which interact with the demands of the individuals workload.
- The task has at least one **performance measure**, which is known by the Operator.

Using these components, Sharples and Megaw specified a framework, shown in Figure 2.4 for defining and evaluating MWL [241]. The framework details the source and interactions between the influencers of the operators MWL. Specifically, the framework details the interaction between the demands of the task and other external (e.g. An interruption to the task) and internal (e.g. Thoughts not related to the task at hand) influences, and their effect upon the operator’s task performance.

In the context of BHCI, the need for modelling MWL becomes clear when we wish to evaluate or measure the amount of work a participant (HCI’s terminology for Operator)

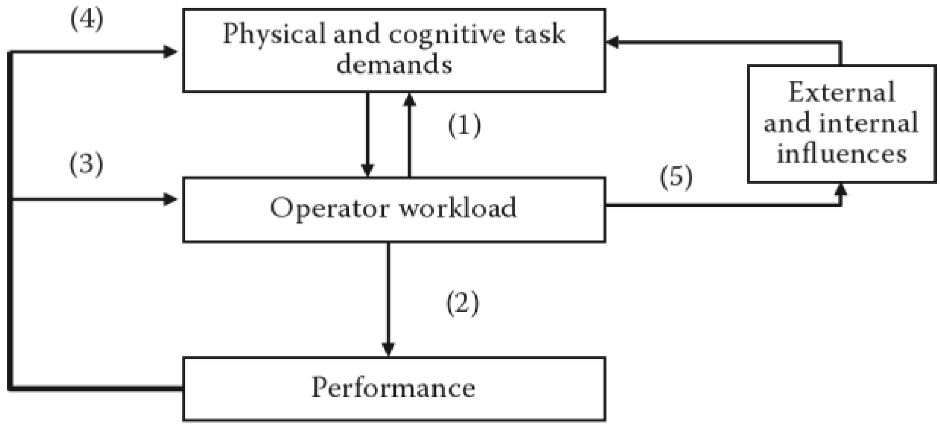


Figure 2.4: A framework for MWL definition and evaluation. Diagram from 'Evaluation of Human Work' - [241]

has performed during a task. The application of BHCI for evaluation in particular will utilise the theoretical concept of MWL in order to characterise the mental demands placed upon a participant. Through the works presented in the following sections, we see that the existing literature provides BHCI researchers with a framework for modelling MWL demands and how we can manipulate the type and amount of workload experienced by the participant. By varying the amount of workload a participant is exposed too, we are able to evaluate how a UI supports these variations in MWL, and identify potential 'weak-spots' in current designs. Additionally, this body of literature provides us with a framework for discussing and framing the results of a BHCI study, a theoretical grounding upon which the physiological measure can built on.

2.3.1 Model of Working Memory

Before we can discuss MWL, we must first establish an understanding of how memory, specifically Short-term Memory (STM), is described by the current literature. Specifically, we discuss Working Memory (WM), a component based model proposed by Baddeley and Hitch which refers to a specific system in the brain which:

> "provides temporary storage and manipulation of information..." - **Baddeley and Hitch** [11].

WM is a model of STM, a finite storage area used to manipulate information currently being processed by the brain [11,12,13,14]. The PFC has been shown to have significant involvement in WM, with a significant number of studies correlating measures of MWL to the PFC [107,148,44,52].

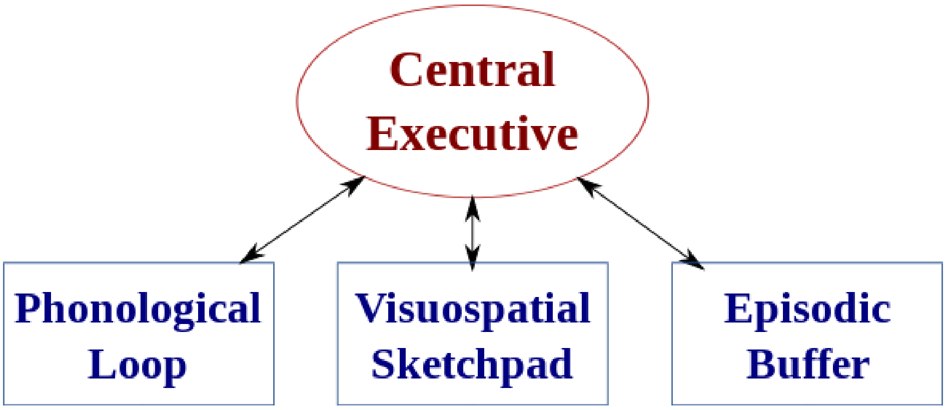


Figure 2.5: Baddeley's Working Memory Model.

According to Baddeley, information within WM can be represented in two forms:

1. **Verbal** - Information which is encoded verbally e.g. Numbers
2. **Spatial** - Spatial information, such as the geographical layout of your home-town or the interior of your childhood home.

Additionally, Baddeley specifies that WM consists of four main components, shown in Figure 2.5 and detailed below.

1. **Central Executive** - The overseer of information flow between the components of WM, specifically mediating information flow between the Visuo-Spatial Sketchpad and Phonological Loop, it's slave systems. The central executive also performs cognitive tasks such as mental arithmetic, problem solving and managing attention.
2. **Visuo-spatial Sketchpad** - The Visuo-Spatial Sketchpad is assumed to be responsible for the processing of spatial/visual information e.g. colours, shapes, maps, etc. The Visuo-Spatial Sketchpad is typically employed in navigation and imagery based tasks such as image rotation.
3. **Phonological Loop** - The component of WM responsible for spoken and written word. The Phonological Loop is specialised on learning and remembering information using repetition e.g. Remembering a telephone number through rote repetition. The phonological loop is assumed to be responsible for the manipulation of verbal/speech based information.
4. **Episodic Buffer** - dedicated to linking verbal and spatial information in chronological order. It is also assumed to have links to long-term memory.

The capacity of the WM model is generally considered to be limited, a phenomena identified by cognitive psychologist George Miller [150]. In his paper "The Magical Number Seven, Plus or Minus Two" Miller documented the capacity of the human memory system. Miller uncovered the limitations of the STM through the use of a 1D absolute-judgement task, in which a person must recall previously presented information. Through manipulating the number of items that participants were expected to memorise, Miller identified a decline in performance as the number reaches 5-6. This work would go on to become the most cited article in Psychology (25,000+ citations).

In addition to modelling STM, Baddeley describes the concept of Long-term memory (LTM), which identifies a different storage location to working memory [13]. LTM is (conceptually) unlimited in space and responsible for storing information that is no longer in WM. Transfer between STM and LTM is mediated by the Episodic Buffer and the Central Executive.

Using Baddeley's model, we have a high level abstraction of how memory works. This abstraction will provide BHCI researchers with a conceptual model upon which study tasks can be designed, taking into account the type of memory storage (spatial versus verbal) being utilised. A researcher interested in understanding the effects of a *verbal* protocol upon a participants WM, for example, can utilise the model to develop tasks to specifically target this modality. In tandem with the findings of Miller, the researcher can manipulate the number of verbal items (e.g. tones) a participant is expected to memorise whilst completing the verbal protocol, and observe the effect on recall performance. Equally, a researcher might want to identify the effect a verbal protocol might have on spatial memory, in a similar manner. Again, the model can provide insight into task design and development. The comparison and analysis between these two studies could also be contextualised and explained in relation to the model.

2.3.2 Modelling Mental Workload

Below we present a variety of models that attempt to detail how the human brain responds to changes in task types and demands. These models provide the BHCI researcher with a way of comprehending the demands imposed by study tasks and inform on the design and framing of a study and it's results. Our current understanding of the human brain does not allow us to simply quantify a value for MWL as a participant completes a task - this would be the ideal scenario, especially in the application of BHCI. Instead, we rely upon the knowledge provided by these models to explain why we observe variations in different measures of MWL, which we present below.

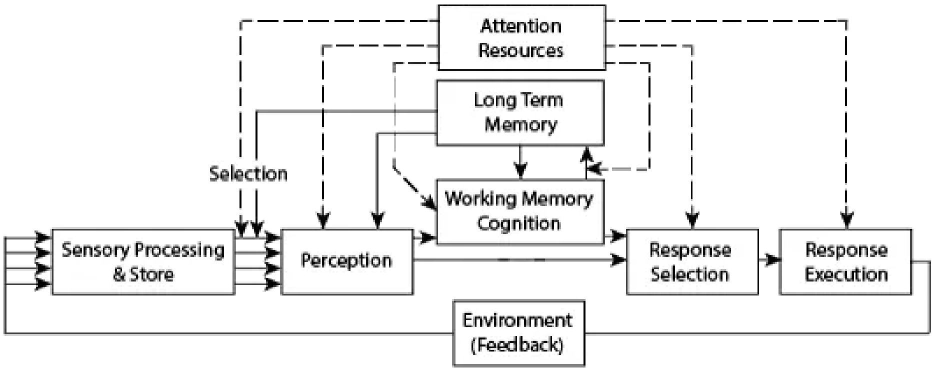


Figure 2.6: Human Information Processing Model Lifecycle, from Wickens - [239]

2.3.2.1 Human Information Processing

Information Processing models have been a significant source of our knowledge and understanding of MWL. Wickens was a significant influence in the development of these models, and was able to capture how basic psychological processes interact with task demands. Wickens' "Human Information Processing Model", shown in Figure 2.6, identifies the interaction between Sensory Perception, WM and Response. Wickens describes that necessary resources are limited and aims to illustrate how elements of the human information processing system such as attention, perception, memory, decision making and response selection interconnect.

Wickens describes three different 'stages' at which information is transformed: 1)a perception stage, a processing or cognition stage, and a response stage. The first stage involves perceiving information that is gathered by our senses and provide meaning and interpretation of what is being sensed; 2)The second stage represents the step where we manipulate and "think about" the perceived information. This part of the information processing system takes place in WM and consists of a wide variety of mental activities.

Wickens also proposed the Multiple Resource Model [237], illustrated in Figure 2.7. Through the MRM, Wickens describes the aspects of cognition and the multiple resource theory in four dimensions:

- The **STAGES** dimension refers to the three main stages of information processing system as described above.
- The **MODALITIES** dimension indicating that auditory and visual perception have different sources.
- The **CODES** dimension refers to the types of memory encodings which can be spatial or verbal.

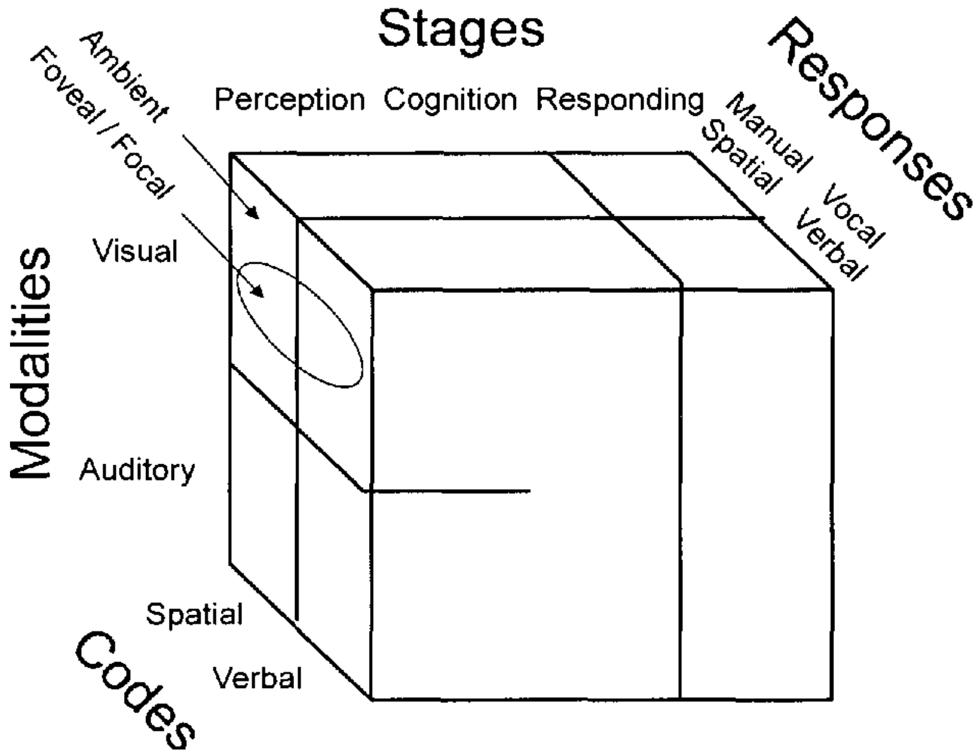


Figure 2.7: The 4-D multiple resource model, by Wickens.

- The **VISUAL PROCESSING** dimension refers to a nested dimension within visual resources distinguishing between focal vision (reading text) and ambient vision (orientation and movement).

One of the key roles of the MRM is to demonstrate the hypothesised independence of modalities and use this to design tasks. Wickens' model provides a high level view of the available resource types and can be used as a basis for explaining results obtained from a study. Wickens' model(s) have been used extensively in the field of Human Factors, across a diverse range of tasks and study protocols, with replicated/repeated findings. We will apply Wickens' model in a similar manner for BHCI related work. Wickens' MRM differs from Baddeley's model in the respect that it does not associate itself with a particular neurological region, rather it attempts to characterise the entirety of human resources.

2.3.2.2 Limited Resource Model (LRM)

MWL can be described as the amount of resources an operator uses when performing a specific task. These resources are limited; therefore, a problem arises when a task requires the operator to use more resources than are maximally available. This state is known as a human operator overload, and normally results in a significant drop in performance.

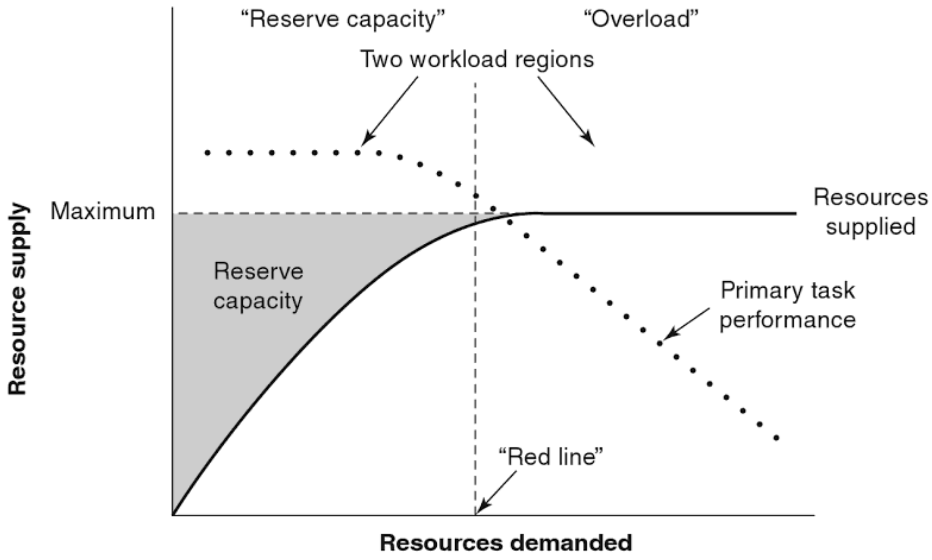


Figure 2.8: The relationship between task performance and resource supply, from Megaw - [239]

Therefore, it is important to consider the operators optimum level of workload (not overloaded) throughout the task.

One other model that is of interest is the Limited Resource Model (LRM) which describes the relationship between the demands of a task, the resources allocated to the task and the impact on performance [146]. The graph shown in Figure 2.8 is used to represent the LRM. The X axes represent the resources demanded by the primary task and as we move to the right of the axis, the resources demanded by the primary task increase. The axis on the left indicate the resources being used, but also the maximum available resources point (if we think of working memory that is limited in capacity). The right axis indicate the performance of the primary task (the dotted line on the graph). The key element of this model is the concept of a limited set of resources which, if exceeded, has a negative impact on performance. However, it does not distinguish between resource modality, therefore we propose to use both the limited and multiple resources models to inform our work.

BHCI researchers can utilise the LRM in accordance with the continuous measure obtained from a direct brain-monitoring device. The combination of this physical measure and theoretical model, will allow researchers to track how resources have been allocated in accordance with the LRM. This combination will allow BHCI researchers to detect situations in which the participant is 'overloaded', the state where the resources allocated are near the maximum available resources and the spare capacity is minimal.

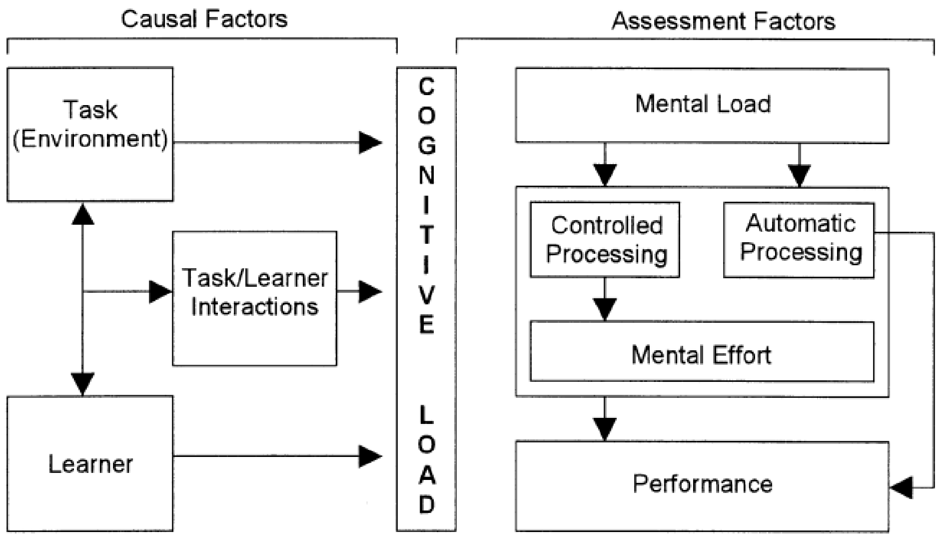


Figure 2.9: Factors determining the level of cognitive load. Diagram from Kirschner - [111]

2.3.2.3 Cognitive Load Theory (CLT)

Cognitive Load Theory (CLT) is a model that attempts to detail how individuals acquire new concepts or knowledge. Developed by John Sweller in 1988, CLT utilises the [Working Memory] model described by Baddeley, modelling the interaction between STM and LTM when forming new schemata (learning). In order to learn, a schema must form permanently in LTM in a transitional process which is facilitated by STM. CLT models this relationship by exploring the limitations of WM, specifically through the modelling of the limited storage capacity presented by the STM, most notably 7$\pm$2 [150]. CLT is therefore concerned with the limitations presented in STM and how this effects an individuals ability form new schemata during learning.

Depicted in Figure 2.9, **Causal** and **Assessment** factors play a role in the Cognitive Load of an individual [168]. Causal factors will include the existing abilities of the participant in relation to the task. Task complexity and the Environment in which the study is conducted are also factors. Assessment factors are those imposed by conducting the study and it's task. These include mental load, effort and performance as measurements of Cognitive load.

CLT differentiates between the type of Cognitive Load according to three forms: **Intrinsic**, **Extraneous** and **Germane**.

Intrinsic load specifies the inherent level of difficulty imposed by the task or material presented to the participant during a study. Sweller states that each type of task has it's own inherent difficulty e.g. Solving a jigsaw of 5 pieces versus a Jigsaw consisting of 50 pieces [37].

Extraneous and *Germane* load are linked to the presentation of the materials during a study. *Germane* cognitive load is imposed upon the *participant* as they construct new schemata (through learning the task) in LTM [218,37]. The construction of sufficiently complex and rich schemata will require more effort when the task has complex requirements (High Intrinsic load). *Extraneous* load is that affected by the presentation of the task and instructional material. In the context of HCI, we are are especially interested in approaches to designing interfaces that reduce the extraneous load of a user, allowing them to focus on the inherent demands imposed by the task (Germane and Intrinsic). As such, the aim of any interaction designer should be to elicit minimal levels of Extraneous cognitive load [217,67].

2.3.3 Measures of Mental Workload

We present a selection of the most commonly utilised measures of MWL below and detail their relative strengths and weaknesses - especially how these relate to BHCI.

2.3.3.1 Task Performance

Task performance is a common measure that is captured and analysed by research scientists when running a study. The literature indicates that certain types of task performance measures can also be indicators of a participants MWL. The techniques employed in this measure are presented in two forms: Primary and Secondary task performance.

Primary task measures are measures calculated from the completion of the main study task and can include, for example, errors made, speed and reaction times. Whilst the collection of these measures are inherent in any study employing a performance based task, the relation of this measure to MWL is direct. There are a number of issues with using primary task performance as an indicator of MWL, as Sharples and Megaw detail [241]:

1. Poor performance could be indicative of task demands being too high, but equally, task performance does not reflect task demands and we cannot know the amount of spare capacity available to the operator using this measure.

2. There is an observation effect in conducting the study. The operators will allocate additional resources to completing aspects of the task that are captured by the performance measure, whilst possibly ignoring other aspects of their duty.

3. Measures are prone to interpretation, especially when related to MWL.

A more commonly employed form of a performance based measure of MWL is the application of Secondary-task measure. In using this technique, a researcher will ask a participant to complete a task in addition to the primary task. In doing so, a researcher will be able to quantify the spare capacity in MWL that a participant has as they complete the primary study tasks. Typically, a dual-task methodology will be employed in one of two ways [30]:

1. Researchers observe the reduction in the primary task performance in a condition containing the secondary task - compared against the single task condition.
2. Observe the reduction in Secondary tasks performance in the dual-task condition - again, compared against the single task condition.

The dual study methodology has been successfully applied across thousands of studies. Ryu and Myung utilised a dual task methodology in combination with three physiological measure (Electroencephalogram (EEG), Electrooculogram (EOG), and Electrocardiogram (ECG)), whose signal were combined using weighted coefficients [193]. The study tasks utilised visual (object tracking) and verbal (mental arithmetic) memory, and the combined signal systematically increased with task difficulty - indicating that the methodology successfully manipulates MWL.

2.3.3.2 Subjective

Subjective measures are a popular tool amongst researchers in the field of HF and HCI, thanks to the measure possessing a number of positive qualities, including:

- **Easy to Implement and Established Measures** - Pre-existing tools and techniques have been developed, peer-reviewed and their validity has been established.
- **Low-Cost** - Typically paper or computer based measures that require few resources or investment.
- **Non-Intrusive** - Subjective measures require no apparatus to be worn by the participant, and can usually be conducted post-task, meaning there is minimal task interference.
- **Established Statistical Analysis** - A well established processing for analysing and reporting statistical data derived from these measures are available in the existing literature.

- **Effective** - A number of subjective measures have been shown to be sensitive in their reporting of MWL.

But as noted by Sharples et al., the measures do take into account participant feedback at face value and that true validity of these measures remains relatively elusive [241].

Subjective Workload Assessment Technique (SWAT) [189], NASA Task Load Index (NASA-TLX) [80] and Workload Profile (WP) [224] are examples of popular subjective measures that are heavily utilised in HF and HCI user studies. Typically, the application of these measures will require participants to complete a questionnaire related the task they had previously undertaking (or are in the middle of completing). SWAT for example, requires participants to rate workload in three levels (*low, medium* & *high*) across three dimensions (*time, mental effort* and *psychological stress*).

NASA-TLX, is perhaps the most widely used subjective measure of an individuals' perceived workload. Originally developed for use in the field of Aviation, the measure has since been adopted across a number of disciplines including HF and HCI.

NASA-TLX specifies six dimensions that are used to assess MWL.

1. **Mental Demand** - How much mental/perceptual activity was required.
    - *Examples*: Memorisation, Arithmetic, Searching
2. **Physical Demand** - How much physical activity was required.
    - *Examples*: Pushing, Pulling, Controlling, Clicking
3. **Temporal Demand** - How much time pressure was felt as a result of the pace at which the task elements occurred.
    - *Examples*: Tasks with visible/known time limits (countdown).
4. **Performance** - How successful did the participant feel they were.
    - *Examples*: Tasks with an end-goal, how close were they to obtaining it.
5. **Effort** - How hard did the participant work (mentally and physically) to complete the task.
    - *Examples*: For physical tasks, the participant might perceive perspiration as being an indicator of their effort.
6. **Frustration** - How frustrated did the participant perceive themselves becoming.
    - *Examples*: Insecurity, discouragement, irritation, stress and annoyance (presence/absence of these)

Figure 2.10 shows a paper version of the questionnaire, although, typically the presentation and completion of this form is done via a computerised method.

We see from Figure 2.10 that twenty step bipolar scales are used to obtain ratings for each of the six dimensions. Each dimension is scored from 0 to 100 on each scale. A global score is obtained, by weighting and combining the six individual scale ratings. The weighting process requires the operator to perform a pairwise comparison between all pairs of the six dimensions, with the selection in a pair adding weight to the chosen dimension. Having obtained a weighting for each dimension, the score for each is multiplied by its respective weight to obtain a workload score between 0 and 100.

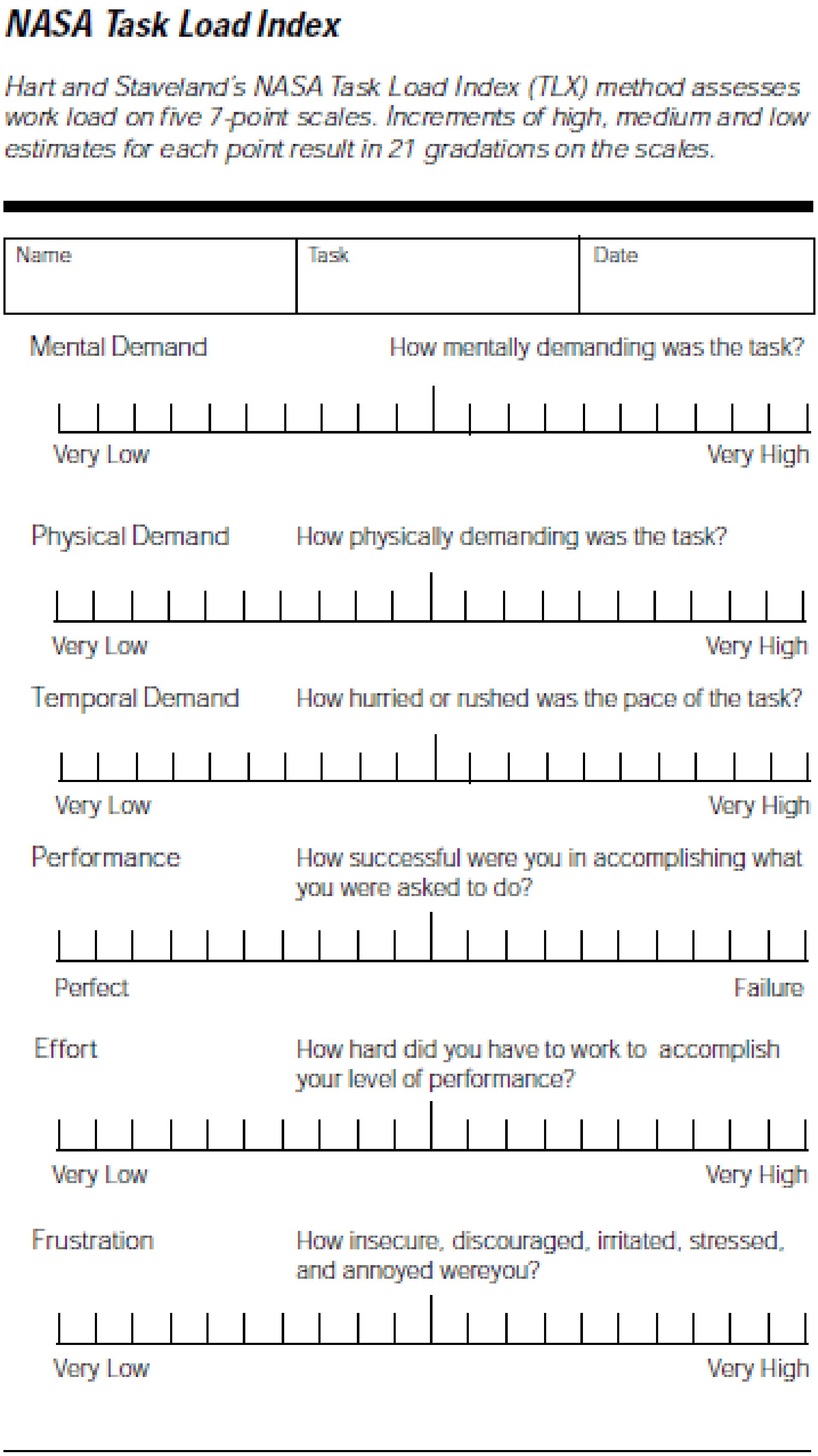
The NASA-TLX questionnaire is the result of significant program of research performed at NASA by Hart and Staveland [80]. The value, sensitivity and reliability of the measure has been proven time-and-time again, for a diverse set of tasks across a number of different research disciplines. Originally developed for use in aviation [17,200,248], but has seen adoption across a broad landscape of research fields, including: Medical [249,255,91], HCI [162,192,2,73], Automotive [212,213] and many others.

2.3.3.3 Objective Measures of Mental Workload

MWL is a theoretical concept of work that occurs in an operator under task condition. Measuring MWL therefore is a distinguished by measuring change in how hard a particular region of the brain is working, and relating the change in this work onto our theoretical models of MWL (Wickens and Baddeley). Measuring these changes in an operators workload can be achieved by observing their psychophysiology, in a manner that is typically classified as being either *direct* or *indirect* forms of observation.

Indirect measures typically refer to techniques that do not directly measure changes in the brain itself. Rather, these measures track changes in related, often easier-to-measure, psychophysiological indicators that are then attributed to changes in MWL. Examples of these measures include: Endogenous Eye Blinks [215], Pupil Size [105,16,185,240], Electrodermal Activity [104,57]. The advantage of these measures is their relative ease and simplicity in application, but remain prone to additional influences not directly attributable to the brain.

Direct measures, conversely, derive their measurements directly from changes in the brain itself. This allows us to measure changes in the brain, without depending on inference or relationships from other measures which have been associated with MWL. A number of techniques exists for estimating or deriving levels of MWL from indirect measures of



NASA Task Load Index

Hart and Staveland's NASA Task Load Index (TLX) method assesses work load on five 7-point scales. Increments of high, medium and low estimates for each point result in 21 gradations on the scales.

Name	Task	Date

Mental Demand — How mentally demanding was the task?

Very Low — Very High

Physical Demand — How physically demanding was the task?

Very Low — Very High

Temporal Demand — How hurried or rushed was the pace of the task?

Very Low — Very High

Performance — How successful were you in accomplishing what you were asked to do?

Perfect — Failure

Effort — How hard did you have to work to accomplish your level of performance?

Very Low — Very High

Frustration — How insecure, discouraged, irritated, stressed, and annoyed wereyou?

Very Low — Very High

Figure 2.10: A paper version of the NASA-TLX questionnaire.

the brain, as presented above, in this work however, we specifically concentrate on direct, brain based measures.

Below, we review the literature for the three brain monitoring technologies below, detailing each technology's suitability for application in a BHCI setting. We additionally provide Table 2.1 as a summary and reference of a high level overview of the technology choices available and their properties in relation to BHCI.

2.3.3.3.1 Functional Magnetic Resonance Imaging (FMRI) Functional Magnetic Resonance Imaging (fMRI) is a direct brain monitoring technique that measures relative changes in blood flow/delivery to the neuronal tissue within the brain [142]. Since neurons do not have their own internal energy stores, they are dependant on the rapid supply of energy, in the form of oxygenated blood, in order to continue 'firing' (working). As the demands upon the brain increase in line with task demands (e.g. A complex mathematical task is presented to the participant), the relative amounts of oxygenated and de-oxygenated blood and it's supply varies accordingly [225]. It is this variation in oxy(HbO2)/deoxy(HbR)-genated blood that fMRI uses to calculate the location/source of activity within the brain [130]. This calculation is performed using an image contrasting technique called *Blood-oxygen-level dependent contrast* imaging, or **BOLD**. Ogawa et al. proposed that such an approach would work *in-vivo* subjects [165], a fact later confirmed by Kwong et al. in 1992 [199].

fMRI has become the primary form of measurements for thousands of cognitive, motor and affective function based studies utilising this imaging technique [184,31]. Favoured for it's extremely high spatial resolution, fMRI is considered as the 'gold standard' imaging technique for identify interaction of regions of the brain with particular cognitive function.

For application in BHCI however, fMRI presents a number of practical limitations. First, as we can see from Figure 2.11, an fMRI is a large machine requiring significant space and financial resources. The purchase and operational cost of an fMRI is significant, with a price in the millions of dollars to purchase and an operational cost of $500 an hour, making the technology prohibitively expensive to most HCI based researchers [241]. The use of fMRI also introduces a significant impact upon the ecological validity of the environment in which a participant must conduct a study. Being confined to a large, noisy and restrictive machine is far from a typical environment in which a user would normally operate. A number of proposals to reduce this impact upon ecological validity have been suggested by researchers interested in using this highly accurate imaging technique. Campbell et al. for example, proposed the application of Virtual Reality as a way of reducing the impact upon ecological validity, through the simulation of a user's natural environment, but this remains a work in progress [32]. fMRI is very sensitive to

Table 2.1: Comparison of direct brain-monitoring technologies that can be used in BHCI based research.

Method	Technology	Measures	Portability	Cost	Invasiveness	Spatial Resolution	Temporal Resolution
fMRI	Magnetic	Oxy/De-Oxy Hemoglobin	Low	Very High	Significant	High	Low
fNIRS	Optical	Oxy/De-Oxy Hemoglobin	Moderate	High	Moderate	Moderate	Low
Medical EEG	Electrical	Post Synaptic electrical activity	Moderate	High	Moderate	Low	High
Consumer EEG	Electrical	Post Synaptic electrical activity	High	Low	Low	Low	High

Figure 2.11: A participant laying ready to be scanned in an fMRI machine. Image Credit - UCL.

motion artefacts, and participants are usually required to be strapped into the machine, restricting head and body movement significantly in order to preserve the quality of the measure [72].

Finally, fMRI is fundamentally dependant on a large electromagnet which prohibits the existence of any metal based objects in the room - a significant restriction for HCI studies which are dependant on metal/electronic hardware e.g. phones, laptops and screens - none of which can be present whilst the machine is operational. This restriction has proved to be significant barrier for a wide set of research studies and has required researchers to develop novel new forms of input to circumvent this restriction. Examples workarounds include a plastic, full-size (typing) keyboards [99] and an entirely plastic (musical) keyboard for researchers investigating the neuroscience of improvisation and creativity [129]. Generally, fMRI is considered to be impractical for use in general BHCI based applications and should only be considered as a primary measure when a study is *primarily* interested in understand the regions of the brain that are activated, rather than the general 'amount' of affect that we tend to measure in BHCI [219].

2.3.3.3.2 Functional Near-Infrared Spectroscopy (FNIRS) Functional Near Infrared Spectroscopy (fNIRS) is an optical based, non-invasive imaging technique for measuring Cerebral haemodynamics in the human brain (similar to fMRI, above, but with a lower spatial resolution). fNIRS measures the haemodynamic response - the delivery of blood to active neuronal tissues and it is designed to be placed directly upon a participants scalp, typically targeting the pre-frontal cortex (PFC). It has the properties of

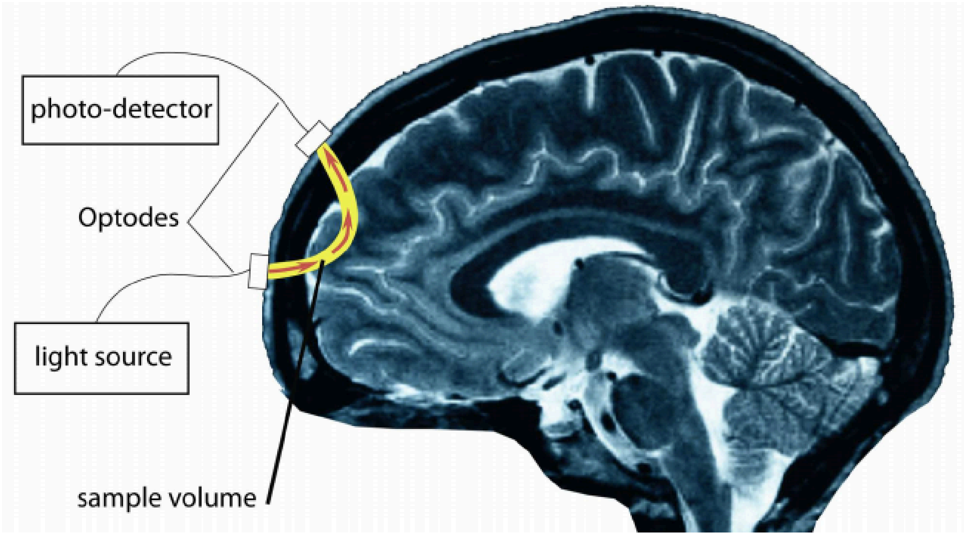


Figure 2.12: The banana-shaped path taken by Infrared Light (700-900nm). Image Credit - [127]

being non-invasive, portable, inexpensive and suitable for periods of extended monitoring relative to other Neuroimaging techniques.

An fNIRS system will typically utilise infra-red light and a phenomena known as the "optical window" which allows light at a particular frequency (700-900nm) to pass through skin, blood and bone without 'scattering' [102,206,85]. The path of the transmitted Infrared light at this frequency is however affected by the nervous tissue and chromophores that reside within the Cerebral cortex of the brain, causing the light to follow a 'banana'-shaped path (this is the actual technical term - [127]) back to the surface of the scalp as depicted in Figure 2.12.

The absorption of light by chromophores (an atom whose presence is responsible for the colour of a compound) within the nervous tissue of the brain affects the amount of light being emitted back to the scalp's surface, for detection by a photo-detector (as demonstrated in Figure 2.12). This property has been demonstrated to allow researchers to measure haemodynamic responses [233,36].

The properties of the optical window allow researchers to apply fNIRS in a non-invasive (does not require implants), low risk (harmless Infrared light) method of brain monitoring. fNIRS devices are typically portable, reasonably affordable (significantly less than fMRI, but generally more than EEG), easy to apply in a non-invasive manner with little training.

Recent research has shown that because blood-flow in the brain is less affected by body movement, fNIRS may be a more appropriate brain sensing technology for evaluation relative to other brain monitoring technologies [121,173,87] - a significant factor in technology choice for application in BHCI. Equally however, researchers should consider the

Figure 2.13: The Biopac fNIRS device being worn by a participant.

fact it takes several seconds for blood to flow to the brain [234] , meaning fNIRS has been largely discounted for use in real-time interaction systems.

A detailed review of commercially available fNIRS instrumentation was conducted by Scholkmann et al . [196]. Additionally, Scholkmann et al. provide a detailed overview of the methodological aspects, different methods of calculating oxy/de-oxy ratios and approaches towards data analysis. The review should be considered essential reading for BHCI practitioners using fNIRS.

A limited number of HCI researchers have successfully applied fNIRS in a BHCI style study (we detail this work below in BHCI in the Wild) [173,210,5,71]. fNIRS has been deployed in a number of studies and has caused minimal Interference, with many reporting ecological validity whilst using fNIRS [178,210,5]. Solovey et al. demonstrated that fNIRS was able to distinguish between common human behaviours (typing, mouse movement, head and facial movement) and a verbal memory task and provides the researchers with a set of guidelines for applying fNIRS in a HCI setting [210]. We replicated and extended the work of Solovey et al. [210] , further contributing to the body of work establishing the reliability and usability of fNIRS in a HCI setting [137]. Peck et al. provide a broad overview of the application of fNIRS in 'the real world' and discusses the impact upon ecological validity [172].

2.3.3.3.3 Electroencephalogram (EEG) Electroencephalograph (EEG) is an electrical based brain-monitoring technology that measures electrical activity on the scalp

Bandwidth	Rhythm	Function
0.0–0.5 Hz	Sub-delta	Normal: artifacts Pathological: interictal and ictal rhythm with focal seizures
0.5–3.5 Hz	Delta	Normal: sleep, hyperventilation, posterior slow waves of youth, elderly Pathological: encephalopathy, structural lesions involving white matter
>3.5–<8 Hz	Theta	Normal: drowsiness, children, elderly Pathological: encephalopathy, lesion (lower correlation than delta)
8–13 Hz	Alpha	Normal: posterior dominant rhythm, mu rhythm, "third" rhythm Pathological: ictal rhythm with seizures, alpha coma
13–30 Hz	Beta	Normal: effect of medication, drowsiness Pathological: breach rhythm, drug overdose (continuous and high amplitude), ictal rhythm associated with seizures
30–80 Hz	Gamma	Normal: voluntary motor movement; learning and memory. Pathological: seizures
80–250 Hz	Ripples	Normal: cognitive processing and episodic memory consolidation Pathological: interictal and ictal seizure frequency and possibly epileptogenesis
250–500 Hz	Fast Ripples	Normal: ? Pathological: seizures
500–1000 Hz	Very Fast Ripples	Normal: acquisition of sensory information Pathological: seizures

Figure 2.14: The spectrum of EEG waveforms. Table is provided by Tatum [220].

of a participant [161,219]. The measure has temporally high resolution, as the recording of electrical activity is almost instantaneous from the observed synapses firing, however, EEG has relatively poor spatial resolution as the observed measure of electrical activity is the summation of many post-synaptic firings, derived from a large cluster of neurons [181]. EEG sensors can therefore be positioned in general regions of interest (e.g. frontal, parietal lobes) but is not suited for precise mapping of neuronal-activity analysis.

Since the late 1920's, EEG has been used in both clinical and experimental settings to study the electrical response to stimuli in participants. EEG is typically recorded on the scalp of the subject, allowing for a relatively non-intrusive application, although implanted electrode applications of EEG exists and are typically utilised in the monitoring and diagnosis of individuals suffering with Epilepsy [164,96].

EEG recordings are composed of a collection of waveforms, classified by different frequencies with a range of between 1 and 40Hz and with a voltage range of 10 to 200 microvolts [41]. From this range of frequencies, power analysis can be performed to obtain a frequency spectrum for each recording, to identify a set of constituent frequency bands which are classified according to the Figure 2.14.

Recent work has discovered interactions between these waveforms and MWL. **Alpha (8-12hz)** has a number of strong inverse relationship with MWL, that is, as levels of MWL increase, Alpha levels will decrease. Pfurtscheller et al. identified a relationship between Alpha waves and idling in the cortical region of the brain [175]. Similarly Laufs et al. identified evidence to support the idea of a 'default mode of brain function' as proposed by Raichle et al. through observations upon Alpha based activity [120,187]. More direct observation of MWL and Alpha waves have identified correlations between low levels of MWL and high levels of Alpha [28,58].

There is also evidence in the existing literature to suggest that **Theta(4-8hz)** waves have a relationship to MWL. Theta has been shown to increase as task requirements increase [152], indicating a direct relationship between Theta levels and MWL. Esposito simultaneously recorded EEG and fMRI signals during an n-back task and identified sustained changes in the BOLD signal (discussed above - fMRI) and the synchronization of EEG theta waves to task demands, such that, as task requirements increased, so did theta levels [54]. Jensen and Tesche identified a parametric increase in the Theta band (specifically 7.5-8hz) for a visual n-back test based on number memorisation [101]. In his reviews of the literature on EEG and MWL, Klimesch has summarised the evidence to support the relationship between Alpha, Theta and MWL [112].

Smith and Gevins were able to confirm the relationship in theta and alpha band activities in a flight simulator task[208]. In their study, 3 versions (easy, medium and hard) of a piloting task were administered to 16 participants via a desktop PC. Theta and alpha bands were found to vary systematically with task difficulty. High MWL conditions had increased theta band activity and decreased alpha band activity relative to low MWL tasks.

The recent commercialisation, miniaturisation and portability of EEG technology has led to a number of researchers investigating the application of EEG in their work [219;]. One consideration when using EEG is it's susceptibility to motion derived artefacts [18]. *Muscles movements* are the most common form of artefact sources, with jaw muscles (e.g. jaw clenching) being a common sources of interference. These artefacts are typically shorter in duration and their irregular frequency allow us [145,29] to isolate and identify these artefacts, specifically using a independent component analysis (ICA) based approach. *Eye movement* is another example of a motion artefact, with eye blinks being especially easy to classify [221,35] and remove it's effect from the EEG signal [103,88]. The eyeball itself acts as a source of electrical potential, with the cornea being positively charged and the retina being negative. When rotated (eye movement), a large-amplitude AC field is generated which EEG sensors, especially those close to the eye are able to detect. Not all artefacts can be identified or mitigated in this manner however, and the signal provided generally by EEG is considered to be noisy, especially in uncontrolled settings.

2.4 BHCI in the Wild

The push for interdisciplinary research, affordability and an increased awareness of the quality and richness of the data obtained from brain-monitoring devices, has attracted many 'non-traditional' researchers to using these technologies as a part of a their research. As such, there are many examples of brain-monitoring devices being utilised in new non-clinical settings. Although these examples might not be direct analogues of the BHCI approach we are promoting through the work presented in this thesis, they do nevertheless provide a rich set of work upon which BHCI can build and draw upon. We will present these works below, which have been categorised according to their application area.

That being said, we also note examples of direct applications of BHCI in the existing literature. Perhaps the most prominent figure whose work falls into the classification of BHCI research is Robert Jacobs[1] research group at Tufts University, Boston USA. Jacobs, a prominent figure in HCI for the past three decades having championed the application of eye tracking in HCI research has now taken to a similar stance on the application of the brain monitoring technology - fNIRS. Through his group of researchers, Jacobs has set out to demonstrate the value of utilising fNIRS in the context of HCI. In the following section we mark works from Jacobs group with a dagger (†) indicator, to give the reader a sense of the diverse research being produced by Jacobs and his research group at Tufts.

2.4.1 Control

2.4.1.1 Communication and Direct Control

Some of the earliest applications of brain-monitoring technologies have explored the potential of using the brain as an control centred input device for a digital system. The allure of controlling an object, moving a cursor, controlling a vehicle and generally providing a high resolution communication channel via the brain has been a primary focus for researchers in the field of Brain-Computer Interfaces (BCI). One of the major aims of BCI research is to provide a communication channel for the disabled. The ability to control artificial limbs, communicate and navigate around an environment via the brain alone is clearly a very valuable and important branch of research in affording additional independence to the physically disabled.

Typing on a physical keyboard is the primary manner in which text-entry is performed on a computer. Similarly, the control of a digital cursor is performed by the relative positioning of a physical mouse. Both forms of input require physical interaction from

[1] Robert Jacobs Homepage- https://www.cs.tufts.edu/~jacob/

their users - a significant barrier excluding many physically disabled individuals from being able to use a computer system. BCI has the potential to provide a communication channel that extends control to the physically disabled. One of the earliest systems that enabled form of interaction was a spelling device was presented by Birbaumer et al. in 1999 [23]. Birbaumer et al.'s approach depended on the subjects ability to regulate activity in the Cerebral cortex of the brain in order to provide a binary response (positive or negative). Using this form of input, the speller would present users with 2 'word banks', with a negative response selecting one and a positive response the other. This processes was repeated until the desired letter was obtained. The system was tested on 2 patients with overall accuracy across a number of metrics reaching the 70% mark. Deploying the system however required significant training periods and allowed only for slow, limited and predefined forms of communication. Piccione et al. in a manner similar to Birbaumer et al., developed a P300 control signal based 2D cursor control using a four-choice (Up,Down,Left,Right blinking arrows) paradigm [176]. Five disabled and seven able-bodied subjects were able to successfully control a cursor with relatively short training periods, with average performance being 68.6% for patients and 76.2% for healthy controls. Gao et al. demonstrated an environment controller consisting of an LED stimulator, a digital signal processor, and programmable infrared remote-controller [64]. Gao et al.'s approach was based on using a visual stimuli from the LED stimulator which consisted of 48 green LEDS, each blinking at a unique frequency between 6 and 15Hz. Each LED frequency was then mapped to the infrared controller, which was programmed according to the patients environment e.g. LED with the blinking frequency of 6Hz was linked to trigger the same frequency of IR light (on the programmable infrared remote-controller) as the power button on the Televisions remote-control. Gao et al. utilised EEG and SSVEP (a phenomena similar to P300) and applied signal processing (via a digital signal processor) to classify the incoming brain signal accordingly. Using this approach a transfer rate of 68bps was obtained.

2.4.1.2 Passive Control, Adaptive Systems and User State Monitoring

A significant amount of initial research into the application of brain monitoring control was focussed on providing direct control, primarily as a communication channel for disabled users. There is also work however investigating 'passive' forms of control, where the mapping between mind and machine is less obvious to the user of the system. Described by Girouard as "interfaces that use brain measurements as an additional input, in addition to standard devices such as keyboards and mice" and "applications that pay attention to the user", passive control provides HCI researchers with a interesting range of

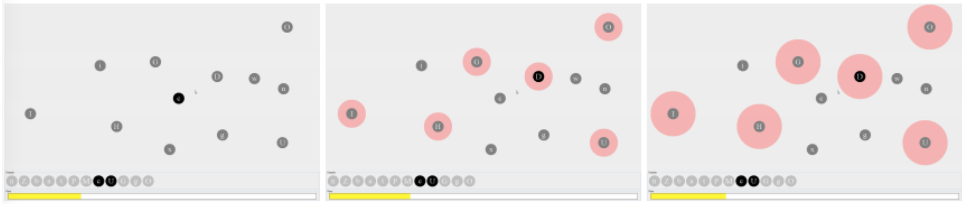


Figure 2.15: The target expansion conditions with visual indications of expanded target widths presented to participants [5]. Note - The participants would not see the bounding area (indicated in red).

potential application areas, including: adaptive interfaces/systems, user state monitoring and neurofeedback [68,87].

Lee and Tan present an EEG based system capable of classifying different types of tasks being completed by the user at a given time [122]. In their 2-part study, Lee and Tan investigate the ability of the system to detect both cognitive (rest, mental arithmetic, and mental rotation) and non-cognitive (relaxation, PC game with and without opponents) based activities. The system was able to obtain 84% and 92.4% task classification accuracies (respectively). The authors discuss the potential applications of such a system, describing an adaptive system that can support users depending on the task it believes them to be completing. Girouard et al. †, expands upon this idea by presenting an fNIRS system that dynamically adapts the background music depending on the task being completed by the user [70]. In a successive (dual) task study consisting of watching a video and playing a game of Tetris, the system was successfully able to determine which task the user was completing at a given time, and adapted the background music to suit the task. The adaptation of background music according to task engagement had a positive impact upon participants satisfaction, according to a post study questionnaire.

Zander and Kothe propose a fusing of BCI technology with cognitive monitoring in order to provide state information about the users' intentions, situational interpretations and emotional states into a real-time system and discuss the general-purpose applications that could arise from this pairing [252]. Zander, building upon this pairing of BCI technology with realtime cognitive monitoring, has gone on to broadly study passive forms of interaction, including: adaptive instruction in digital environments [66], detecting auditory errors [253] and documenting the differences between active and passive BCI's and their influence on the future of HCI [254].

Afergan et al.† present an fNIRS based system that dynamically adjusts the difficulty of a UAV based simulation task according to the users state [5]. The system is trained to detect extended periods of boredom or overload and adjusts the number of UAVs the participant must control accordingly. The dynamic system resulted in 35% less errors over the baseline condition (no dynamic adjustments) and demonstrates the real-time use of fNIRS. Yuksel et al. developed Brain Automated Chorales (BACh), a fNIRS based system that dynamically adapts the levels of difficulty in a musical learning task based on the users MWL [250,251]. Feedback from participants in the study indicate that the system was of an aid to them and suggested that they felt “more creative”. Afergan et al. † present a brain-based target expansion system that dynamically adjusts the size of ‘high importance’ target according to the participant current level of ‘multitasking’ [6]. A SVM classifier was used to identify a particular state of multitasking, prior to beginning the main trial of the study, the interface of which is shown in Figure 2.15. Participants were required to complete a dual-task study consisting of a n-back and visual search task. Results indicate improvement in the adaptive expansion condition across a number of measures when compared to a baseline of no expansion and a static (constant) expanding condition. Peck et al. used fNIRS as an indicator of users preference in a movie recommendation system [171]. A brain recommender system classifier was build using a support vector machine (SVM), a machine learning based approach. The classifier would take a 25 second recording of fNIRS data as well as the user’s rating (1-5 stars) for 6 films (3 favourite, 3 worst) selected by the participant. For each film, the participant was shown the associated IMDB page and then asked to rate the film. In a 14 person study, the recommendation system was better at recommending higher rated films that were personalised to the individual (over a control condition).

An example of the direct application of passive control informed by brain based measures for HCI is the issue of appropriate user interruptions. Knowing when it is appropriate to interrupt a user is an outstanding issue in HCI today and one that directly benefits from the application of BHCI. Inappropriately timed interruptions can increase errors, reduce efficiency and affect the ‘flow’ of the users’ output [95]. With the ever increasing amount of mobile devices, smartwatches and the Internet of Things (IoT) devices, the symbiosis between these decoupled devices and the user becomes ever more important [59]. Knowing how hard a user is working could be an indicator for the users willingness to be interrupted - without it affecting their workflow too negatively.

Solovey and Jacobs† discuss the possibility of fNIRS being used in this capacity, noting the technologies ability to distinguish between various levels of multitasking [209]. Chen and Vertegaal introduce Physiologically Attentive User Interface (PAUI), a prototype system that uses Heart Rate Variability (HRV) and EEG to regulate email, instant-messaging and phone-based notifications according to the user’s MWL [39]. PAUI classifies 4 different user states in which the level of notifications vary from none being permitted to all being permitted. A six person study indicated that PAUI could identify the ‘correct’ user state 85% of the time. Mathan et al. examined the possible military applications of such a system [141]. Through the observation of three military personal during a scenario based (entering and clearing buildings) training mission, the system was able to accurately reproduce the findings of others - indicating the system was capable of accurately measuring MWL. The authors conclude with a discussion for the potential of the system to modulate the delivery of battleground messages according to the individual's MWL. Tremoulet et al. found that the delaying of notifications until a user's MWL was low (assessed by EEG, HRV and GSR) increased the number of tasks the user could complete [223]. Afergan et al.† present Phylter, a physiological-based notification filtering system that "…that sends pertinent notifications to a user only when the user is in the proper cognitive state to handle additional information. The system uses physiological sensing as a means to time, suppress, and modulate information streams in real time" [4].

The current body of work in the field points to the infancy of applying BHCI for these passive forms of interaction. We see a significant focus placed on understanding the data that's being obtained from these BCI devices, with a number of the works above highlighting their data-processing system as their primary contribution, over that of the actual interaction itself. This leads to a significant body of work remaining unexplored, as our understanding of how to design, develop and support these forms of interactions has not yet been sufficiently explored in the literature.

2.4.1.3 Games and Entertainment

In recent years a number of consumer-grade EEG based brain-monitoring devices have entered the market place with the primary objective of providing a input peripheral to a computer based game. Examples of these headsets include the Emotiv EPOC, Neurosky Mindwave and Muse. Indeed games have been utilised as a training method for developing explicit forms of control using BCI [179].

Studies have investigated the application of brain based control in familiar and well known games. Reuderink et al. present "Affective Pacman" an adaptation of the original 'Ghost avoidance' game Pacman which measure players "Frustration" as they complete a level of the game, and would broadly be classified as an evaluation measure in this gaming context [190]. Tangermann et al. implemented direct control of a virtual pinball machine with players controlling the machines paddles via EEG [113]. Compared to a pseudo-random and no control conditions, the study results indicate that "fast and well-timed control" beyond the possibility of chance were obtained.

Pires presented a brain controlled game inspired by Tetris that allowed players to effectively control and position falling blocks as desired [180]. The study integrated three differing types of control based on P300 ERPs and sensomotor based imagery. The game is being used as a form of neurofeedback for the treatment of children with attention-deficit and hyperactivity disorder (ADHD), since it requires the focussing and calming of the mind to solicit precise control over the game.

Figure 2.16: The brain-controlled, Tightrope walking game 'MindBalance' where users must control the balance of an animated character in a 3D environment [117].

Laar et al. used the popular online game - World of Warcraft (WOW) to investigate the effect of introducing brain based control has upon user experience [115]. In a experiment consisting of 42 participants, Laar et al. augmented the shape and function of avatars in the game according to the alpha band power calculated from a consumer grade EEG device (Emotiv EPOC). Using a within-subjects study design, participants played two versions of the game 1) 'Regular' WOW; 2) with brain-based avatar adaptations. Post study questionnaires indicate that participants indicated that they felt they had less involvement and control over the game, but did not experience less fun with the introduction of the adaptations.

There are also examples of purely brain-based games, that is, games that have been explicitly developed for use with a brain based controller. Lalor et al. developed 'MindBalance', an EEG controlled game whereby players must attempt to balance an animated character in a tightrope walking game (control in 1D, balance - left or right) [117]. The game is set in a 3D modelled environment with dynamic camera positioning and lighting responding to the characters movements in the environment (Shown in Figure 2.16). The game utilises chequerboard patterns on either side of the characters balancing pole to allows for players to maintain the character's balance on the tightrope. A machine learning based processing pipeline is first trained to the players response to seeing each checker-board pattern, then during the game, the player will focus on the desired pattern (left or right) in order to maintain the characters balance. Leeb et al. developed a virtual

Figure 2.17: The interface for the software system CUES, which augments browser interactions recordings with EEG, audio and web event data [177].

reality game requiring users to steer a penguin character down a snowy mountain side [126]. The steering of the penguin was controlled via a standard games controller, but the ability to jump was controlled via EEG and was triggered when participants imagined a "brisk dorsiflexion of both feet" (jumping). 14 participants partook in the study, but only 7 achieved the required classification performance to complete the entirety of the study. The results indicate that the transfer of skill is possible (learning the ability to 'jump') despite the visually complex and rich environment in which the game was set. Doud et al. demonstrated the ability to control a virtual Helicopter in 3D dimensions using visual motor imagery using EEG and sensorimotor rhythms (SMRs), with 5 participants obtaining an average performance score of 85% through the piloting of the helicopter through rings in the environment [47]. Similarly, LaFleur et al. demonstrated a similar approach using a physical Quadcopter drone [116].

2.4.2 Evaluation

A number of HCI researchers have proposed the application of brain-monitoring technology in the evaluation of user's interactions with computer systems [114,69,77]. Wilson argued the value of using EEG or fNIRS based devices in informing and evaluating the design of Search User Interfaces (SUI) [242]. Earlier work conducted by this author (prior to initiating the work presented in this thesis) presented the software tool: Cognitive User Evaluation System (CUES), a system designed to aid in the collection and analysis of data collected from the Emotiv EPOC during web based interaction sessions (Shown in Figure 2.17)[177].

Nacke compared the user experience of using a Nintendo Wii remote and a standard PS2 Controller to control a computer game "Resident Evil 4" [157].
Using a combination of EEG and subjective measures (gaming experience questionnaires), Nacke was able to identify a relationship between the EEG readings and input modality. Nacke identified that Alpha and Delta power correlate with negative effect and tension for the PS2 controller, while the Delta and Theta power correlate with self-location for the Wii remote. Peck et al. † applied fNIRS monitoring in evaluating the effectiveness of various types of visualisations (Bar chart versus Pie chart) [173]. Participants were presented with a number of visualisations and asked to estimate the size difference between sections in 2 different graphs (presented sequentially). Results of the study indicate that fNIRS is able to capture the impact of visual design and suggest that there are no universal differences between bar and pie charts. Cherng et al. investigated the viability of using EEG in the evaluation of graphical icons [40]. In part 1 of their study, Cherng et al. investigated the relationship between the EEG data and how closely a presented icon matched to a textual description of a function. Results from the study showed that particular amplitudes of the EEG signal (N1) were larger for icons that were 'close' to the presented textual descriptor. In part 2 of their study, the researchers presented a number of icons (sequentially) with participants instructed to press a key when they believed they had seen an icon consistent with the function - 'Calender' (adaptation of the oddball study design). The results for part 2 indicate that 'far' icons had the lowest N1 amplitude.

2.5 Problem Statement

From the literature review above, we have established the state of the art in both the underlying theory and the application of BHCI in practice. The aim of this concluding section is to provide a clear link between the reviewed literature and the research questions/themes we present at the end of the last chapter. In doing so, we document the gap in the current knowledge of the field and detail how the work we present in this thesis aims to address these gaps and advance the state of the art.

Through the initial two studies (TAP and LEAP), we will explore the themes of **Method** and **Input Control**.

In our TAP study, we use the **method** set out by Solovey et al. [210] to apply BHCI in the evaluation of a widely used verbal protocol (Think-Aloud). fNIRS has been utilised for HCI based evaluation in prior works. We documented the study performed by Peck et al. [173], who successfully applied the recommendations set out by Solovey et al. to evaluate different types of chart visualisations. However, there is no documented application of fNIRS being used to evaluate TAP or other HCI research methods in the existing literature. Our TAP evaluation study therefore is an important contribution in verifying the integrity of using a TAP in an HCI study. Performing this work will allow us to identify the significance of the interference from the inclusion of TAP in a HCI study, and allow us to verify or question the integrity of studies that have utilised TAP.

Similarly, the modality of Speech and it's effect upon the fNIRS signal is not documented in Solovey et al's. recommendations. This leaves the unknown of knowing whether it's inclusion will interfere with the measurements obtained from the fNIRS, and affect the precision of a studies findings. Speech is a fundamental part of HCI, and the rise of voice based agents such as Apple's Siri and Amazon's Alexa, make the study of speech evermore relevant in HCI going forward. In validating the inclusion of Speech in fNIRS based BHCI studies, we are enabling other HCI researchers who want to use a BHCI approach for studying speech based interactions.

In our LEAP study, we apply BHCI to evaluate different forms of **Input Control**, designing a study that examines the effect of an augmented reality based controller against more common physical and mouse control. As indicated above, evaluation using BHCI is not novel, however, it's application in studying the impact of various types of input control is. Similarly, there is little work documenting the differences between these forms of input control in general, indicating that this may be a contribution to HCI in general. Studying the impact that different forms of inputs have upon users is another demonstration of applied evaluation using BHCI, and is analogous to existing tasks that are performed using subjective techniques. Finally, the specification of a suitable task, one that facilitates our basic study requirements must also be developed. We were unable to identify an existing task that would allow us to compare across these different types of control, whilst allowing us to vary the difficulty of the task.

These two initial studies also address a gap in our existing understanding of what makes a BCI suited to facilitating Natural Interaction, and what properties of a BCI we should seek when designing and developing such experiences - **RQ1**. Through applying different forms of BCI technology (fNIRS and EEG) in user studies exploring Natural Interaction, we will gain insight into what the technical considerations should be when developing a Natural Interaction experience. Current work prioritises the process of translating the signal from a BCI into something meaningful (signal processing), whereas, here we focus on understanding what is required of the BCI technology in order to facilitate Natural Interactions from the point of view of the user. This is important in understanding how we can develop and deploy experiences that are comfortable and engaging for users.

In our final study, #Scanners, we explore the theme of **Novel Interaction** by developing, deploying and analysing a new form of cinematic experience using passive BHCI based control. The development of a passive BHCI based cinematic experience has not been documented in the current literature, despite the presence of similar concepts such as adaptive systems (e.g. Girouard et al's. adaptive background music [70]). In exploring the theme of **Novel Interaction** we set out to answer our remaining research questions which centre around understanding and facilitating this type of interaction. Specifically, this will focus on contributing to our understanding of how these experiences can be developed, and how they are received and experienced by users.

As detailed above, a significant focus in the current literature is placed on identifying techniques for sense-making of the signal that is obtained from the device, rather than the more humanistic elements that arise from the interaction itself. Whilst the initial two studies focus on understanding the technical requirements of the BCI (**RQ1**), here we focus on understanding the interaction itself. Solovey et al. provide an example of how one may apply fNIRS for HCI based research studies [210], but it is unknown whether these considerations extend to the passive forms of interaction we propose here. **RQ2** interrogates the current lack of detailing around the role BCI can play in developing these types of experiences.

Finally, our understanding of how participants might experience this type of interaction is also not documented in the current literature, yet another result of the lack of work investigating this form of interaction. Through **RQ3** and **RQ4** we aim to identify key considerations in the design and development of natural interaction experiences, and provide a clear set of guidelines for enabling this.

Chapter 3

Methodological Approach

As a scientific method, HCI is founded on the principles of empirical research, specifically relying on the observation of human participants. The existing literature provides a diverse array of methodologies that could be applied in exploring the research agenda set out in the motivating chapters of this thesis. In this section we present a selection of available methodologies that are relevant to exploring our research agenda. We detail the relative advantages and disadvantages of each and relate them to the requirements of each individual component of our research agenda. We conclude by detailing and justifying the approach that we have chosen to apply in this work.

3.1 HCI Research Methods

According to MacKenzie [135], research methods in HCI fall into 1 of 3 categories:

1. Observational Research
2. Experimental Research
3. Correlational Research

MacKenzie characterises each of these methods according to their trade-off between relevance vs. precision. In this context, relevance relates to the study of real world phenomena and holds an inherent practical value, as it observes real study phenomena. Precision, is the measure of identifying precisely how much a particular variable of interest has changed.

Widely used in HCI research, Observational methods focus on the study of behaviour exhibited in natural settings (as opposed to an artificial lab). This approach tends to be qualitative with data collection techniques involving interviews, case studies and think aloud protocols. In applying this approach, researchers hope to gain insight into how human qualities such as emotion, feeling, sentiment etc, are affected by the interaction between the participant and interaction under study. Observational methods therefore are highly relevant, as they pertain to actual human behaviours observed in their natural habit. This, however, comes at the cost of precision. Measures of emotion, feeling and sentiment are inherently difficult to quantify and are therefore difficult to map precisely, especially in an uncontrolled natural environment. Usability evaluation is a classical example of applied observational research in HCI. In performing usability evaluation, a researcher will attempt to identify problems that affect a user interface's usability and elicit feedback on how the design of the interface could be improved to better facilitate the needs of a user.

Experimental research methods are typically conducted in a controlled, laboratory setting with typically, a single variable being manipulated during the course of the study. In the context of HCI, this could be the controlled manipulation of an interface component in various configurations during a lab based study. The researcher would measure a quantifiable response during this manipulation and record the differences between the various configurations. Examples of possible measures include: EEG, fNIRS, task speed, task accuracy, etc. This approach, sometimes referred to as the *scientific approach* is common in the medical and psychological sciences, where precise attribution of effect needs to be tied to a particular cause. This is therefore a precise method, since we are able to calculate the precise amount off affect attributed to a particular variable manipulation. This precision, naturally, affects the relevance of these findings however, since the setting in which these measures are collected are not normal or natural to the user. In HCI, we would call this a compromise in the *ecological validity* of the task.

Correlational research is a statistical based method for identifying a relationship between 2 sets of data. Identifying the relationship between a participant's IQ and their social network activity could be one example of applying this technique. This approach often accompanies an Observational or Experimental method e.g. Analysing a post-study questionnaire. This data is not directly collected under a controlled setting, but is nevertheless reflective of a user's preference. This technique represents a balance between relevance and precision.

MacKenzie, using the relationship between relevance and precision, has mapped each of these methods on a chart shown in Figure 3.1.

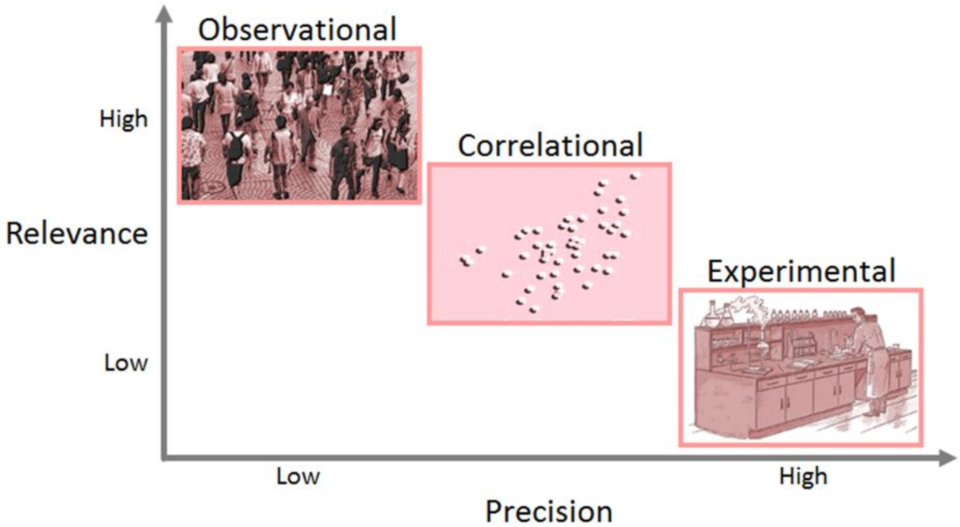


Figure 3.1: HCI methods: Relevance vs. Precision. Image Credit - MacKenzie [135]

3.2 A Hybrid Approach

From the high level overview of the studies presented in this thesis, the reader will note a mixing of the above methodologies in our final studies. We decided upon implementing a hybrid/mixed methodological approach in our studies, primarily to ascertain a balance between the relevance and precision of the results - relative to the respective study aims. Specifically the first two studies *TAP* and *LEAP* will be conducted using an *Experimental Research* methodology and our final study, *#Scanners* will be conducted using an *Observational Research* methodology. #Scanners will specifically use the *Performance-Led Research in the Wild* methodology proposed by Benford et al. which we detail below [19].

In our TAP and LEAP studies we apply a *Experimental Research* methodology, based in a Laboratory setting in exploration of **RQ1**, as well as the **Methods** of applying BHCI and evaluating different forms of **Input Control**. To achieve this we use Mental Workload (MWL) as the basis of comparison between task study conditions. MWL has a significant body of literature documenting the application of the measure across a variety of BCI technologies in a number of different environments. MWL provides a useful abstraction for reasoning and modelling how a participants mental activity was affected during and between studies. We will therefore apply MWL as the basis of evaluation for these initial studies.

One possible approach in organising this research, especially in exploring **RQ1**, would have been to apply the same task using different BCI technology. This would give a strong basis for comparing the relative properties of each BCI for the same task, providing a controlled comparison of the technology, without the potential influence of a different task. Equally, such an approach would allow us to validate the findings of one technology with the findings of another - giving further validity to our findings. This would have been a valid approach, and is an example of important work that needs to be performed at some point in the maturation of a BHCI as a field of research. However, as per our stated research themes, one of the intentions of this body of work is to demonstrate the broad application of BHCI, especially as it relates to traditional HCI research. In addition to answering our stated research questions, the thematic interests serve to contribute both insight and interest in applying BHCI.

In our final study, #Scanners, we employ an observational research method in the exploration of the remaining research questions (**RQ2-4**) and the theme of **Novel Interactions**.

In developing this natural form of passive interaction, we move away from focussing upon MWL, and more towards measures that are associated with entertaining experiences. We employed MWL in our TAP and LEAP studies since they focussed on evaluation. However, this doesn't necessarily imply that MWL would provide the basis of an entertaining cinematic experience, especially one where there's potential for the viewer to exert control. We also wanted to explore an interplay between different measures to provide a dynamic, multi-layered experience to the viewer. With this in mind, the artist and researchers involved in #Scanners explored a variety of different measures provided by consumer grade EEG devices (out of the box), in order to develop this interaction. The sacrifice here is in precision. We cannot know, for example, that the measure of "Attention" we receive from the BCI used in #Scanners is either meaningful or accurate. We may however observe how the viewer interacts with this measure, to see whether they discover this ability to control and whether they chose to exert it. Again, this is a shift in the focus from traditional BCI, where the focus is typically placed on the understanding of the signal. Here, we are concerned with the application of this signal and how it is perceived by the user. We want to develop, explore and understand a novel new form of interaction that our literature review indicates is truly unique and unknown territories in BHCI, and HCI more broadly.

The choice of methodology for #Scanners and the related research questions we attempt to answer through it's study, are critically important therefore. We chose to use the Performance-Led Research in the Wild (PLRW) methodology in exploring #Scanners. PLRW is an observational research methodology that HCI researchers can apply in the study of novel performances and interactions [19]. Benford et al. provide a framework for evaluating novel performances which have been created by an 'artist' and are 'supported' by HCI researchers, much like #Scanners. The framework consists of 3 activities - practice, studies and theory, which interact in a complex manner through nine different relationships, shown in Figure 3.2.

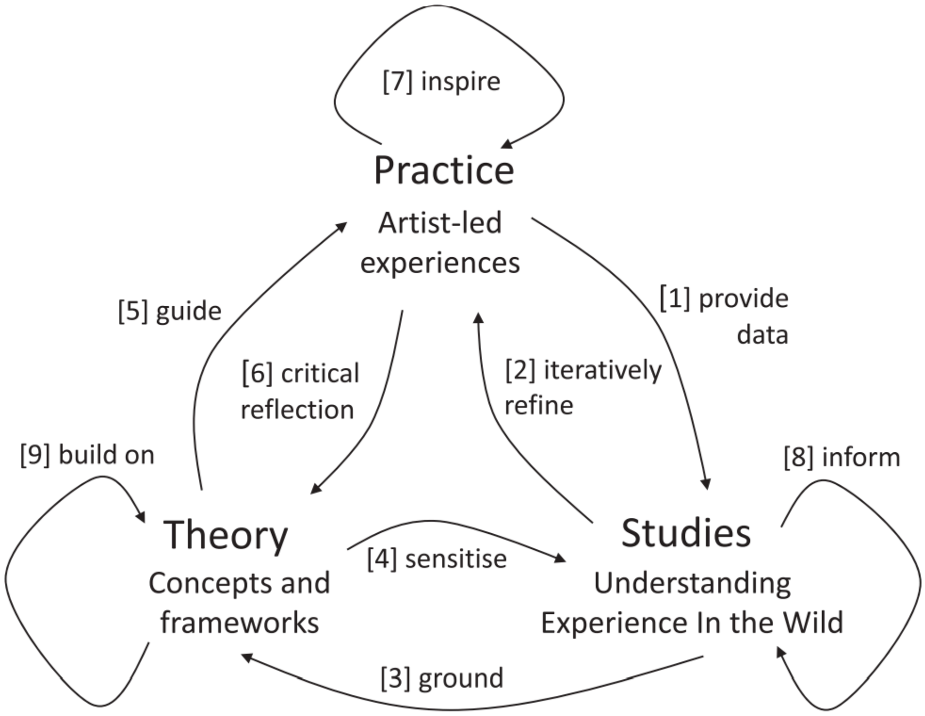


Figure 3.2: Overview of Performance-Led Research in the Wild. Image Credit - Benford et al. [19]

The appeal of the framework is it's accounting of historical projects that have benefited from it's application. The documentation of PLRW provided by Benford et al. includes a number of projects that are similar in structure and interaction to the work we explore in #Scanners. Broncomatic, a breath-controlled bucking-bronco ride where the movement of the ride is partially controlled by the rider's breathing [138]. This exhibits a number of elements that we will encounter in #Scanners, including the interplay between entertainment and research as well as exhibiting a form of passive/indirect interaction between the participant and experience. Benford also documents the overarching challenges of balancing artistic and research interests a balance we must be careful to maintain in #Scanners, to ensure success from both an entertainment and academic point of view.

We believe our chosen approach towards a mixed methodology, as presented here is both necessary and appropriate in pursuing and answering the research questions set out in our research agenda. In choosing a mixed method approach, we are able to selectively sacrifice the variables of precision and relevance in accordance to the particular aspects of the research agenda that are being analysed in a given study, as documented in this section.

Chapter 4

Evaluating fNIRS in BHCI via Verbal Protocols

The work presented in this chapter was performed in collaboration with the following co-investigators: Mr Horia Maior, Mr Martin Porcheron, Dr Max L. Wilson and Professor Sarah Sharples.

4.1 Introduction

In this chapter, we explore the application of functional Near Infrared Spectroscopy (fNIRS) as a brain-sensing technology within a HCI style user study. We conduct this work in contributing towards the research question **RQ1**, which aims to identify the suitability of BCI devices/technology in these settings for enabling natural interactions. In this chapter we attempt to answer **RQ1** for the BCI technology - fNIRS.

fNIRS has received recent focus in HCI research for its amenability for more ecologically valid study conditions [210,87]. It is understood that fNIRS permits more natural movements typically associated with using a computer without being subject to significant artefacts in the data.

Our initial investigation into the application of fNIRS follows the work of Solovey et al [210]. In their study, Solovey et al. empirically examined typical human behaviours (head/facial movement and keyboard/mouse interactions) to identify if these behaviours had a significant effect upon the measurements that were obtained from the fNIRS. From the study results, Solovey et al. provide a set of recommendations that HCI researchers should consider when using fNIRS in HCI. These results are summarised in Figure 4.1.

Table 1. Summary of considerations. Legend: ✓ indicates acceptable, **C** indicates to correct, and × indicates to avoid or control.

Considerations	**Result**	**Reference**	**Correction Methods**
Forehead movement	×	Exp 4	
Major head movement	×	Exp 3	Use chin rest
Minor head movement	**C**	Exp 3, [20]	Filter
Respiration and Heartbeat	**C**	[4, 20]	Filter
Mouse Clicking	✓	Exp 2	Collect signal during a clicking only task
Typing	✓	Exp 1	
Ambient Light	**C**	[3]	Wear isolating cap
Hemodynamic Response	✓	[1]	Expect 6-8s response
Ambient Noise	**C**	[22]	Minimize external noise
Eye Movement and Blinking	✓	[16]	

Figure 4.1: A summary of the recommendations provided by Solovey et al. Figure Source - [210].

These recommendations by Solovey et al. provide a foundation upon which the design of this study is built. We may assume, for example, that interaction via mouse clicking, typing and general humanistic movements (eye movement/blinking) do not require any special consideration in this study. Solovey et al. do recommend however that certain naturalistic movements be controlled, including general head movement - which have affected how we designed and deployed the study. For example, participants were briefed to "restrain from major head movement" whilst completing the tasks. This demonstrates the value of such guidelines, and reinforces our interests in exploring **RQ1** and **RQ4**.

Solovey et al's. recommendations are also a basis of our interest in exploring the **Methods** of applying BHCI. Through conducting this study we contribute to the body of work validating the recommendations outlined by Solovey et al. and demonstrate the application of BHCI research, through the use of fNIRS. Additionally, we also explore the use of Speech, a modality not yet documented by Solovey et al.

Although the suggestion is that fNIRS can be used more easily within natural, ecologically valid user study conditions, it is not clear the extent of which fNIRS can be used for enabling sustained, comfortable forms of natural interactions. Current research is still limited to performing controlled simple Working Memory(WM) tasks (e.g. [173,210]), in controlled study settings. In-line with answering **RQ1**, we explore the extended use of fNIRS under 'normal' (minimally controlled) HCI study settings. Additionally, participants complete a task using a common, speech based verbalisation protocol. Speech, an enabling component of natural interaction was not investigated by Solovey et al. We further contribute to the literature in this regard.

The Think-Aloud Protocol (TAP) is a widely used research method [144], utilised in a variety of research fields including Human Computer Interaction (HCI). In the context of HCI, TAP is typically used as an evaluation method to elicit insights into participants thoughts and strategies during usability and user studies. TAP, however, has also been used in other settings, such as cognitive psychology and social sciences [38], to understand phenomena such as user mental models, expertise, and problem solving. TAP is a verbal protocol in which participants are required to verbalise their thoughts and actions as they perform some interaction task. Since TAP will use resources from verbal working memory, it is fair to assume that the inclusion of spoken protocols will potentially affect cognitive processes due to use of available resources. Consequently, TAPs may affect performance in tasks, and also measures of workload during studies.

As well as being a core part of user studies, verbalisations are also closely related with WM, as both the interpretation of words in the task and the integration of thoughts involve the phonological loop [237]. Consequently, to integrate fNIRS measurement within a typical user study that might involve a TAP, we have to be aware of how one will affect the other.

There are various forms of TAP, including retrospective, which occurs after a task has been completed, and concurrent, which occurs during a task. Of concurrent forms of TAPs, there is both invasive, which involves directly questioning participants, and passive, which simply encourages participants to maintain verbalisations about their thoughts and actions. Because fNIRS measurements are taken during tasks, this paper focuses on concurrent TAPs.

In this chapter we present a user study that examines the impact of:

a. Nonsense Verbalisations
b. Passive Concurrent Think Aloud
c. Invasive Concurrent Think Aloud

All conditions were compared to a baseline of silent non-verbal working memory. We then present the results of the study, discuss the findings in terms of what we can learn about the impact of TAP on MW in general. We also provide some recommendations for using fNIRS in both general BHCI based user studies and detailing it's suitability in enabling natural interactions (**RQ1**).

4.2 Related Work

4.2.1 Think-Aloud Protocols

Ericsson and Simon's seminal work on verbal reporting of participants thought process is the most cited amongst Think Aloud Protocols [167]. Prior to this work, consideration was made to the type of verbalisation produced by participants under study conditions [84]. In their original discussion of TAP, Ericsson and Simon [53] distinguish between 3 distinct levels at which verbalisations occur. Levels 1 and 2 are described as being valid representations of a participant's mental state, since they are verbalising information stored in short term memory and are representative of the participant's current state. Level 3 requires access to long term memory and influences what would otherwise be their typical state. Ericsson and Simon's version (Levels 1 and 2) of the protocol is strictly non-intrusive, and researchers implementing the protocol are restricted to simply using the verbal prompt -"Keep talking"- to avoid influencing the participant, and ensuring that the reported data relates solely to the task. To distinguish between other forms, we refer to this type of TAP as Passive (PTAP) for the remainder of this paper.

In practice, however, researchers generally misreport or incorrectly implement the TAP they are using [144]. Practitioners of TAP often prefer to question participants at level 3 to obtain coherent, actionable utterances relating to the system under evaluation, instead of inferring results from level 1 and 2 utterances. Researchers have attempted to formalise this level of questioning [49,84]. We characterise these approaches under the umbrella term Invasive TAP (ITAP). With ITAP, researchers are free to probe the user's mental model, but Ericsson and Simon would disregard the findings at these levels stating that they have been influenced. Under ITAP, a practitioner is able to prompt the participant with more probing questions - "Why did you do X?".

4.2.2 Working Memory (WM)

In an attempt to characterise and model the cognitive processes involved when a participant is undertaking a TAP, we draw on research into WM and MWL, as presented in our Literature Review. We use these theoretical models to characterise the mental processes and processing centres (WM models) that are utilised during the TAP task, and how much cognitive work is affected by the utilisation of particular resources (MWL models) as we vary the type of TAP being utilised. Here we apply these theoretical models to the study tasks to rationalise how different TAPs may affect participants' MWL and ultimately their task performance.

We can relate a number of concepts described by Ericsson and Simon to the working memory model described by Baddeley. For example, Ericsson and Simon note that verbalisations at level 1 and 2 occur within Short-Term memory (STM). We can further characterise this with Ericsson and Simon stating that TAP will utilise the Phonological loop as it is verbal in nature. Tasks under which the TAP is performed may also interact with other components of the working memory model. Tasks involving imagery or mental rotation, for example, will utilise the visuo-spatial sketchpad since they are spatial, whereas verbalising occurs in the phonological loop. For such tasks under TAP conditions the two concepts of the model will be activated, with the central executive mediating information flow between the two. The episodic buffer may also have a role under ITAP conditions, since the protocol will require access to memories that are not in the immediate short term memory. We would not expect the Episodic buffer to be utilised in the PTAP condition.

In addition to the WM model, we can also consider the Information Processing Model and Multiple Resource Model (MRM) proposed by Wickens. Through his models, Wickens states that necessary resources are limited and aims to illustrate how elements of the human information processing system such as attention, perception, memory, decision making and response selection interconnect. We are interested in observing how and when these elements interconnect under TAPs.

We can use Wickens' MRM to model the effect TAP will have on participant cognition. One of the key roles of the MRM is to demonstrate the hypothesised independence of modalities and use this to design tasks. We know for example that the inclusion of TAP will introduce additional Auditory resource requirements, since the participant will hear their own verbalisations. This in turn will require additional Perception from the participant and will draw on their Verbal coding resources and Vocal Verbal responses.

4.3 Experiment Design

The primary aim in conducting the work presented in this study is to evaluate the use of fNIRS in a BHCI context. In conducting this study we hope to answer **RQ1** and establish whether fNIRS is suited to facilitating natural forms of interactions. To explore and answer **RQ1**, we have developed a task which seeks to evaluate a commonly used verbalisation task - TAP. In performing a TAP, participants engage in a common form of Natural Interaction - speech.

In developing this study task we also investigate how verbalisation and TAPs affect cognition and the thought process during user study tasks. We produced three research questions which are specific to this task:

TAP-RQ-1) How can we identify the impact of TAPs on human cognition and mental workload using fNIRS?

TAP-RQ-2) What are the most suitable measures to sense such an impact?

TAP-RQ-3) How can we sense the reduction of available resources due to integrating a TAP concurrently with a task?

To answer these research questions, a theoretical understanding of TAPs, human cognition, mental workload and the interconnection between these concepts is required. One way is to look at the relationship between TAPs and concepts like working memory and workload. Therefore, it is useful to describe both ITAP and PTAP in relation to a theoretical model of WM. Wickens' Multiple Resource Model [237] can describe the relationship between the available resources and task demands. When performing a task, a person perceives both their own verbalisation and/or external stimuli as an Auditory modality. During "think aloud" we also process information (make decisions, store memories, retrieve memories, etc.), and output them as a response (e.g. as a Vocal Verbal encoding). Therefore, TAPs might have an impact on all three stages (perception, cognition and response) of the Multiple Resource Model. According to the model, a TAP is a verbal/linguistic activity, therefore the codes of its cognition stage is Verbal. Consequently, we chose a task (described further below) that was easy to verbalise and involves continuous use of the phonological loop, such that different verbalisation conditions would interact with the task.

Primarily, we wanted to compare the different concurrent TAPs against a baseline of not verbalising. In order to check whether simply using your voice creates an artefact in the fNIRS data, as opposed to thinking and talking, we also included a second baseline of repeatedly verbalising the word 'blah'. Type of verbalisation, as primary independent variable, created four conditions:

1. Task Only (Baseline - **B1**)
2. Task + "Blah blah blah" (Baseline - **B2**)
3. Task + Passive Concurrent TAP (**PTAP**)
4. Task + Invasive Concurrent TAP (**ITAP**)

We designed a repeated measures, within-participants study to compare these conditions, where participants solved eight mathematical problems. Conditions and tasks were counterbalanced using a Latin-square design.

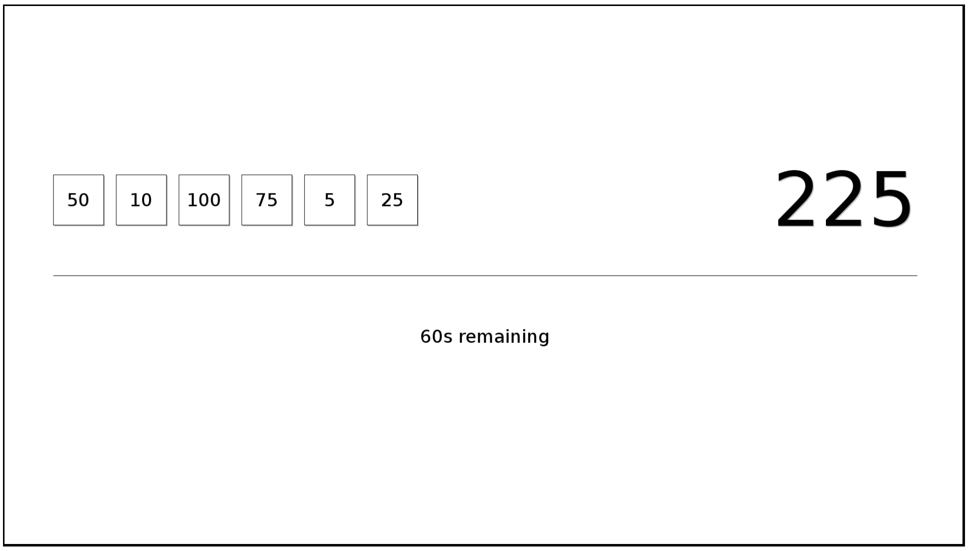


Figure 4.2: A Screenshot of the study task.

4.4 Hypotheses

We had a number of hypothesis that we sought to investigate whilst conducting this study relative to performance, cognition, and participants' grouping based on mathematical performance (High and Low performing groups):

HP - There will be a significant difference in performance between verbal conditions.

HC - There will be a significant difference in cognition between verbal conditions.

HP and HC were drawn from Wickens 4D MRM [237]. Both TAP and mathematical tasks should primarily use verbal working memory in the modality, encoding, and processing dimensions. Consequently, the demands imposed by various verbal conditions may affect the total workload element, and workload may then affect performance.

HC.S - There will be a significant difference in cognition between verbal conditions for high performing participants.

HC.W - There will be a significant difference in cognition between verbal conditions for low performing participants.

Depending on how well participants performed during the four conditions, we distinguished between high performing participants (top half) and low performing participants (bottom half) [147]. These groups were formed to investigate whether TAPs have a different impact on cognition relative to the participants grouping.

In order to determine how TAPs affect the different stages of the MRM, the task had to be chosen carefully such that verbalisation could potentially interrupt the process. The first criterion, therefore, was that the task should primarily use the phonological loop, and thus be a verbally oriented task. Second, the task had to involve continuous use of the phonological loop, and so a simple and discrete memory task was not sufficient. Third, the task had to be verbalisable for the TAPs, which also meant that a memory task was not sufficient. Fourth, the task also had to have various levels of difficulty to enable control over the primary task mental demands; according to the resource vs demands model [146] harder tasks would increase demand and thus reducing participant's resources for engaging in the TAP. Finally, performance on the task had to be measurable in order to determine the effect of verbalisations. Based upon these five criteria, we decided on using a mathematics task. Participants were provided with a set of six numbers and had to get as close as possible to a target final number. This problem is a variation on what is commonly known as the countdown problem, based on the mathematical challenge presented to contestants of the popular UK TV quiz show "Countdown". Each number may be used only once (although there is no requirement to use every number), and participants have 60 seconds to reach as close to the target number as possible, using four operators: addition, subtraction, multiplication and division.

36 versions of the task were created to be used across the four conditions, at four levels of difficulty (easy, quite easy, quite hard, and hard). Under each of the four conditions, a participant would be asked to solve 9 problems. These problems were balanced according to their difficulty, such that there was at least 2 of each class of difficult in each condition. Problems were counterbalanced between conditions. To classify difficulty, one researcher and two independent judges rated the difficulty of each problem. Difficulty was judged in four categories: . Inter-rater agreeability was confirmed with a Cohen's Kappa test, where the researcher achieved scores of 0.6419 (substantial agreement [118]) with the first independent judge, and 0.8571 (almost perfect agreement) with the second. This agreement was used to ensure that problem difficulty was balanced between conditions.

4.5 Participants

Twenty participants (14 male, 6 female) with an average age of 28.55 years were recruited to take part in the study. Participants were recruited from the University of Nottingham, and included a mix of staff members and students from a range of disciplines. All participants had normal or corrected vision and reported no history of head trauma or brain damage. The study was approved by the school's ethics committee. Participants provided informed consent, and were compensated with £15 in gift vouchers.

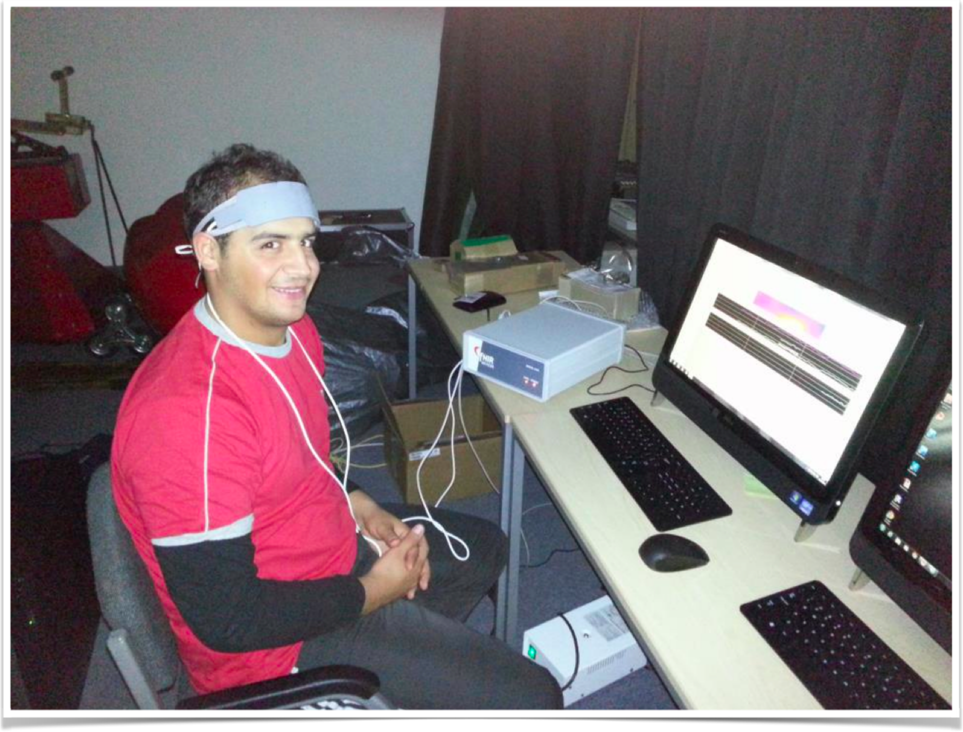


Figure 4.3: A participant, wearing fNIRS in the environment where the study was conducted.

4.6 Procedure

Participants began by providing informed consent and demographic information. We then attached the fNIRS to the participant and adjusted the gains for each channel as specified by the device's user manual. This adjustment is to account for individual differences between participant's skin pigmentation and bone density (which affects the path of the emitted infrared light). The adjustment accounts for these individual differences and ensures a consistent measure between participants.

Participants were first introduced to the task that they would be completing during the study. They were given two practice runs of the task (under baseline conditions) to familiarise themselves and reduce the impact of learning in their first condition. Once comfortable with the requirements of the task, participants were fitted with the fNIRS brain imaging device, which was placed upon their forehead targeting the PFC. At this point participants entered the recorded section of the study. During this stage, participant input was captured, verbalisations were recorded via microphone, and brain data was captured on a separate, calibrated machine.

Participants partook in four conditions which were counterbalanced using a Latin square rotation. Each condition began with a tutorial and practice session. The tutorial session was used to train the participant on how to verbalise according to the specific TAP being used in the particular condition. The practice session would then serve as an opportunity to trial the technique prior to beginning the test itself and thus reducing the interference on the first task in each condition. Each condition included eight of the tasks described above.

For each of the eight tasks in each condition, participants were given sixty seconds to attempt the problem. All calculations were performed mentally; pen and paper was not provided. After the sixty seconds had elapsed (or if the participant decided to proceed prior to this), participants were prompted to enter the number they had achieved during the calculation period. To avoid participants simply entering the target number, they were prompted to recall their solution. The solutions provided by participants were recorded by the researcher on paper and later digitalised.

After each condition, participants completed a standard NASA TLX form to subjectively rate their mental workload during the task. Each condition concluded with a thirty second rest period where the participants were asked to remain still, relax and empty their mind of thoughts.

The study was conducted in an office-like environment, with the participant sat at a desk with a standard desktop computer and 20" monitor. The room was quiet but not sound-proof, as the space was proximity to peoples normal offices and working environments. This was an important consideration as many brain based studies are conducted under strictly controlled lab settings. The office environment provides a more naturalistic and ecologically valid setting.

Within each task, participants were given 60 seconds to perform their calculations, and then input the number they had reached when they either achieved their goal or ran out of time. To avoid participants simply entering the target number every time, participants had to explain to the researchers how they reached their answer. In the two TAP conditions, the method of solution was the focus of the verbalisations, but for the silent and nonsense conditions, participants explained their calculations after each task was completed.

After a condition was completed, participants filled in a NASA TLX form to subjectively judge the mental workload of the task condition, before briefly resting and moving on to the next condition.

Having completed the study, participants were given a short debriefing period, where the researchers informally questioned the participants regarding their experience during the study. Specifically, the researchers focussed on the comfort and impact that participants experienced whilst wearing the fNIRS device. The aim of this line of questioning was to establish evidence for and against using fNIRS for extended periods in HCI centred studies - contributing towards answering **RQ1**.

4.7 Measurements and Equipment

We collected various types of data during the study. The data can be categorised into two groups:

- Performance during the study (**P**)
- Cognition (**C**)

4.7.1 Task Accuracy - *P*

The primary measure of task performance was measured according to participants accuracy in solving each task. Accuracy was calculated using the distance from the target answer for each of the 36 problems across the four conditions. Because the target varied, we used measured distance from the target as a percentage, which was subtracted from 100%. 100% represented the correct answer, 95% as being 5% from the target, and so on. As the results tended towards the target, task accuracy was analysed. To provide incentive to submit actual rather than ideal answers, we also measured whether participants could recall the solution to their answer.

4.7.2 Task Time - *P*

Task time was measured for each of the 36 problems performed across the four conditions. We note that participants were not encouraged to solve the problem in the shortest possible time, rather, they were asked to get as close possible to the target. It is for this reason that task accuracy (above) is considered the performance measure for this study, but we are still interested in exploring how task time was effected under each condition. As task time was weighted towards the sixty second limit, time was also analysed as non-parametric.

4.7.3 NASA-TLX questionnaire - *C*

We used the NASA-TLX questionnaire, a subjective workload assessment tool [80], based on the weighted average ratings of six sub-scales including, in order:

- Mental Demand
- Physical Demand
- Temporal Demand

- Performance
- Effort
- Frustration

Each participant was asked to self rate their mental workload using the NASA-TLX once after each condition. We additionally investigated each of sub-scales independently.

4.7.4 fNIRS data - *C*

fNIRS data was recorded using an fNIRS300 device and the associated COBI Studio recording software provided by Biopac Systems Inc. The headband shaped device is a sixteen-channel transducer for continuous Near Infrared Spectroscopy (NIRS). The headband consists of four infrared (IR) emitters operating on a range between 700 to 900 nm, and ten IR detectors. The device is placed on the PFC targeting the Brodmann area 10 (BA10). Oxygenated haemoglobin (HbO) and de-oxygenated-haemoglobin (Hb) are both strong absorbers of light, whereas skin, tissue and bone are mostly transparent to NIR light, this property is typically referred to as the **optical window** [98]. The tissue is radiated by the light sources and the detectors receive the light after the interaction with the tissue. See Figure 4.4 for an illustration of how the headband is positioned, and to visualise the path that the light follows during operation [46].

Preprocessing was performed to transform raw data from the device into oxygenation values using the Modified Beer-Lambert law (MBLL) [233]. We also applied filtering algorithms to remove high-frequency noise, physiological artefacts such as heartbeats and motion derived artefacts. To perform this preprocessing step we used the Matlab Toolbox, NIRS-SPM [247]. We performed de-trending using a discrete cosine transform with a high frequency cut off of 128 seconds. The baseline, which was recorded prior to beginning each condition, was subtracted from the signal using NIRS-SPM , and low pass filtering was performed with a Gaussian filter with a width of 1 second.

We also considered the delay induced by the haemodynamic response [233] by omitting the first 10s of the trial when processing the data [173].

The Biopac fNIRS device used in this study provides 16 channels of brain data readings. A channel is defined by the relationship between a source and detector pair as shown in Figure 4.4. From the MBLL we receive Hb, HbO and TotalHb (Hb + HbO) values for each channel. Measures were synthesised by combining specific channels averages to form a single measurement. Channels 3,4,5,6 were used to represent the left side and channels 11,12,13,14 formed the right side in these measurements. An overall measurement was produced by averaging the data from all 16 channels (see Figure 4.4).

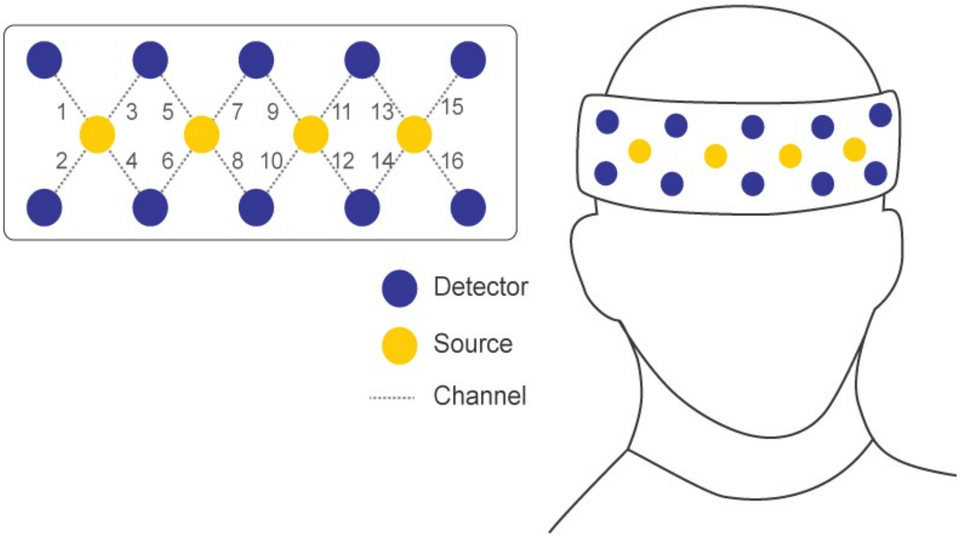


Figure 4.4: Sensor layout for the Biopac fNIRS used. Figure provided by Hyosun Kwon.

4.8 Experiment Software

When designing the study we placed a strong emphasis on automating the running of the study and collection of the associated data. With the exception of the brain data, all other measures were collected from a single program. Experimental data was synchronised using network time. We developed this program using PEBL: The Psychology Experiment Building Language [151]. The language provides a convenient set of features including accurate experiment timing and predefined psychology/study procedures such as demographic questionnaires. Of particular relevance to this study was the pre-defined, computerised version of NASA-TLX.

4.9 Task Results

In this section we present the findings relating to the study task specifically. We provide a more detailed discussion of the application of fNIRS to BHCI based studies in the following section.

We began by checking for ordering effect. A one way repeated measure ANOVA showed that participants performed significantly slower in the first condition they experienced, while average time to complete the subsequent conditions was even ($F(19, 3) = 2.816, p < 0.05$). An LSD post-hoc ANOVA test also showed that average scores also improved between the first condition they experienced and the last ($F(19, 3) = 2.271, p < 0.05$).

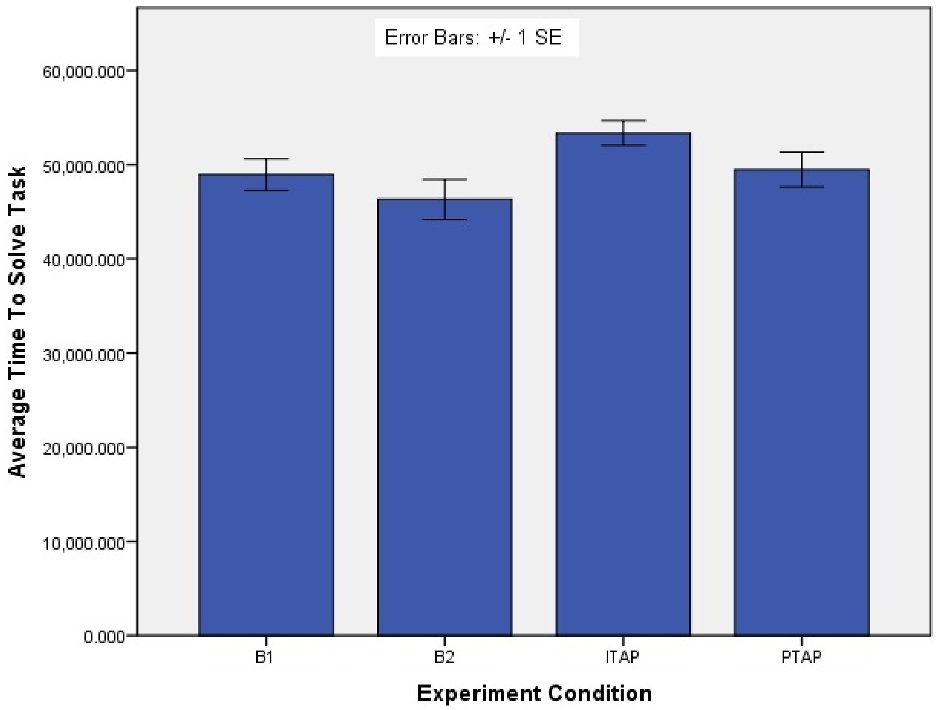


Figure 4.5: Mean time to solve a set of tasks.

4.9.1 Performance

Performance is characterised as the distance of the participants solution to the target answer as a percentage. Against hypothesis HP, our analysis showed no significant difference in task accuracy between conditions. We found no significant difference in performance between any of the four conditions, however, under the TAP conditions, participant performance slightly improved. There was also no difference in the number of tasks correctly calculated in each condition. We hypothesised that, ITAP under time pressure would cause performance to drop, but instead these results support the findings of McDonald et al. who found that neither form of TAP affected performance [143].

A significant difference was found in terms of time to complete tasks (Figure 4.5). As perhaps expected, participants took significantly longer to solve tasks in the ITAP condition ($F(17, 3) = 9.895, p < 0.01$) relative to the other three conditions (B1: $p < 0.005$, B2: $p < 0.001$, PTAP: $p < 0.05$). PTAP was not significantly different to B1 or B2. This time difference was likely created by the additional time required to explain decisions being made. %Performance can also be characterised by the time participants took to solve a problem. Participants were not asked to solve the tasks in the shortest amount of time, but were encouraged to get as close to the target answer as possible. As such this metric is a measure of the participants natural behaviour under a given condition.

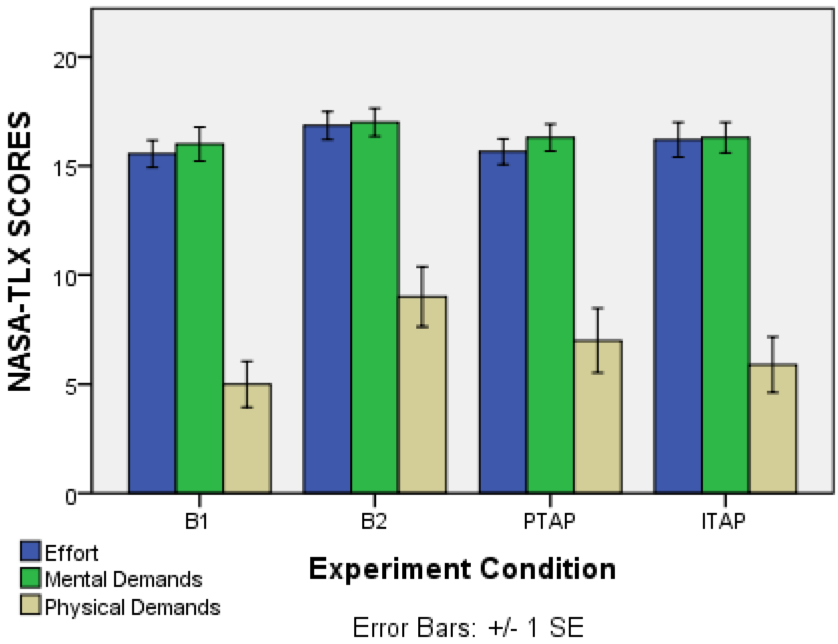


Figure 4.6: Mean Values for three sub-scales (Mental Effort, Mental Demands and Physical Demands) of NASA-TLX.

4.10 Mental Workload: Subjective measure

Participants were asked to subjectively rate their perceived mental workload using the NASA-TLX assessment tool at the end of each condition in the study.

In support of hypothesis HC, we found significant differences between conditions in the NASA-TLX scales:

- Mental Effort
- Mental Demands
- Physical Demands

Against our own intuitions, each of these measures demonstrated higher demands for B2 compared to the other conditions (see Figure 4.6). A Wilcoxon signed-rank test showed that B2 created significantly more mental effort than B1 ($Z = -2.058, p < 0.05$), and it required more mental demands ($Z = -2.292, p < 0.05$). The difference between B2 and PTAP was only $p = 0.075$ and there was no significant difference between ITAP and the other conditions. Participants also rated B2 as being physically more demanding than the other conditions (B1: $p < 0.05$, PTAP $p = 0.067$, and ITAP $p < 0.05$). This is to say that participants found the additional utterance of a nonsense word whilst solving the maths problems induced a greater physical demand than other conditions (see Figure 4.6).

Correlations between performance scales from unweighted NASA-TLX and performance data were found. This includes a negative Pearson correlation between NASA-TLX Performance scale and distance from target $r = -252, n = 80, p = 0.024$, indicating that participants were rating their performance as worse, when in fact it was better. Two positive Pearson correlations between NASA-TLX Mental Demands and Temporal Demands when compared with time to solve a problem were also found: $r = 0.340, n = 80, p = 0.002$, and $r = 0.408, n = 80, p = 0.001$ respectively.

In terms of Temporal Demands, participants reported higher temporal in ITAP compared to the Blah Baseline ($p = 0.025$), which supports the time performance results above. We found significant difference between baseline 2 (Blah) and other conditions ($p = 0.01$), with inter-condition significance of - (2-1, $p = 0.012$), (2-3, $p = 0.067$), (2-4, $p = 0.024$) when participants subjectively rated their physical burden during the task. Similarly we saw that under condition 2, mental effort was greater when compared against other conditions, significantly so between C2 and C1 ($p = 0.04$) as well as between C2 and C3 ($p = 0.075$), again indicating that the inclusion of a nonsense word increases mental workload.

4.11 Mental Workload: fNIRS

Apart from the subjective ratings (NASA-TLX scores) for mental workload measures, we used an fNIRS device, trying to sense information related to participants cognition. Further supporting HC, our analysis found a significant difference in brain region activation in both right and left inferior PFC during the experiment conditions.

As shown in Figure 4.7, OverallHbO were significantly higher during B2 compared to all other conditions (PTAP: $p < 0.05$, ITAP: $p = 0.064$). We also noted an effect on the rest time at the end of each conditions: values at rest after B2 were significantly higher than values at rest after B1 ($p = 0.05$).

Peck et al [173] found a negative correlation between fNIRS levels of Hb and the subjective ratings from NASA-TLX Mental Demands scale. Tasks that created more mental effort were accompanied by lower levels of Hb. We were unable to confirm these findings across all participants, however we found a positive correlation between performance data (distance from target) and fNIRS overall Hb, $r = 0.228, n = 80, p = 0.04$. This possibly complements Peck's correlation assuming that when mental demands are high to the point of overload, performance decreases and therefore Hb follows. This assumption ties well with the Limited Resource Model presented by Megaw [146].

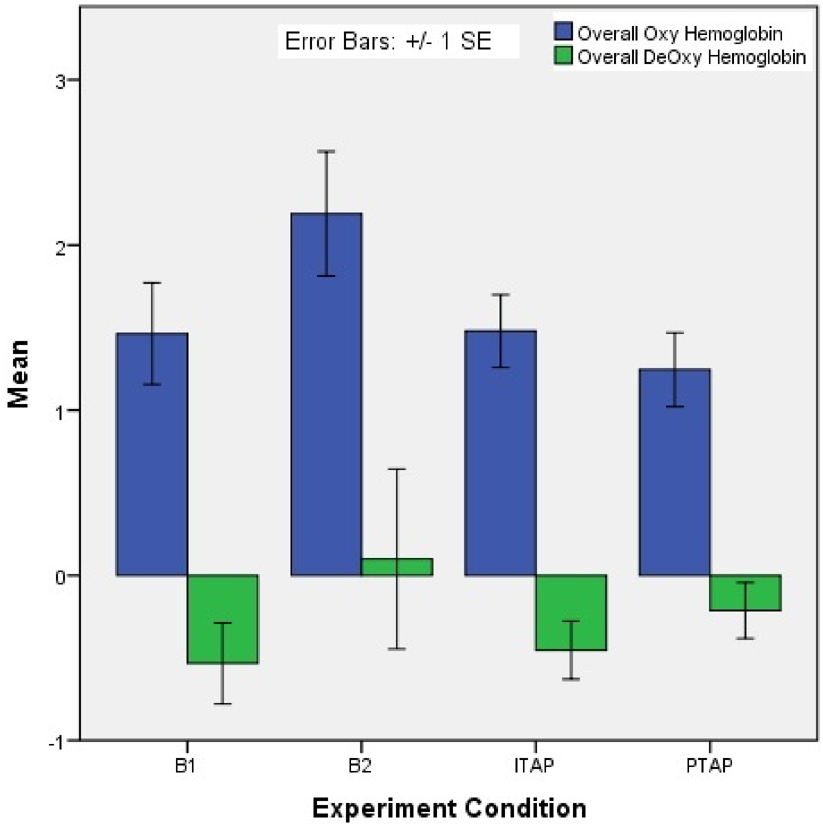


Figure 4.7: Overall HbO and Hb levels for each condition.

There was also a strong positive Pearson correlation ($r = 0.474, n = 80, p < 0.001$) between the fNIRS readings Hb left and Hb right.

4.12 Mathematical Skill

Peck et al [173] found differences in participants depending on their ability to analyse both pie and bar charts. Similarly, we believed that mathematical propensity would affect an individuals performance under differing TAPs, with the assumption that high performers would better cope with TAPs, while lower performers would be impaired as a result of reduced resources (from TAP).

4.12.1 High Performers

The high performing group rated ITAP as being more mentally demanding (Wilcoxon signed-rank test with $Z = -1.89, p = 0.059$) and requiring more mental effort (Wilcoxon signed-rank test with $Z = -1.98, p = 0.048$) when compared against PTAP. A Spearman negative correlation for the strong mathematicians between the NASA-TLX Mental

Demands scale and the fNIRS Hb levels on the right side of the PFC ($r = -0.348, n = 40, p = 0.028$) confirms Peck's [173] findings. High performers also demonstrated a positive Spearman correlation between distance from target and fNIRS Hb on the left side of the PFC ($r = 0.344, n = 40, p = 0.03$). Weighted NASA TLX score also positively correlated with time taken to solve a problem ($r = 0.399, n = 40, p = 0.01$).

4.12.2 Low Performers

For the low performing group we observed an agreement between weighted NASA-TLX score and fNIRS overall Hb. There was the same significant difference from a Wilcoxon Sign Rank test in both NASA-TLX and fNIRS ($Z = -1.78, p = 0.074$) between PTAP and ITAP. Participants workload measured with both NASA-TLX and fNIRS is marginally higher in PTAP than ITAP. This result is opposite to what was observed with the high performing group.

4.13 Using fNIRS in HCI Studies

In addition to collecting quantified data (presented above), we additionally gathered subjective feedback from participants relating to their experience of wearing the fNIRS during the study. This data was obtained, informally, at the end of the studies with the participants being prompted with the following questions:

- Did you find the fNIRS comfortable to wear?
- Did you forget that you were wearing the device?
- How long do you feel you could comfortably wear the device for?
- Would you want to wear the device, day-to-day?

Generally, participants reported that during the initial/early stages of the study, the device was reasonably comfortable and that they generally 'forgot' that they were wearing it. However, as the study progressed, a number of participants reported discomfort in wearing the fNIRS, with a few requesting a 'break' or 'adjustment' of the device between conditions (this was allowed and did not affect the validity of the data collection, since baseline activity is recorded between each condition). Few expressed interest in wearing the device day-to-day in it's current form factor.

One interesting note is that that a number of participants stated 'forgetting' the presence of the device under the task conditions, and later, between conditions, they would once again realise it's presence and discomfort in-between conditions. This indicates that the device was on the 'border' of discomfort, with the requirements of the task serving as just enough of

4.14 Discussion

4.14.1 fNIRS, Language and Mental Workload

Activations in the left side of the PFC are known to occur during semantic, relative to non-semantic, tasks that relate or involve "the generation of words to semantic cues or the classification of words or pictures into semantic categories" [62]. Due to the physical placement of our fNIRS device on participants foreheads, we can discount the interaction between Broca's area, an area of the brain with functions linked to speech production [55], as it does not fall within the reach of our device. Because fNIRS was sensitive to the B2 condition, we developed two premises (interpretations) of the results:

- fNIRS is particularly picking up the part of the brain that is activated during B2 and therefore the signal received by fNIRS is higher, or
- fNIRS is picking up an indicator related to mental workload and that B2 induces more workload. The reason behind this is the non-compatibility and non-complementarity of B2 with the mathematical reasoning task, rather than the compatibility of verbalisation protocols from PTAP and ITAP with the mathematical reasoning task.

One way to distinguish between these two is to look at the participants performance data and subjective ratings (the NASA-TLX scores) together with fNIRS. If the first premise is true, you would not expect a difference in mental workload (in the subjective scores) between the verbalisation conditions. Additionally, you would not expect any relationship between performance or NASA-TLX data with fNIRS readings. We found significant difference between verbalisation conditions in NASA-TLX scores and we also found correlations between fNIRS data with both performance and NASA-TLX. If fNIRS would pick up information related to language generation, you would expect significant difference in fNIRS data between verbalisation conditions and the silent condition (which we did not find, see Figure 4.7). With this in mind, we propose that fNIRS is not an indicator of how many words you are saying, but is sensitive to mental workload and human cognition (therefore provides support for the second premise).

Using the fNIRS alone we were unable to identify the significant differences we were expecting. However we found the fNIRS data to be complementary to existing measures such as performance and NASA-TLX. Considering the number of marginally significant results($p < 0.075$), we believe that increasing the number of participants would increase power, reduce type II error, and positively impact our findings.

4.14.2 High and Low Performers

If generalisable, our findings suggest that for high performers PTAP is the more suitable protocol and that ITAP is better suited to low performers. One possible explanation for this is that high performers have an existing procedural structure in which they operate, so interrupting this procedure (as is experienced under ITAP) potentially interferes with their natural behaviour. For low performers, however, such structure is not present and verbalising via PTAP is potentially troublesome, as they are being forced to verbalise a process that is absent or unnatural for them. The introduction of carefully chosen prompts, however, may encourage non-experts to describe how they are struggling and provide useful insight into how researchers may help these types of users in the future.

4.14.3 Running a TAP

One of our task based research questions was to investigate two think aloud protocols (namely PTAP and ITAP). The study results should be seen as a positive indicator that both TAPs do not significantly affect or influence participants ability to solve the tasks presented in the study. We used a high demand tasks and participants performance was not negatively affected in any way. Contrarily, we observed a slight improvement in participants' performance under TAP conditions, confirming with McDonald [143] that using the TAPs during the task did not have a negative influence on participants' performance.

Reflecting on Wicken's Multiple Resource Model, using multiple resources that are complementary and compatible with the task in hand might have a positive impact on performance in the case of non multi-task resource overload. Between the four conditions, participants performed the worse in Condition B2 where they had to repeatedly say 'Blah' during task solving. This was due to a higher workload generated by the condition, sensed with both fNIRS and NASA TLX subjective scale.

The TAPs conditions differed when compared between the expertise level of participants. The high performing group rated ITAP as being more mentally demanding requiring more mental effort when compared against PTAP. This result was also confirmed with the fNIRS data. Conversely for the low performing group, PTAP was the one that was more mentally demanding.

4.14.4 Using fNIRS to Measure Mental Workload

In this study, one of the aims was to evaluate the cognitive impact of various TAPs using fNIRS as a novel measurement. fNIRS was chosen for its non invasive application, portability and relative resilience to motion artefacts.

fNIRS benefits from having the properties of being both an objective and continuous measure allowing for accurate, time correlated recording during evaluation and testing studies, especially when compared to the subjective one time snapshot rating achieved via NASA-TLX. We must also note the potential negatives associated with this type of technology. fNIRS is an emerging technology and as such does not have the associated supporting research proving its correctness. Studies have correlated the measurements to those observed with fMRI [222], specifically the BOLD signal. Additionally, in the current state of technology, fNIRS can only be used to detect a level of workload (high or low), leaving a distinct lack of mapping between the readings recorded with fNIRS and the actual cognitive or emotional states. For example, detecting frustration under a evaluation study would be a useful measure, but is not currently obtainable from fNIRS.

Another point of interest, that can possibly be considered a shortcoming of this study is the exclusion of performing the study task without wearing the fNIRS device. Doing so would allow us to determine whether fNIRS affected performance or behaviour in anyway. We did ask however, as a part of the informal post study interview, whether participants felt that they were influenced in some way by wearing the device; no one reported such an effect. This does leave the potential for a follow up study to examine whether there was indeed an effect.

4.15 Conclusion

4.15.1 Study Conclusion and Contributions

Through this study we sought to:

a) to investigate how **verbalisations** might affect the use of fNIRS in increasingly ecologically valid user studies.

b) to provide insights into how different forms of verbalisations affect mental workload and performance in user studies.

c) to investigate the application of BHCI (through fNIRS) for evaluating TAPs.

In order to achieve our aims, we compared nonsense verbalisations with different forms of concurrent TAP: passive and invasive. One of our primary findings was that non-complementary verbalisations, as opposed to complex verbalisations, created higher levels of mental workload. In particular, nonsense verbalisations created higher mental workload, across measures, than Invasive TAP where participants discussed their mathematical problem solving options. Consequently, we can conclude that the use of TAPs in user studies is fine as long as the discussion uses words relating to solving the task. We saw a slight increase in MWL for Invasive TAP compared to Passive TAP, indicating that some Invasive TAP verbalisations may not have been directly conducive to solving the task. None of the nonsense verbalisations supported the task.

The findings about non-complementary language were hidden within the subjective, reflective, self-assessments included in NASA TLX; ratings had high variance, and results were only evident in some of the sub-scales. Further, we saw no difference in task performance between conditions. The objective measure obtained from the fNIRS however, provides a clear indication of the participants' mental workload whilst completing the study tasks. Because there were no differences between the silent baseline and TAP conditions, we can conclude

a) that fNIRS measurements were not largely affected by verbalisation itself
b) that fNIRS can be used to determine mental workload objectively during tasks if verbalisations remain task-related.

Overall, we provide three main contributions:

1) we provide novel insights into the underlying cause of increased mental workload created by TAPs **during** tasks
2) we provide a novel example of using fNIRS to measure cognition during a more complex task than prior work
3) we provide an example to show that fNIRS is suitable for use with tasks that involve verbalisation.

Our results make a positive step towards pro-actively using fNIRS as an BHCI evaluation tool within realistic HCI user studies. Other notable BHCI work using fNIRS for evaluation have been presented by Peck et al., who were able evaluate different types of visualisation techniques via fNIRS[173]. This work contributes to the body of evidence suggesting that fNIRS is a suitable imaging technique for conducting BHCI based evaluation.

4.15.2 Thesis Conclusion and Contributions

The primary aim of this study, in the overall narrative of this thesis, was to investigate the suitability of using fNIRS in facilitating natural forms of user interaction - in line with answering **RQ1** and exploring the **method** of applying fNIRS to BHCI studies.

In answering **RQ1**, we found that the fNIRS device we used, which is amongst the most portable and user friendly available on the market, was suited for use in HCI study settings. This is evident from our ability to ascertain clear statistical differences between different tasks directly from the fNIRS measurements. The device did, however, present a number of limitations that will ultimately interfere with the ability to develop and facilitate Natural Interactions.

The first of these was the reported discomfort that participants experienced in wearing the device for extended periods. Participants also noted that they “forgot” about the discomfort during the tasks, as a result of being engaged in the task. Therefore, it could prove significantly more distracting for less cognitively engaging tasks, which may limit the suitability of fNIRS for a number of application areas.

Second, we as the researchers designing and deploying the study noted a number of technical and physical restrictions in applying fNIRS. The fNIRS device, and all other devices at the time of the study, were wired - meaning that participants are required to remain seated and in close proximity of the data collection machine. Also, the recommendations outlined by Solovey et al. are themselves somewhat restrictive and will significantly limit the capability of developing Natural Interaction based experiences. We believe that fNIRS is well suited to BHCI evaluation and usability testing, but we advise that studies utilising fNIRS should aim to keep sessions below 1 hour in a single sitting.

In exploring the **method** of applying fNIRS to BHCI centred research, we have also contributed further to the guidelines for this particular application. In the motivating sections of this chapter we referred extensively to the guidelines produced by Solovey et al. for the considerations researchers must make when using fNIRS in a HCI setting [210]. We found the recommendations outlined by Solovey et al. to be extremely informative in both the design and deployment stages of the study. The guidelines provide the community with detailed notes and recommendations on applying the technology to their studies and provide a convenient point of reference for researchers to validate their chosen approach.

Speech, a common form of natural interaction and humanistic behaviour was not explored in Solovey et al's. original work and therefore represents a gap in the current knowledge of applying fNIRS within BHCI. In conducting this work, we have demonstrated that we were able to obtain accurate measurements of MWL, despite the presence of verbalisation during the task condition. This provides us with the necessary evidence to suggest to the community that speech is a safe behaviour during an fNIRS based BHCI user study. This is a significant contribution to our understanding of the **methods** of applying fNIRS to BHCI studies.

Solovey et al's. recommendations also provide actionable and insightful guidelines for designing and deploying fNIRS based studies, and represent a suitable basis for developing a study methodology. However, for the our intended application - using BCI in facilitating natural forms of interaction, fNIRS, in it's current form factor, is not the most suitable technology. Specifically, the technical limitations of the technology which require a participant to be relatively still and tethered (via cable) and the ecological impact of wearing the device is too significant to allow for comfortable interactions.

4.15.3 Contribution to HCI Craft Knowledge

In addition to providing significant contributions to the goals and research questions of this thesis, it is important to also acknowledge the contributions these findings provide to the craft of HCI in general. These contributions are broad in their scope, but serve as important and valuable demonstrations of the value that we can obtain from the use of a BHCI approach.

4.15.3.1 Validation of a widely used protocol

TAPs are a recognised research tool designed to elicit insight into participants thought processes and decision-making as they complete a task. The protocol has been utilised in thousands of HCI and HF studies. Identifying the validity and cognitive impact of such an important and widely used protocol is therefore a significant contribution to the craft of HCI. Our findings indicate that, the act of verbalising during a task does not significantly affect MWL, except when the verbalisations do not relate to the task at hand e.g. repeating 'blah blah blah' whilst completing a task, validating the use of TAP in user studies. This indicates that TAP are suitable for use in HCI studies. To the best of our knowledge, this is the first, direct-brain based evaluation of TAP's that has been documented in the literature.

4.15.3.2 HCI Evaluation using fNIRS

We have demonstrated that fNIRS is well suited to evaluation in HCI, through the manner that we evaluated TAP in this study. fNIRS is a reliable measure of MWL and should be considered a valuable tool in the evaluation of HCI in the future. Above, we documented some of the limitations of broader applications of fNIRS, but the technology remains well suited to conducting HCI evaluation. We believe this contribution is significant to the future of evaluation in HCI. Currently, evaluation is typically performed using a subjective approach, usually through the use of protocols similar to TAP. The introduction and detailing of a quantitative, objective measure that is derived directly from the brain of a participant, is an exciting prospect for researchers conducting this type of evaluation. The documentation of this work is therefore important in establishing study procedures and encouraging others to adopt this approach.

Chapter 5

Evaluating EEG in BHCI via Gesture based Input

This work was conducted in collaboration with the following co-investigators: Mr Donglai Pan, Dr Eugene Ch'ng and Dr Max L. Wilson.

5.1 Introduction

Continuing from the findings presented in the previous chapter, we resume our investigation into answering research question **RQ1**. Also, through this study, we evaluate various forms of **Input Control**, a common activity in HCI in order to investigate the impact of a particular form of control has over another. We further explore the **methods** of applying BHCI in the context of evaluation. We are especially focussed on preserving the ecological validity of this study, aiming to reduce the impact of using a BCI. In doing this we aim to demonstrate the application of BHCI in performing a common HCI based activity and the value that can be added to HCI in integrating a BCI device in the manner we propose with BHCI centred research.

Previously we identified that the fNIRS-100 device was a reliable and robust measure of MWL, but due to the invasive nature of the device, it is not well suited to facilitating the natural forms of interactions and significantly interferes with the ecology of the study environment. In this study, we wish to reduce this ecological impact by using less invasive BCI technology. Specifically, we attempt to answer **RQ1** by using a portable, non-invasive Electroencephalography (EEG) device.

Figure 5.1: The Muse EEG headset used in this study.

A number of affordable, lightweight and portable EEG devices were available on the consumer marketplace when we were sourcing an appropriate device for application in this study. In identifying an appropriate device, we focussed on the following properties:

- **Lightweight and Comfortable** - During natural interaction experiences, it is likely that participants will be required to wear the device for many hours, without removing it. A lightweight device with considered product design and aesthetics suited to extended use are a significant factor here. Particular consideration was paid to the 'tension' necessary to secure the device to the participants head, with gravity based solutions (those that sit on the head) being prioritised over those that required tension to remain secured to the participants head.
- **Wireless and Portable** - Another limitation we identified in our initial study was the impact of tethering the participant through the wired operation required by our fNIRS device. In selecting a suitable EEG device, we focused on solutions that were wireless, and allowed for portable usage i.e. it allowed the wearer to walk whilst using the device.

For this study, we decided upon using the the Muse, provided by InteraXon.[1] The Muse is a $250, consumer grade EEG device that is readily available for purchase online. Shown in Figure 5.1, the Muse consist of 4 dry sensors situated across the forehead, allowing us to observe the PFC - the area where MWL is believed to occur. The Muse has a built-in lithium-ion battery and uses Bluetooth to communicate with the recording device, allowing for truly mobile operation. The device lightweight, weighing just 61 grams and is constructed from a comfortable, sweat-proof silicone-rubber. For placement upon the participants forehead, the device uses slight lateral tension to secure to remain secure. No additional supporting apparatus is required to retain the device, allowing for comfortable and unobtrusive operation. We believe these properties, as identified from our initial study, make the Muse a suitable device for answering **RQ1**.

[1] Muse EEG - http://choosemuse.com

In our evaluation of fNIRS, we studied the effect of speech (via verbal protocol) has upon a participant's MWL during HCI study setting. In doing so, we were able to evaluate the impact that using TAPs in a HCI study has upon the participant, and whether it is possible that their inclusion could affect the ecology of the task - on a cognitive basis. Our primary finding was that, as-long-as the participant's verbalisations related to the task at hand, the inclusion of TAP did not significantly impact their MWL. So, in addition to exploring **RQ1**, we also validated the use of common HCI protocol via BHCI. We wanted to achieve a similar, dual contribution through the task we explore in this study also. Specifically, we want to evaluate the impact various **Input Controls** have upon the participant. As such, to evaluate the application of EEG to answer **RQ1**, we also wished to explore another form of natural interaction - gesturing.

The translation of human motion into precise action within a computer system is the primary form in which HCI is performed today. Consider our daily interactions with a (physical) mouse, whose relative positioning and button presses are translated precisely onto an on-screen (digital) cursor. As the state of technology advances, the resolution of the input device that translates man and machine interaction is also required to advance in order to fully utilise the capability of this new technology. Advances in Virtual Reality (VR), Augmented Reality (AR), speech and gestural recognition are examples of the development in technology that have advanced the state of input technology in order to fully utilise the capabilities of these new technologies.

To investigate these forms of interactions, through the application of BHCI, we used a high resolution, AR-centred input device - Leap Motion. The Leap motion is a USB hardware device that supports hand and finger motions as input into a computer system. Through this study, we apply EEG as a brain-sensing technology to quantify the MWL experienced by a participant as they complete a jigsaw based puzzle task. We developed 3 versions of the study task, each of which utilises a different form of **input control** for moving the pieces of the jigsaw puzzle:

1. **Physical** - Using one's hands in a familiar manner to solve a physical (tangible) jigsaw puzzle.
2. **Computer** - A digital version of a jigsaw puzzle is solved using a standard computer mouse.
3. **Augmented** - A digital version of a jigsaw puzzle is solved using an Augmented Reality controller (the Leap Motion controller, which we discuss below).

Specifically, the study task is designed to investigate the impact upon the participants' MWL, with special interest in understanding the effect of direct manipulation when using

3D input devices, in comparison with indirect controllers and direct manipulation in the physical world, has upon participants. Ideally, natural 3D interaction would be as close to physical interaction as possible, and be easier than manipulation of objects indirectly via a mouse. In particular, the study task seeks to compare the Leap Motion controller for completing a 3D puzzle, with control using a normal mouse, and with solving the same puzzles physically on a desk. We displayed these 3D-environment puzzles on a normal 2D display, to remove unfamiliarity with being inside a Virtual Reality environment. The physical puzzle was presented and solved on the desk at which the participant was seated. To examine Mental Workload, we utilised 2 types of measure:

1) **Objective** - For this initial work, we utilised a commercially available, affordable and non-intrusive EEG based brain monitoring device to quantify activity levels in the PFC.

2) **Subjective** - We used the NASA TLX [80] questionnaire as an industry standard for taking retrospective subjective workload measures of the task. The questionnaire served a dual purpose in this study, the first was to provide a measure of MWL between conditions and second, to provide a ground truth of the objective measure, in order to establish the accuracy of the BHCI measure utilised in this study.

In conducting this research our primary aim is to establish the suitability of using EEG brain-monitoring technology for enabling natural interactions - in line with **RQ1**. To perform this investigation we explore the application of BHCI as a tool for evaluating this new form of 3D interaction - via the use of a Leap Motion controller. With the recent developments in VR and AR technologies, it is likely that higher resolution, gestural based interaction will become a prominent form of HCI. Understanding the effects upon the user, on a cognitive level, is an input data-point upon which user experience designers can develop these types of VR/AR experiences.

5.2 Related Work

Much of 3D object interaction up to 2012 was tied to limited non-gestural controllers, like wands, which present a rather unnatural way of interacting within a 3D environment. From a Human Factors (HF) perspective, being within a realistic-looking environment yet not having the freedom of using one's hands and fingers can be quite frustrating. The principle of direct manipulation [201] involves representation of the object of interest, rapid incremental reversible actions and physical action instead of complex syntax (or confined freedom of interaction in the case of 3D object manipulation). Advances in input techniques, including the Leap Motion, opened up 3D direct manipulation applications in various domains (e.g., [15,205]), with high accuracy to the point of it being used in operating room procedures [160]. Performance and Accuracy, however, are not the only considerations for good input technologies.

Shneiderman and Plaisant [203] highlight that the MWL involved in achieving a task with a computer is made up of both the Semantic Workload required to achieve the task, and the Syntactic Workload required to use the technology. Cognitive Load Theory (CLT) [216] refers to these as intrinsic load and extrinsic load, respectively. For us, as users, this means that while direct manipulation with 3D objects, via a Leap Motion controller for example, may be more accurate, the input technology may demand significant additional Working Memory resources away from the primary task; resources that may be crucial to important tasks in operating room procedures, for example.

5.2.1 Leap Motion

The Leap Motion is positioned in the Micro Gesture Recognition domain [42]. The gesture-based, direct manipulation controller essentially removes the VR interface, providing users with the experience of manipulating 3D objects, albeit without tactile feedback. The device works by emitting infrared light (IR) t from 3 IR LEDs and tracking the reflections of this light (from the user's hands/arms) via two on-board IR cameras. A software algorithm is then able to use the reflected light data to estimate the user's hand gestures as they interact with the device. An early analysis [235] on the accuracy and robustness of the Leap Motion Controller showed a deviation between a desired 3D position and the average measured positions below 0.2 mm has been obtained for static setup and of 1.2 mm for dynamic setup. A later study also showed significant accuracy for the controller in both static and dynamic tracking [76]. Incremental updates of the SDK have since significantly improved the usability of the controller.

Various usability studies exist in evaluating the usefulness of the Leap Motion ([156,197,26,56,81,94]). Al-Razooq et al. focused on "travel" in virtual environments using the Leap Motion Controller [156], for example, found that the limited workspace of the Leap was more effective for search but less effective for path following tasks, and the accurate finger movements can contribute to the usability of continuous speed control techniques in navigation. Seixas et al. compared two selection gestures (hand grab and screen tap) for the Leap Motion controller in 2D pointing tasks (also compared with the mouse)[198]. Their results suggested that the hand grab gestures had a higher accuracy, but that the mouse outperformed the Leap Motion in gesture, movement time, and error rate.

Figure 5.2: The Leap Motion input device used in this study. Source - Amazon US

A more exhaustive study was conducted to assess the Leap Motion as compared to the mouse, for point-and-click tasks using a basic Fitts' analysis, as well as the MacKenzie et al.'s [134] seven movement accuracy measures. Results suggested that the Leap Motion is a viable device for point-and-click tasks, but inferior to the more familiar baseline device on standard Fitts' assessment measures. However, specific cursor events may be superior, with users of the Leap Motion re-entering targets less often than the mouse. The two devices showed no differences on continuous navigation paths between on-screen targets. Ehrler et al. [51] performed evaluated three different ways of interacting with the Leap Motion - one copying the traditional mouse interaction paradigm, one assigning a different hand gesture for each possible action, and one using a single gesture but allowing switching between interaction modes. The results showed a clear preference for using a limited number of gestures.

Little work so far has investigated the cognitive impact of using Leap Motion as a input modality. Adhikarla et al. designed and evaluated a direct, 3D gesture interaction using the Leap Motion, and found a strong user preference for the free hand interaction (over the 2D version) and a NASA-TLX indicated low cognitive differences between conditions [3]. Bruder et al. also employed NASA-TLX to evaluate the mental demands between 2D and 3D touch using the Leap Motion, but found no significant differences between conditions and overall scores (sub-scales were not investigated).

5.2.2 Puzzle Solving

To characterise and model the cognitive processes involved when a participant is using these alternative forms of input for the task of Puzzle Solving (specifically - jigsaw puzzles), we draw on research into Working Memory (WM). We can relate a number of processes utilised during puzzle solving to the WM model described by Baddeley [14]. For example, memorising the reference image will utilise Short-Term memory (STM). Since a Jigsaw Puzzle will involve imagery or mental rotation, it will also utilise the Visuo-spatial sketchpad. The Central Executive will mediate the information flow between STM and the Visuo-spatial sketchpad.

In addition to the WM model, we can also consider the Information Processing Model [238] and Multiple Resource Model (MRM) [237] proposed by Wickens. Wickens describes that necessary resources are limited and aims to illustrate how elements of the human information processing system such as attention, perception, memory, and decision making. Wickens describes three different 'stages' at which information is transformed: a perception stage, a processing or cognition stage, and a response stage.

The *first* stage involves perceiving information that is gathered by our senses and provide meaning and interpretation of what is being sensed. When solving a jigsaw puzzle, we are initially presented with a random assortment of pieces that we believe to constituents of the larger 'target' image. During this first stage, we may seek or obtain validation of the fact that the presented pieces are indeed constituents of the target image with the recognition of features/elements within the target image being identified in the presented pieces.

The *second* stage, which utilises WM, is where we manipulate and "think about" the perceived information. In the context of a jigsaw puzzle, we may begin to identify clusters of pieces that we believe are related and may begin to categorise pieces accordingly.

The *third* stage involves responding to the situation. At this stage, we begin connecting pieces and forming the "target" image.

In line with the separation of verbal and spatial in Baddeley's conceptualisation of WM, Wickens also separates verbal and spatial in his description of Multiple Resources. By using a jigsaw puzzle, and controlling it spatially with 3D direct manipulation, our participants will be primarily utilising spatial resources.

5.3 Experiment Design

To examine the cognitive demands involved with interacting with objects in a 3D environment, we studied participants using a Leap Motion[2](LM) - a commercially available device designed to reproduce natural-feeling interaction in a 3D environment by directly

[2]Leap Motion - leapmotion.com

mapping the shape and movements of your hands in real time. To understand the cognitive demands associated with this interaction modality, we chose to compare it against both mouse interaction, as a more familiar interaction technology, and with interaction in the real world. These alternative conditions allowed us determine the amount of cognitive demand that was inherent in the task (solving the puzzle in the real world), the amount associated with working in a 3D environment (the mouse condition), and to estimate where replicated direct-manipulation in the 3D environment sat in reference to those points.

5.3.1 Experiment Task

In conducting this study we hope to answer **RQ1** and establish whether EEG is suited to facilitating natural forms of interactions. To explore and answer **RQ1**, we have developed a task which seeks to evaluate a new form of gesture control - Leap Motion. In order to evaluate the cognitive demands of using the LM as an input device, we needed to design a task that could facilitate this type of investigation. We developed the following criterion for the task:

1. The task must feature a significant amount of hand based control, since the LM measures relative hand positioning.
2. The task was required to engage the participant primarily at a mental level, specifically targeting structures within the pre-frontal cortex, since our brain based measure targets this area (described below).
3. The task also had to have various levels of difficulty to enable control over the primary task mental demands.
4. Performance on the task had to be measurable in order to determine the effect of the LM as an input.
5. The task had to be easily performed in both the real world and the 3D environment.

Based upon these five criteria, we decided on using a jigsaw puzzle based task. A jigsaw puzzle is a tile based puzzle which requires the interlocking and tessellation of oddly shaped pieces to form the target image. The individual pieces of the puzzle must be put into it's correct place using one's hands (*Req-1*). The task requires participants to memorise as much of the image as possible in order to recreate the image that they are presented initially. For the simple images, it is likely the participant would be able to

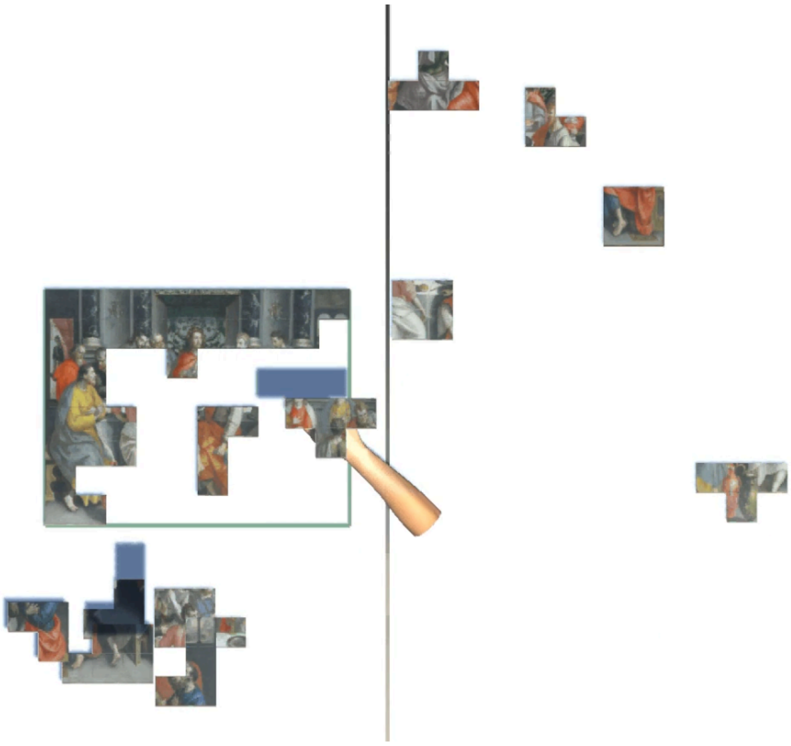


Figure 5.3: The computerised version of the jigsaw puzzle task. Participants would use this version of the puzzle during the Mouse and Leap-Motion conditions of the study.

remember the entire image. For more detailed images a variety of memorisation strategies could be employed including: identifying key constituents of an image, memorising the corners, identifying colour patterns, etc. Regardless of the strategy employed, the information will be stored in the participants WM therefore utilising STM. Additionally, the participant must perform pattern-matching and decision making to place the pieces in their respective places (*Req-2*). A number of these processes (STM and decision making) are known to occur in the PFC [207]. Task difficulty can be controlled by varying the number and size of pieces that form the puzzle (*Req-3*). Task performance can be measured by both time to complete and number of pieces in the correct position (*Req-4*). Finally, we could easily replicate a physical puzzle in a 3D environment (*Req-5*), using the Unity Game Engine,[3] shown in Figure 5.3.

5.3.2 Independent Variables

The primary independent variable was *interaction condition*, creating three major conditions:

1. Real-world physical interaction (P),
2. Mouse-interaction with the 3D environment (M), and

[3]Unity Game Engine - unity3d.com

3. Replicated direct-manipulation with the Leap-Motion in the 3D environment (L).

A second *independent variable* was task difficulty - Easy (E) with 5 pieces and Hard (H) with 20 pieces. Easy versions of the puzzle consisted of 5 pieces and were designed to be solved by all participants. Hard versions of the puzzle consisted of 20 pieces and were designed to be challenging, but achievable by most participants under all three primary conditions. This created a 3 x 2 design repeated measures design, detailed in Table 5.1.

Table 5.1: Task Conditions for this Study.

Task ID	Task Name	Description
PE	Physical Easy	A physical jigsaw puzzle, presented on thick cardboard. 5Pieces.
PH	Physical Hard	A physical jigsaw puzzle, presented on thick cardboard. 15 Pieces.
ME	Mouse Easy	Computerised jigsaw puzzle, solved using Mouse. 5 Pieces.
MH	Mouse Hard	Computerised jigsaw puzzle, solved using Mouse 15 Pieces.
LE	Leap Easy	Computerised jigsaw puzzle, solved using Leap-Motion. 5 Pieces.
LH	Leap Hard	Computerised jigsaw puzzle, solved using Leap-Motion. 15 Pieces.

5.3.3 Hypothesis

As noted in the Introduction to this study, ideal interaction in a 3D environment would feel as natural as in the real world, and would thus involve similar mental workload demands. We created two rejectable hypotheses about the expected actual differences between using the Leap Motion and the other conditions.

HP - There will be a significant difference in performance (Accuracy and Time-to-Solve) between Leap Motion and all other conditions, at their respective difficulty levels.

HC - There will be a significant difference in cognitive demand between Leap Motion and all other conditions, at their respective difficulty levels.

HP and HC were drawn from Wicken's MRM [237]. We expect that the task, the interface, and the interaction technologies should all utilise spatial resources in WM, and thus should interfere with each other without using verbal resources.

5.4 Measurements and Equipment

Below we detail the measurements collected during the study. Measurements were categorised as measuring either *Performance(**P**)* or *Cognition(**C**)*. We note, that video recordings of both the participant and their puzzle solving (physical and digital) were recorded, but not used in the analysis of this study. The data was collected with the view of exploring the idea of Cognitively Queryable Video (CQV), a concept discussed in Future Work.

5.4.1 Task Accuracy - P

Task accuracy was measured according to the number of pieces that were correctly placed in the overall puzzle, across all 6 conditions of the study.See Table 5.1 for the 6 task conditions.

5.4.2 Task Time - P

Task time was measured for each of the puzzles completed across the 6 conditions. **Note** that participants were **not** encouraged to solve the problem in the shortest possible time, in-fact, they were not made aware of the time restriction itself, unless they specifically asked - in which case they were alerted to the existence of a time restriction but that it was only in order to finish the study within a reasonable time and that they should attempt to solve the puzzle at a comfortable, natural pace. In this situation they were encouraged to continue completing the puzzle until they were stopped. The reason for this design decision was that we wished to observe the natural interactions of the participant as they would complete a puzzle normally. We justify this in because we are interested in understanding the potential application of LM in everyday tasks, so adding a time constraint may burden the participants in a manner that is not representative of everyday task completion.

5.4.3 NASA-TLX questionnaire - C

A computerised version of the NASA-TLX questionnaire was used to measure perceived workload after each of the 6 task conditions. The questionnaire identifies the weighted average ratings of six sub-scales including: Mental Demand, Physical Demand, Temporal Demand, Performance, Effort and Frustration; the second part of the NASA TLX questionnaire was also included to measure weightings i.e. the relative contribution of each type of workload to the overall perceived level of MW.

5.4.4 EEG Data - C

In collecting the EEG data for this study, we used an Android Mobile phone, with data being captured via an widely used recording application (Muse Monitor[4]). In the API provided by the device manufacturer, a convenient set of functionality for recording and streaming the data from the EEG device is available. Through the Muse Monitor application, we wirelessly recorded the data from the Muse headset at it's native sampling frequency (256Hz), but enabled additional notch filtering (at 50Hz) to account for potential power-line interference. Recordings consisted of Absolute Power of delta (1-4Hz), theta (4-8Hz), alpha (8-13Hz), beta (13-30Hz), gamma (30-44Hz) frequencies per channel, the contact quality of each channel to the participants scalp, as well as the Raw EEG power (in Microvolts, in a range of 0:~1682). The API also provides indication of events (Blinking, Jaw Clenches etc), which are also captured in the recording. One continuous EEG recording was collected for each participant across the entirety of the study. Individual condition start and end points were labelled using the "label" functionality provided by the Muse Monitor application. Each condition was split into individual files during the processing stages of the EEG using the label data.

We performed general, EEG signal processing to reduce the effects of common artefacts associated with human movement and physiology (Eye blinks/movement, myoelectric activity, verbalisation artefact, etc). We performed a high-pass, low pass and bandpass filtering of the raw-data to remove the influence of artefacts. The Raw-data was then smoothed using a Hamming Window (Window Size of 2 Seconds). Despite excluding participants with particularly poor data (this process is described below), it is inevitable that some data quality issues will remain within the chosen-samples data. To address this, we utilised the recorded contact quality indicators, and linearly interpolated values where the contact quality was poor. The interpolation process affected 0.92% of overall data. For the analysis process, we were primarily interested in the computed FFT absolute values, and where presented, statistics based upon these measures were performed upon a re-sampled (2Hz) dataset, this was performed in order to improve data-processing performance. Each individuals data was then split in accordance with each conditions relative study time. FFT values were normalised (between 0 and 1), on a per-condition basis.

[4]Muse Monitor - http://www.sonicpenguins.com/

Figure 5.4: The environment in which the study was conducted.

5.4.5 Procedure

Participants were first introduced to the task that they would be completing during the study, and informed consent was gained before the study began. Since all participants reported little-to-no prior experience using the LM, we provided an unlimited practice session where participants were able to practice with this form of input on a simplified version (1-piece) of the puzzle task. The study progressed only when the participant reported they were comfortable with the style of interaction. Once comfortable with the requirements of the task, participants were fitted with a Muse EEG headband, which was placed upon their forehead targeting the PFC. We were especially particular at this stage in the study to ensure that good contact quality was reported (via the Muse Monitor application) between the device and participant. The process of establishing good contact (as reported from the device itself) varied in time and approach between participants, with the physical shape and size of the participants head proving to be a significant factor in how easily we were able to establish a good contact quality. After this point, all data recording devices were initiated and the participants were alerted to the fact that recording had started. We began by collecting baseline reference EEG data for each participant, using an initial 2-minute Rest Condition (RC), where participants were simply instructed to relax, focus on a single point in the room and remain still.

Participants then partook in all 6 conditions, which were counterbalanced using a Latin square rotation. For each of the 6 task conditions, participants were given 15 seconds to memorise the image that they would be required to recreate in the forthcoming jigsaw puzzles. This differed from a ‘regular’ jigsaw puzzle, where an individual would have a reference throughout the solving period. We designed the study in this way in order to elicit additional MWL during the solving stage of the study. Since participants were required to memorise the reference image, they will utilise their Short-Term Memory resources more than simply having the reference image available to them. Participants were

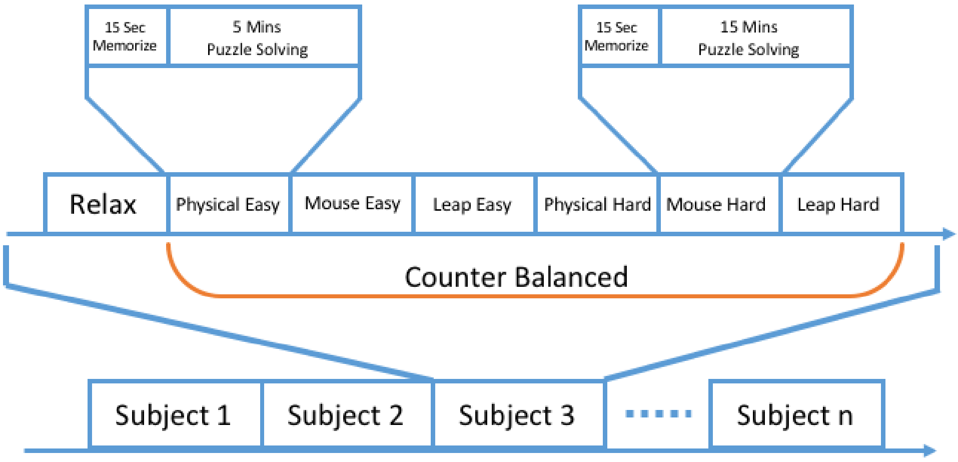


Figure 5.5: The Experimental Procedure.

verbally encouraged by the study conductor to do their best to memorise as much of the image's detail as possible, and it was explicitly stressed that they would not see the image again. Participants then had a time limit of 5 minutes to complete the easy conditions and 15 minutes to complete the hard conditions. After completing each task, or running out of time, participants filled in a computerised version of the NASA-TLX questionnaire to establish a subjective rating of their MWL during the task. Each condition concluded with a brief rest period, before continuing to the next allowing the user a brief moment of respite.

Having finished all six conditions participants were then asked to complete a brief questionnaire relating to their gaming (2D/3D) experience, previous exposure to VR/AR technologies, and their experience with puzzle solving.

The study was conducted in an office environment, with participants sat at a standard desk and chair with a 20 inch monitor placed in front of them (See Figure 5.4). Dependant on the version of the task being sat, the environment was adapted to suit the requirements of the task. For ME/MH, a standard desktop mouse was placed on the desk, whereas for LE/LH, a leap motion controller was placed in front of the participant; the desk was cleared for PE/PH conditions to make room for the physical puzzle.

We note that the environment was representative of a typical office desk, and would be representative of somewhere where an individual may chose to complete a jigsaw puzzle in 'real' life. In accordance with our interest in the **methods** of applying BHCI, the environment was designed to be familiar to the participant, and little instruction were giving regarding restriction of movement or considerations particular to the EEG they were wearing.

5.4.6 Participants

To ensure we achieved our target power (0.8) we performed priori-analysis to identify the minimum number of participants necessary to achieve this power; to achieve a significance value of 0.005 required 17.3 participants. Twenty-six participants were recruited to take part in the study, but subsequent analysis of the collected EEG data quality led us to exclude six participants. EEG data was analysed using the sensor quality report from the device. From this contact quality indicator, we set a level of 85% of 'good' contact during the experiment conditions period. The six rejected participant's data did not meet this specified level due to various reasons, mostly relating to the size and shape of their head, e.g. one participant had a particularly small head which led to the sensor slipping down the forehead during the course of the study. From here-on-in, all results presented were collected from twenty participants (15 Male, 5 Female) with an average age of 30.3 (SD = 7.6). All participants had normal or corrected vision and reported no history of head trauma or brain damage. The study was approved by the school's Ethical Committee which applied the ethics evaluation procedure specified by the University of Nottingham (UK). Participants provided informed consent and received no compensation for their participation.

5.5 Results

We began by checking for an ordering effect. To do this, we normalised the easy and hard conditions onto the same scale, by dividing the number of correctly placed pieces and the time taken, by the total number of pieces (E=5, H=20) and maximum available time given (E=5-mins, H=15-mins). Overall, there was a slight, but not significant, increase in performance (accuracy and time-to-solve) as the participants progressed through the study tasks. A one-way repeated-measures ANOVA confirmed that there was no significant difference for normalised task performance (Accuracy - ($F(5, 114) = 0.351, P = 0.881$) and normalised timing - ($F(5, 114) = 0.407, P = 0.843$). We were also keen to identify whether there was a learning affect within the input-type condition, as to whether e.g. the second time using the LM was better than the first. Again using normalised accuracy and times, Table 5.2 shows these differences. The biggest differences were in the higher accuracy on the second experienced LM condition, and in faster performance times for the mouse. These were not significant, however, which implies that participants did not improve significantly with experience with each input type during our study.

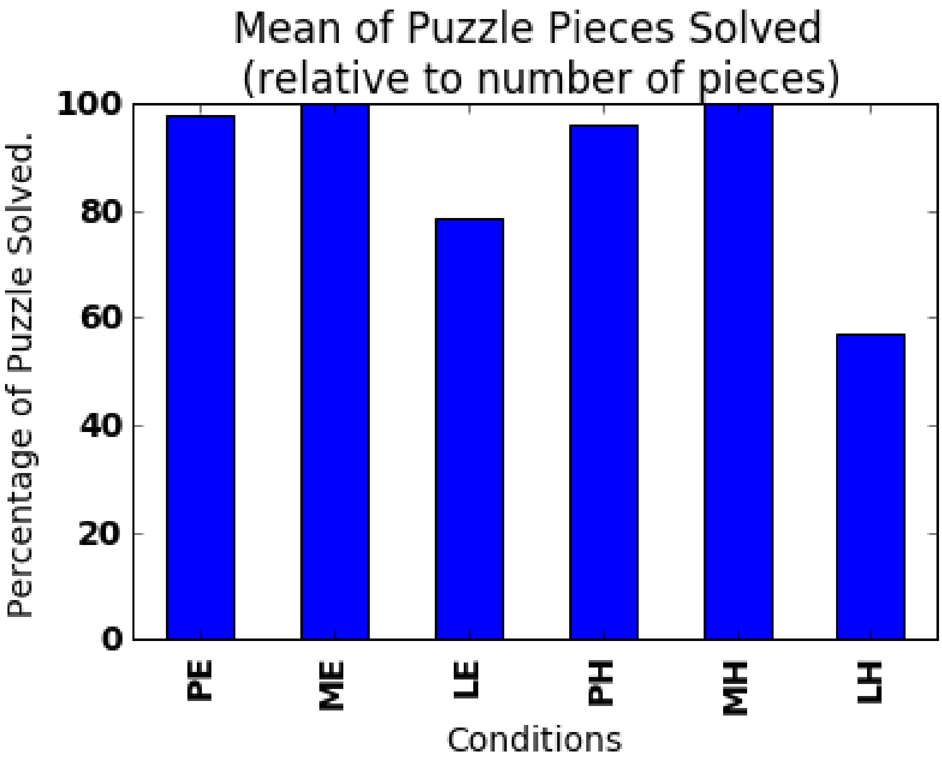


Figure 5.6: Overall participant accuracy for each Condition.

Table 5.2: Weighted Improvement between Trial 1 and 2 of each Task Condition.

Input	Accuracy Improvement	Timing Improvement
Leap-Motion	7%	1% (Faster)
Mouse	0%	7% (Faster)
Physical	3%	0%

5.5.1 Performance

As noted in Experiment Design, we examined performance using both Accuracy (number of pieces in their correct position) and Speed (how long to complete the puzzle. If incomplete, the maximum time limit for the condition was the resulting time: Easy - 5 mins, Hard - 15 mins). We present both measures here, but reiterate that participants were not encouraged to solve the puzzles in the least amount of time, in-fact many did not discover that there was indeed a time limit, with the exception of Condition LH where few managed to complete the puzzle fully in the prescribed time limit.

5.5.1.1 Accuracy

A two-way ANOVA first revealed that there was a significant difference in task accuracy between Easy and Hard conditions ($F(1, 150) = 1046.78, P < 0.001$). In support of our hypothesis HP, we also found a significant difference in accuracy between input conditions ($F(2, 150) = 66.56, P < 0.001$), as shown in Figure 5.6. The interaction between input type and difficulty was also significant ($F(2, 150) = 40.43, P < 0.001$), which implies that the difference in input techniques was exaggerated by the difficulty of the task. A post-hoc Tukey test showed significant differences between Leap Motion and all other conditions (both $P < 0.01$), but not between Mouse and Physical conditions. We conclude from these that performance was worst in Leap Motion conditions, and this decrease was exacerbated by task difficulty.

5.5.1.2 Speed

A two-way ANOVA revealed a significant difference between Easy and Hard conditions for time to solve ($F(1, 150) = 378.08, P < 0.001$). In support of hypothesis HP, we also found a significant difference in time-to-complete between task conditions ($F(2, 150) = 133.41, P < 0.001$), as shown in Figure 5.7. Again, there was an interaction between these two variables ($F(2, 150 = 31.98, P < 0.001$), implying that increased time to solve was exaggerated by task difficulty. A post-hoc Tukey test showed significant differences between all version of the input types (each with $P < 0.01$), implying that Mouse was faster than physical puzzle solving, which in turn was solved faster than in Leap conditions. We note here that the automatic 'snap-to-grid' functionality in the 3D environment was likely the cause of improved performance time for mouse over the physical condition. This also provides us with a possible explanation for the discrepancy between the numbers of solved PE vs ME tasks. That is, everybody solved all ME tasks but not everyone solved all PE tasks.

5.5.2 Cognitive Measures

Below we analyse the cognitive demands experienced by participants, first through NASA-TLX data, as an empirically validated industry standard measure, and then examine whether these differences were observable in EEG data from the commercially available Muse sensor.

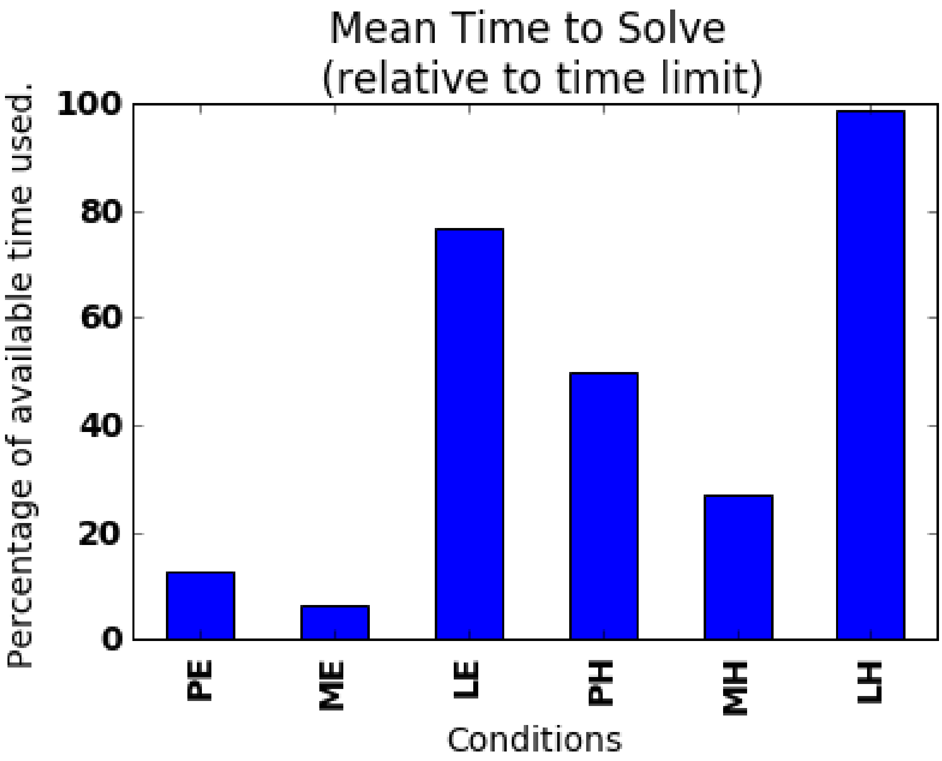


Figure 5.7: Overall Time-to-Solve in each Condition, relative to the available time.

5.5.2.1 NASA-TLX

Figure 5.8 shows the overall, weighted NASA-TLX workload score for each condition. In keeping with our Hypothesis HC, we see that both versions of the Leap Motion condition elicited the highest amount of MWL.

As expected, we saw a significant difference ($Z = 5.4133, P < 0.001$) between Easy and Hard conditions in Weighted Overall Workload, from NASA-TLX. A Friedman test comparing the three input conditions in the Easy task setting found a significant difference ($\chi^2(2) = 30.87, P < 0.001$), with pairwise comparisons revealing significant differences between Leap and the two other conditions (both $p < 0.001$), but not between Mouse and Physical. Similarly, the same comparison in the Hard task setting found a significant difference ($\chi^2(2) = 12.4, p < 0.005$); Leap created significantly more Overall Workload than Mouse ($p < 0.05$) and Physical ($p < 0.01$), but not between Mouse and Physical. More specifically, in our comparison of LH to the baseline of PH, we also saw significantly more Mental Demands ($Z = 11, p < 0.01$), Physical Demand ($Z = 0, P < 0.01$) and Frustration ($Z = 6, P < 0.01$) in LH.

In relation to our performance data, we found that accuracy ($r_S = -0.63, p < 0.01$) and time-to-solve ($r_S = 0.78, p < 0.01$) correlated strongly with Weighted Overall Workload from NASA-TLX. Further, time-to-solve correlated strongly with the Effort sub-scale

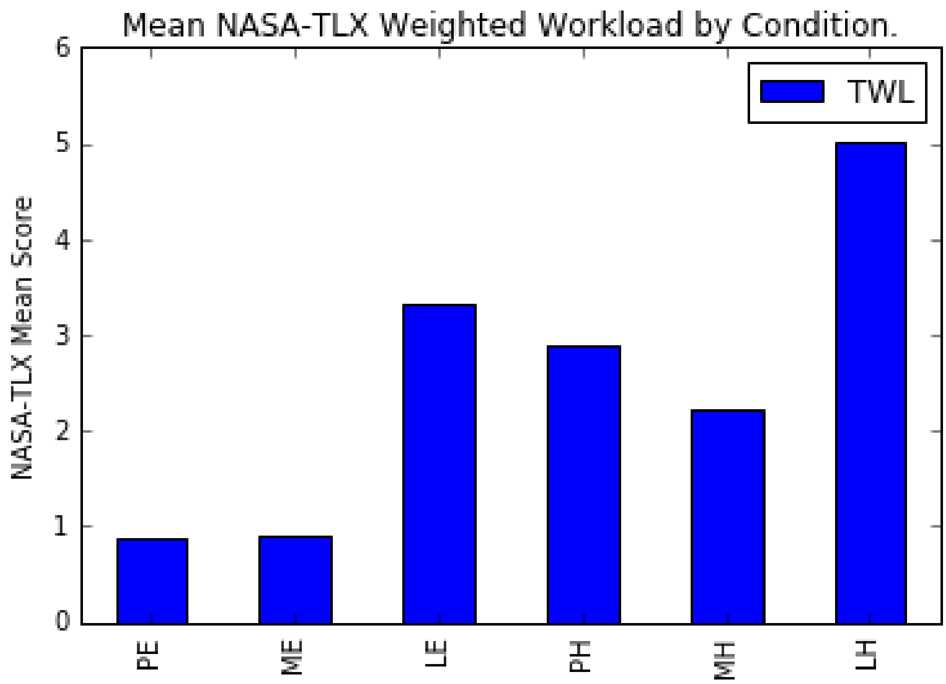


Figure 5.8: NASA-TLX weighted overall Workload Score.

($r_S = 0.63, p < 0.01$), the Frustration sub-scale ($r_S = 0.65, p < 0.01$), and Physical Demands ($r_S = 0.62, p < 0.01$).

5.5.2.2 EEG

The Muse EEG device, used in our study, has four sensors. Our analysis for this study primarily focused on sensors AF7 and AF8 from the Muse EEG headset, since these target the left and right sizes of the PFC, respectively. To quantify MWL, we use the ratio of Theta (4-7 Hz) to Alpha waves (8-13 Hz), a relationship identified by Klimesch, shown to reflect cognitive and memory performance [112]. Klimesch's ratio was chosen thanks to the extensive amount of literature supporting the relationship identified by Klimesch [45] and it's application in HF studies [186].

We were first interested to know whether data from the Muse, a commercialised BCI sensor, correlated with the NASA TLX data, as an empirically validated industry standard for evaluating workload. Overall, we found no correlations between:

a) overall mean across the four channels, nor
b) with AF7 only, with NASA TLX overall scores and sub-scales.

AF8, however, correlated strongly with the overall calculation of NASA workload ($r = 0.36, p = 0.05$), with average MWL levels from this sensor visualised in Figure 5.9.

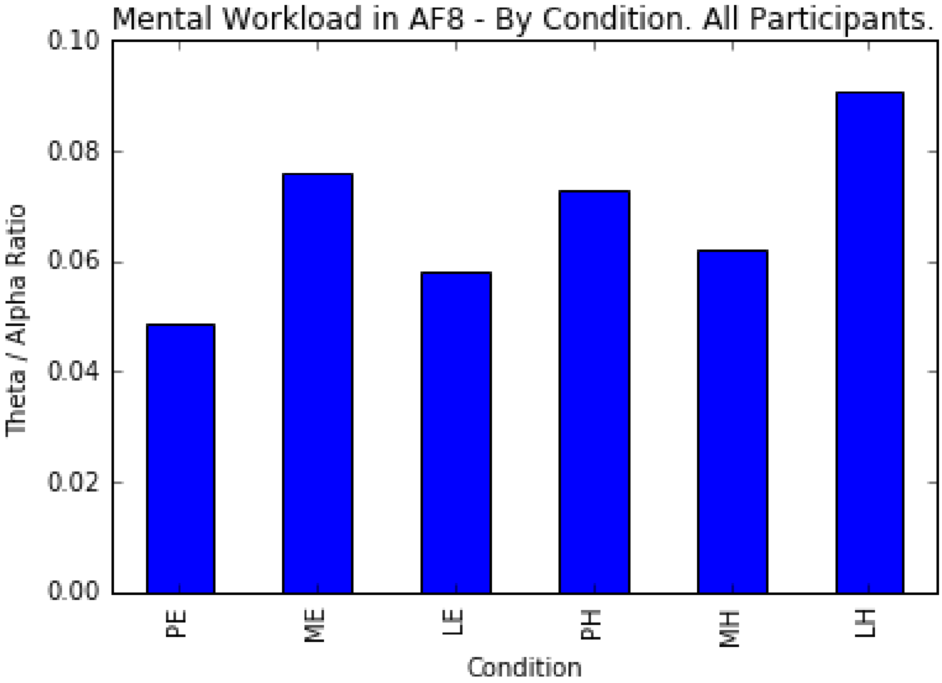


Figure 5.9: Mental Workload in AF8 - By Condition.

We also found that AF8 correlated strongly with 3 particular sub-scales of the NASA workload measure (see the detailing of NASA-TLX in the Literature Review), specifically: Frustration ($r_S = 0.39, p < 0.05$), Mental Demands ($r_S = 0.41, p < 0.05$), and Temporal Demands ($r_S = 0.37p = 0.04$).

To see whether the Muse could determine the difference in MWL between conditions, shown in Figure 5.9, we applied a two-way ANOVA. We examined both overall data from all four channels, and AF8 only. No significant differences, however, were found for either the difficulty or input-type variables. The only significant difference found was in the extreme condition, between Leap-Easy and Leap-Hard ($t(36.99) = 2.0313, p < 0.05$). This implies that data from the Muse is not as sensitive as NASA-TLX, as an industry standard, and would require deeper and richer analyses in order to detect smaller differences in MWL.

5.5.3 Post-hoc Analysis of Performance

In performing the study, we identified several variations in peoples experiences - despite no participants having notable prior experience with the Leap-Motion, some participants performed much better or found it easier than other participants. We began this analysis by considering overall participant performance, using the formula: $Total\ Time\ To\ Solve/Total\ Pieces\ Solved$. We then took the top and bottom quartile for respective high and low performers. We also analysed participants who

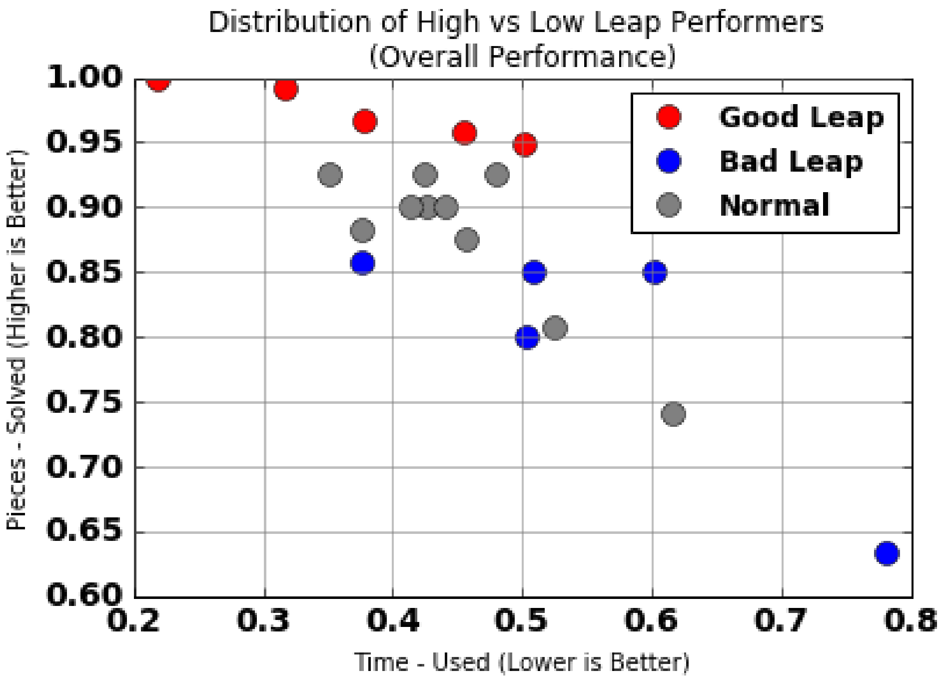


Figure 5.10: Distribution of Participant's performance, highlighting the effect of Leap Condition Performance.

performed particularly well in the Leap-Motion condition. Perhaps interestingly, as shown in Figure 5.10, the top quartile of Leap-condition performers was the exact same five participants as those identified by overall performance.

A two-way ANOVA, comparing input condition and high/low performers found a significant difference in overall weighted NASA-TLX Workload score ($F(1, 54) = 13.80, P < 0.01$). This confirms that high performing participants experienced significantly lower overall workload than low performing participants. As with our overall statistics, there was a significant difference between input modalities ($F(2, 54) = 14.09, P < 0.01$), but as there was no significant interaction between these variables, we find that high performance did not affect the difference between input modalities. Using *overall* workload from the NASA-TLX, however, we did not find that high performers found leap easier, or more comparable to our other conditions.

We continued this analysis by investigating the sub-scales of NASA-TLX, and found that high performers experienced lower cognitive demands in some factors of workload. Higher performers reported significantly less effort in the Leap Condition ($p < 0.05$) than low performers. When comparing Leap with Mouse, we found that, while low performers gave significantly different (all $p < 0.05$) ratings for all sub-scales, high performers did not report significant differences for Effort, Temporal Demand, and Performance. This indicates that high performers found the Leap Motion interaction comparable to the Mouse interaction for three sub-scales, while low performers did not. Further, comparing Leap with the Physical condition, high performers did not report a significant different for Temporal Demand either, indicating that they found all three conditions comparable in temporal demand, while low-performers did not. This could mean that, for low performers, the mental workload associated with these interacting in the 3D environment lowered their capacity to remember the design of the puzzle they were trying to recreate.

5.6 Discussion

Overall, and in support of our study hypotheses, we found that participants *performed* less well in the Leap Motion condition (in terms of accuracy) and experienced high levels of Mental Workload (as measured by NASA-TLX). In accordance to the Limited Resources Model, we would expect performance to drop as Mental Workload increases towards capacity, implying that the Leap Motion brought participants towards this limit.

This study contributes to our thematic aim of evaluating different types of **input control** and contributes further to our understanding of the **methods** of applying BHCI.

Despite us being unable to directly attribute differences between types of **input control** directly from the EEG device itself, the study nevertheless was able to provide interesting findings and insights. These results indicate that although the Leap Motion provides a replicated form of direct manipulation in the 3D environment, the experience is far less natural than real-world physical interaction, and less easy to control, for participants, than the mouse.

5.6.1 Possible Confounding Variables

We suggest that there may be four aspects that affect this distance between desired and actual experience with the Leap Motion: haptic feedback, separation of input and output, familiarity with the input technologies, and familiarity with the output technologies.

5.6.1.1 Haptic feedback

A key difference between the Leap Motion and the other conditions is feedback through the sense of touch; where physical interaction represents the ground truth in this case. The Leap Motion involves participants grabbing thin air, and depending on the visual senses to determine as to whether their hand-actions are being successful. With the mouse, however, although the sense of touch did not directly relate to touching pieces of puzzles, the participants hands were controlling a physical object, within a limited 2D plane. It may be interesting in future work to evaluate haptic feedback technologies for 3D direct manipulation, like tactile gloves[5] [166] and mid-air haptics [33], reduce the Mental Workload experienced with 3D interaction.

5.6.1.2 Separation of input and output

In this study, we kept the display of the 3D jigsaw puzzle within a 2D display, so that experience in a Virtual Reality (VR) environment did not affect our results. This means that, although participants hands were replicated within the 3D environment, this was a duplication of their physical hands that were also in their field of vision. Future work could also investigate as to whether interaction via a Leap Motion using a VR headset for output feels more natural and demands significantly less Mental Workload than on a 2D display.

5.6.1.3 Input Familiarity

One possible explanatory factor for the difference between indirect control with the mouse and replicated direct manipulation with the Leap Motion, is familiarity. Interaction via a mouse, for example, has become an autonomous skill for many people, which could explain the lack of cognitive difference between physical interaction and indirect mouse interaction. Likewise, although participants did not significantly improve between their two experiences with the Leap Motion, the overall lack of experience with the Leap Motion could explain the significant differences found between it and the other input conditions.

If we were to conduct this study again, we would explore the application of a longitudinal study, introducing training periods for the LM. In doing so we would reduce the issue of input familiarity, as described here. We would also observe changes in participants performance and EEG data before and after the training period. This comparison could provide an interesting narrative for the differences between differing levels of expertise and would allow us to explore it's impact upon an individuals EEG data.

5.6.1.4 Output Familiarity

As part of our study, we asked people to report their gaming experience, as a form of estimating their familiarity with interacting within 3D environments on a 2D display. Participants rated their 3D gaming experience on a six part Likert scale (1- No Experience, 6-Very Experienced) after completing the study. Figure 5.11 shows that experienced

[5] Manus Tactile Gloves https://manus-vr.com/

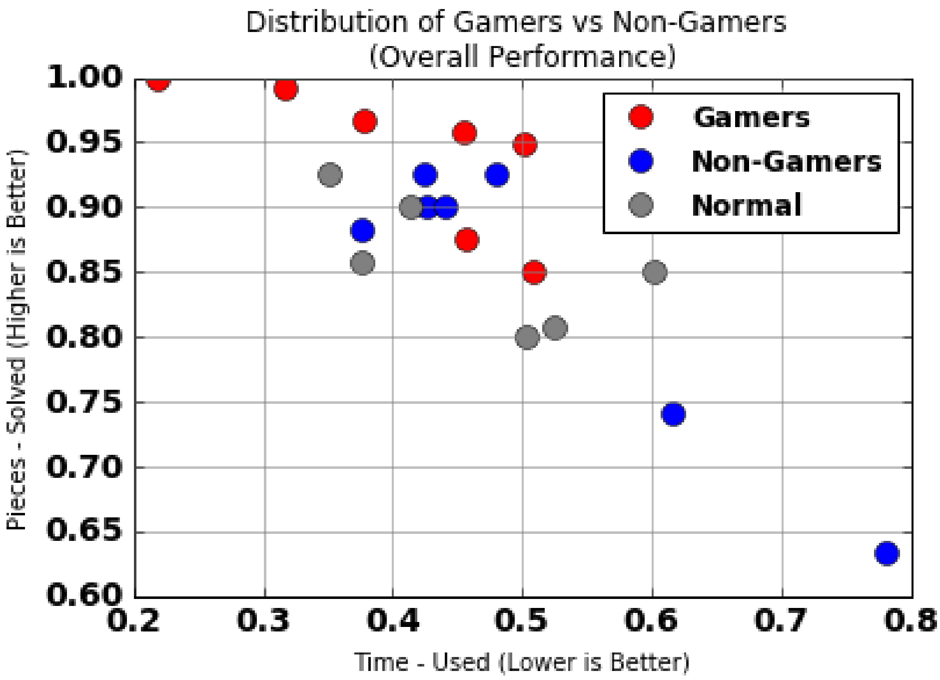


Figure 5.11: Distribution of Participant's performance, highlighting the effect of expertise.

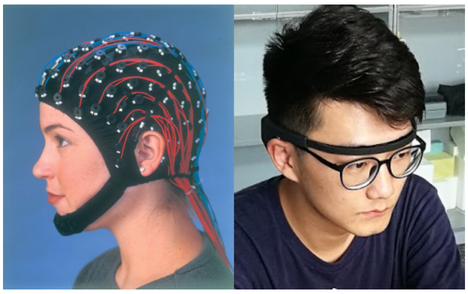


Figure 5.12: Full-Head EEG device compared to the Muse. Full-head image provided by WUSTI

gamers largely overlapped with the top quartile of high performers, and inexperienced gamers were spread across the lower quartiles. Perhaps interestingly, although we found that gaming experience did not significantly reduce any of the cognitive measures, their performance was significantly better than non-gamers ($F(2, 102) = 4.841, P < 0.01$).

5.6.2 Applying EEG to Natural Interaction

The primary objective in conducting the research presented in this chapter was to investigate the application of a non-invasive, direct brain monitoring device in a HCI setting, in answering **RQ1**. We used a consumer grade, commercially available EEG device (Muse EEG), to measure participant's mental activity as they completed the tasks presented in this study. Traditional applications of EEG utilise a medical grade, 'full-head' device, such as that shown in Figure 5.12. Medical grade EEG devices depend on the application of saline solution to ensure conductivity between the sensors and the participant's scalp - a process that involves significant setup time [244]. In contrast, the Muse headset was simple to set up, requiring no saline solution and the intuitive design of the headset meant little re-positioning or adjustments once in place. To capture data, the Muse connects wirelessly over Bluetooth to the researchers mobile device running a recording application, whereas specialised data recording hardware is required for medical grade devices. The deployment therefore of the Muse EEG (and consumer grade devices in general), was straightforward - requiring little preparation, setup time and introduced little discomfort to the participant.

Ecological validity is an important factor in conducting HCI research and enabling natural forms of interaction, and was a stated interested of our investigation into the **methods** of applying BHCI. On this front, the study design was successful. Preserving the context of the ecology in which a study task is based is critical in identifying and qualifying value in the results of a HCI study [136]. From the post-study questionnaire and informal discussions with participants regarding their experience with the device, we learnt that while participants were aware of the device's presence at times of inactivity, they generally 'forgot' that they were wearing it.

> ''I forgot it was on most of the time" - **P9**

Some participants noted the presence of the device in the periods between conditions, but did not note any discomfort:

> ''I didn't notice the device when I was completing the task, but during periods in-between I began to notice it". - **P12**

It is possible that the Muse and other consumer grade EEG headsets, could be suitable for extended periods (> 1 Hour) of monitoring in BHCI style studies. This finding adds to the growing body of literature [219,7] that suggests portable, non-invasive are suitable for integration within a BHCI style user study and for facilitating natural forms of interaction **RQ1**. This is an important contribution, for establishing both the validity of applying BHCI research and for validating the suitability of using lightweight forms of EEG in natural interactions.

As researchers, we noted that the device did not interfere with how we designed or deployed the study and we did not make particular considerations for including the device in our study. We believe, in particular, the ability to record data from the Muse via a mobile device is crucial in enabling interaction designers to develop comfortable engaging forms of interaction - without the restrictions we encountered in our initial study.

The obvious limitation however in using EEG in this context is the quality of the data we receive from the device. The results of our study indicate that despite the task eliciting varying levels of MWL, the Muse was not able to record these variations in a statistically significant manner. A number of factors were likely contributors to this issue:

- **Uncontrolled Study Conditions** - EEG, as we know, is prone to a number of motion derived artefacts: limb movement, speech, eye movement/blinking etc. However, in conducting this study we wanted to explore the application of EEG in quantifying natural user interactions - so a number of these potential artefacts were intentionally left uncontrolled during the study task. This indicates that EEG might not be the most suitable technology in BHCI applications that require precise, reliable and consistent brain activity measures, but may be suitable for applications that require general indication of cognitive state, especially derived from more complex analysis (discussed below).
- **Lack of existing works** - The Muse is a relatively new device to the marketplace and there is little supporting literature documenting it's use for quantifying MWL. Some existing work identified slight statistical differences using simple averaging measures of the wave forms (Alpha, Beta, Theta, etc) provided by the Muse [1], but no clear work linking MWL to data from the Muse currently exists. The results of this study indicate, however, that the Muse might not be sensitive enough to capture these variations in MWL, despite using a recognised measure (Alpha-Theta ratio - [112]) for quantifying MWL from EEG data.
- **Data Analysis** - We must consider the possibility that the Muse EEG device was, in fact, capable of distinguishing between varying levels of MWL, and that calculation of MWL i.e. our analysis, was inappropriate or too simplistic in this instance. One of our stated future works is to investigate the application of more advanced forms of data analysis, such as machine learning to establish deeper insights and meaning into the data we obtain from these devices.

For our stated research question, **RQ1**, we believe that this study demonstrates the suitability of using EEG in facilitating these forms of natural interaction. We believe that despite our inability to distinguish between variations in MWL, we were able to successfully utilise brain-sensing technology whilst preserving the kinds of ecological validity that enables natural forms of interaction.

5.7 Conclusion

5.7.1 Study Conclusion and Contributions

Although existing research has demonstrated the accuracy and potential value of technologies like the Leap Motion for control in 3D environments, this work has specifically evaluated the *cognitive demands* of such technologies - in these environments using commercially available EEG. As we have limited cognitive resources, Mental Workload created by difficulty in using an input technology reduces the amount of resources we have available for the primary task, which can then quickly impact performance. We hypothesised that participants' performance would be reduced, and Mental Workload increased, when interacting via the Leap Motion, in comparison to interaction with a mouse and physical interaction in the real world. Our measures of performance and Mental Workload confirmed these hypotheses, but our EEG measure was unable to mirror these findings. We found, however, that high-performing participants did not experience the significant cognitive differences between input conditions reported by low-performing participants. Overall, the study findings contribute a critical evaluation, grounded in empirically validated models of Mental Workload, of the cognitive demands of 3D interaction techniques and the application of EEG in a HCI study. The significant differences in our results highlight that, despite having high accuracy, such technologies are still a long way from replicating natural real-world physical interaction.

5.7.2 Thesis Conclusion and Contributions

The primary aim of this study, in the overall narrative of this thesis, was to investigate the suitability of using a consumer grade EEG device in facilitating natural forms of user interaction - in line with answering **RQ1**.

From the discussion above we can establish that lightweight forms of EEG are well suited to enabling extended forms of portable and non-invasive use - key properties for enabling natural forms of interactions. Despite the failure to accurately measure MWL through it's application in this study, we note that EEG provides a variety of other brain based measures that have been shown to be accurate and detectable using consumer grade technology. A number of devices on the market (including the Muse) provide pre-classified levels of psychophysiological indicators such as: Attention, Meditation and Stress. As discussed in our literature review section on EEG, there are relationships between these EEG wave (Alpha, Beta, Delta, etc) that are known to relate to certain physiological states e.g. Attention, Meditation, Frustration etc. The problem however, at least with the Muse EEG device used in this study, is that the precise detailing of pre-classified measures are not explicitly documented.

We believe that the form-factor and portability of consumer grade EEG holds greater promise in fulfilling the remaining research questions (**RQ2,3 & 4**). Despite being a potentially poor measure of MWL (given our current knowledge and approach), it nevertheless lends itself to natural interactions in a way that we are unable to attain with alternative devices (in the current market). In this capacity, we will use this technology to develop a natural interaction experience using this brain-sensing technology in the investigation of these research questions.

Additionally, we sought to evaluate different forms of **Input Control** and their impact upon participants. We have demonstrated through the results of this study that Leap Motion provides an analogy to more familiar forms of interactions such as mouse or physical. However, this type of input control has a significant effect upon the MWL of participants, as measured from the NASA-TLX questionnaire. This raises questions as to the viability and practicality of using the LM as a form of input control in future applications.

5.7.3 Contribution to HCI craft knowledge

5.7.3.1 Evaluation of an Novel Interaction Technique

In our motivation for studying the LM using a BHCI technique, we detailed various studies that have examined the impact of using the LM as an input device. Al-Razooq et al. for example found LM was suited to search tasks but not path following tasks [156], whereas Seixas et al. explored the performance impact of using LM in 2D pointing tasks [198]. But, to the best of our knowledge, the physiological and MWL impact of using a LM has not been explored in the literature. In exploring this work we highlight the importance of understanding new interactions techniques on a multitude of levels and the insight this can provide to the community. We have shown that LM induces additional MWL by utilising additional mental resources that are not utilised in more traditional/familiar forms of interaction e.g. Mouse and Keyboard. We acknowledge however that there could be an associated learning affect here and a longitudinal study would be of significant contribution to our understanding of long-term LM based interactions in participants. Nevertheless, our current findings do inform how new users of LM are likely to experience the interaction on a cognitive level, something interaction designers should consider when developing an interface using the LM.

5.7.3.2 Novel Study Task

During the design stages of this study, we realised quite early on that developing a task for studying the LM in a BHCI context was somewhat challenging. Given the number of constraints we described, specifically focussed on being able to vary task difficulty and have analogous tasks in both a digital and physical environment was challenging. We identified a number of works exploring the paradigm of digital versus physical documents, and tasks centred around organisation and information finding [214]. These tasks however did not provide an intuitive way of manipulating task difficult in the manner we have achieved with our study task. We believe the task and study design we present here provides the HCI community with an approach for studying such interactions via a task that has controllable difficult and is suited to LM based interactions.

Chapter 6

#Scanners: The Application of BHCI for Interactive Cinematic Experiences

6.1 Introduction

Through the body of work presented thus far in investigating **RQ1**, **Methods** of applying BHCI and our investigation in different forms of **Input Control**, we have identified a suitable technology for enabling natural forms of interaction using brain-sensing technologies. Specifically we have found lightweight, consumer grade EEG devices to be especially well suited for this use case.

Having identified the properties of a suitable technology, we must now seek insight into the three remaining research questions:

> **RQ2**. How can BHCI be used to develop natural forms of indirect control?
>
> **RQ3**. How are these natural forms of indirect control experienced by the users?
>
> **RQ4**. What design considerations must we make when developing indirect natural interactions using BHCI?

To explore these remaining research questions and our final theme of **Novel Interactions**, we develop an interactive cinematic experience which we use to study the effects of utilising BHCI in this setting. Specifically, we develop, deploy and analyse a novel cinematic experience in which the composition of the film is informed, in real-time, by the viewers physiology. In pursuing this form of **Novel Interaction**, we gain significant insight into answering the remaining research questions. We implemented a form of novel form of **Novel Interaction** where the viewer affects and is affected by the experience presented to them. We present this as one way of developing interaction using BHCI - **RQ2**. In observing and understanding the effect of this interaction upon the viewers, we develop insight into how the experience was perceived - **RQ3**. Finally, we will synthesis these findings and provide a concise resource upon which others will be able to develop future works - **RQ4**.

6.1.1 Motivation

It is a central tenet of usability that the locus of control shall remain with the user, to the extent that this is one of Shneiderman's eight golden rules [202]. Direct Manipulation, as a core principle that underpinned the design of graphical user interfaces of HCI, emphasises that, wherever possible, a user should be able to use an interaction that directly maps to what they are trying to achieve [204]. Using a mouse to move an object, an example of direct-control, provides more direct manipulation over the object than entering desired x and y locations into text boxes.

A great deal of BCI research has focussed on extending this paradigm of direct control into BCI application. Advances in our understanding of the brain and suitable monitoring devices have enabled us to exert control over a system (both physical and virtual) without requiring a physical interaction. This has proven particularly beneficial to individuals with significant motor impairments, who's quality of life is significantly impacted by our current approach to system control [125,245,89].

One interesting thread of recent work however has focused on forms of partial control that lie somewhere between direct and indirect. Nagashima, for example, investigated the use of bio-sensing techniques within media arts [158], whilst many others have used bio-feedback to control gaming, such as using heart rate in fitness games [140,211], relaxation within a shooting game [159], and breathing rate to control the speed of amusement rides [139]. Marshall et al. breathing-controlled bucking bronco [139] directly utilised this limited amount of control; riders had to try to overcome their autonomically-controlled increasing breathing rate associated with the thrilling experience. Their paper further discusses the importance of surrendering control, a topic that also discussed in other areas, such as auto-piloted vehicles [123] and home automation for those observing the Jewish Sabbath [246].

Höök called this interplay between physiological response and physiological control an affect loop [92], in which a user is affected by a system, which is affected by the user, and so on. In this work we're interesting in exploring the application of this affect loop in the context of a cinematic experience.

Research that could broadly be categorised as BHCI, is increasingly being adopted by the entertainment industry [25] both as a tool to understand people's emotional experience [77] and now as a way of controlling emotionally engaging experiences [61,65]. While entertainment may benefit from a BHCI approach, it could also contribute to its development, offering ways of engaging mainstream audiences with the technology and contributing to our understanding of what can and cannot be controlled and how to design for the benefits and limitations of the new technologies. Against this broad backdrop, we report an exploration of using BHCI to create an interactive entertainment experience. Following an exploratory 'Performance-led Research in the Wild' methodology [19], we worked with the artist and film producer, Richard Ramchurn[1], to design, develop and deliver a public experience, #Scanners, a brain adapted film. #Scanners was screened and studied in-situ, qualitatively, in order to reveal wider issues and principles.

6.2 TWAL: Two Way Affect Loops

In investigating **RQ2**, we propose a novel form of natural interaction that we name as Two Way Affect Loop (**2WAL**). A users interaction with a film typically involves a One Way Affect (**1WAL**), in which the film being viewed has an effect on the viewer. Through the application of BHCI however, we can begin to explore 2WAL, an cinematic experience in which the presented media is dynamically affected by the viewers' physiology or behaviour (as-well as the viewer continuing to be affected by the media). In this section we outline our interest in investigating **2WAL**, and outlines the relationship between the experience and viewer of this **Novel Interaction**.

The standard experience of video and other multimedia is well modelled by Reeve's et al. Performance and Spectator Model [188]:

> The relationship between a traditional film and the viewer can be conceptualised as a 1 Way Affect (**1WAL**) since the film has an effect upon the viewer.

Most research in the field of Neurocinematics, for example, attempts to understand how the brain responds to a given film and how its composition can affect the viewer over time [82]. Recent work, however, has examined opportunities for the viewer (spectator) to be able to influence the composition of the presented media. This creates a 2

[1] Richard Ramchurn

Way Affect Loop (**2WAL**), since the viewer would be directly influencing the flow of the film whilst continuing to be affected by the presented film (as in **1WAL**). Beyond allowing behaviours, such as voting, to affect the flow of multimedia, some research has already tried to use physiological measures to create an Affect Loop. Hillard et al., for example, successfully used a **2WAL** in the form of neurofeedback with film during focus and attention training for ADHD sufferers [86]. The study presented participants with fragments of documentary films which were manipulated (varying brightness, size and continuation) according to the participants focus and alertness levels, which were measured via an EEG brain monitoring device. Alpha Labs[2], was an electronic arts installation by the Australian artist George Khut, in which electronic soundscapes were dynamically controlled by changes in participants Alpha and Theta brainwave activity, with the effect being likened to lucid dreaming. Similarly, Carlos Castellanos presented the ''Biomorphic Aggregator''[3] a bio-responsive network data collection and visualisation system where participants physiology is used to affect a data visualisation [34].

Despite initial work utilising physiological methods in **2WAL**, little focus has been placed upon understanding the effect and the implications of its inclusion in film based experiences as well as other interactions. The lack of understanding surrounding **2WAL** provides us with the necessary motivation to explore **RQ2**.

With these research questions in mind, we pursued a research project called #Scanners which would aim to investigate the influence, role and power that **2WAL** may have on both the future of cinematic experiences and their potential impact on HCI in general.

6.3 #Scanners : An Adaptive, Cinematic Experience

#Scanners is a bio-responsive digital arts experience that blurs the lines between cinematics and neuroscience. Using a commercially available wireless EEG device, #Scanners presents a specially commissioned film that is dynamically altered both visually and aurally in accordance with the viewer's levels of Attention and Meditation, as calculated by the EEG device. The system has been demonstrated to audiences across Europe and has allowed us to explore design opportunities around extends of and awareness of control with otherwise passive multimedia experiences.

Following an exploratory 'Performance-led Research in the Wild' methodology [19], we worked with an artist to design and deliver a public experience that was then studied in-situ, qualitatively, in order to reveal wider issues and principles.

[2] Alpha Labs - http://georgekhut.com/alpha-lab/
[3] Biomorphic Aggregator - http://ccastellanos.com/projects/biomorphic_aggregator

Figure 6.1: Neurosky Mindwave Mobile, the consumer grade EEG device used in #Scanners.Image Source - Amazon.com

Richard Ramchurn, the producer of #Scanners, envisaged an interactive film which would be manipulated by the users levels of Attention and Meditation (provided by a consumer grade EEG device), and partially controlled by blinking. In the following sections, we document the artist's rationale behind the work and then report on how 35 people experienced the film when exhibited at a public arts venue. Our study reveals the different ways in which viewers experienced control through BCI (**RQ2**), the various tactics they established (**RQ3**), and how they experienced tensions between voluntary and involuntary control of the film (**RQ3**), versus being aware of their own attempts to control it (**RQ3**). We translate these findings into a two-dimensional design space, exploring partial control, that is - not fully understanding control and not thinking about control (**RQ4**).

6.3.1 BHCI Technology Choice

In exploring **RQ1**, we successfully applied the Muse EEG in a HCI style user study without participants reporting impact upon the ecological validity of the study setting. We were unable to accurately measure MWL between task conditions using the Muse

EEG, but we noted that some devices provide pre-classified measures of emotional or cognitive states - directly from the device itself.

Having reviewed a number of devices available in the current marketplace, we decided upon using the Neurosky Mindwave, shown in Figure 6.1. The Neurosky provides a single, dry sensor allowing for very quick setup and transmits data wirelessly via Bluetooth. The Neurosky provides pre-classified measures of participant's concentration and meditation values via an specialised processor that is built into the hardware of the device (TGAM - http://neurosky.com/biosensors/eeg-sensor/). The calculation and accuracy of these pre-classified measures have been discussed in existing literature [128].

6.3.2 The #Scanners Film

> We note explicitly that the visual, audio, narrative and production presented in #Scanners were purely the work of the 'Artist' - Richard Ramchurn, and we ('Matthew Pike') had no contribution to these aspects of #Scanners. We did however contribute the technology behind the BCI aspect of the experience and contributed to the discussion on how to map the brain of the viewer into the experience. We present the work in this sub-section to provide the reader with context and motivation for file #Scanners.

We worked with Richard Ramchurn, an artist/film-producer to create #Scanners, an interactive film that attempts to deliver a unique immersive viewing experience by having the composition and rhythms of the film match up to the viewers internal rhythms of thought and/or emotion. Richard's inspiration for the experience came from the work of S.Nishimoto and Walter Murch. S.Nishimoto demonstrated the possibility of reconstructing visual experiences from brain activity evoked by natural films using an fMRI machine and advanced machine learning algorithms [163]. Walter Murch postulated that blinking is an automatic response that can reveal rhythms of thought and likens blinking to cuts in film [153]:

> ''If it is true that our rates and rhythms of blinking refer directly to the rhythm and sequence of our inner emotions and thoughts, then those rates and rhythms are insights to our inner selves and therefore as characteristic of each of us as our signatures."

In addition to stating the role of blinking in expressing our inner emotions and thoughts, Murch also likens film to dream; thoughts to a shot; and a blink to a cut - a set of relationships we were interested in exploring with #Scanners. More recent research,

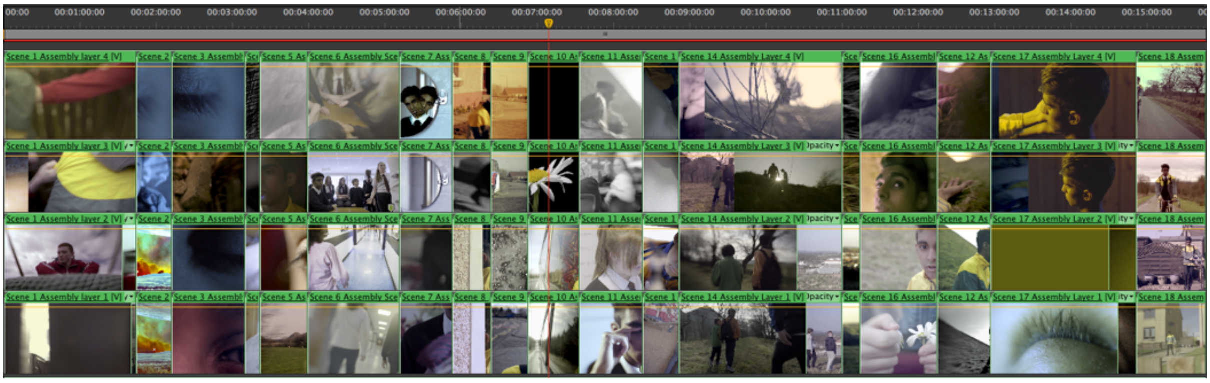


Figure 6.2: The #Scanners film shot as four layers synchronized through 18 scenes, with scene cuts aligned between layers.

published since #Scanners began, has directly linked blinking to 'changing scenes' within dreams [8], which is a concept that closely mirrors the experience that the artist envisaged.

Inspired by these ideas, Richard shot and edited an unusual film structure that consciously played with the notions of dreams and reality. The overall film ran for 16 minutes, progressing through 18 scenes. However, each scene was filmed as four distinct layers, two showing different views of the central protagonist's external Reality and the other two showing different views of their internal dream-world. The structure of 18 scenes in 4 layers is shown in Figure 6.2.

BHCI control was then used to move the viewer back and forth between Reality and Dream as the film progressed and also to control the mix of the two layers within each of these. This utilised 3 predefined outputs from the Neurosky EEG device: 1) Blinking, 2) Attention and 3) Meditation. Beyond pre-product research [131], the suitability and accuracy of these outputs for HCI have been discussed by others [75].

To match the experience design, Meditation was associated with Dream, and Concentration was associated with Reality, and Blinking would change between Dream and Reality, as shown conceptually shown in Figure 6.4.

Two groups of layers were established:

1. **Dream** - Engaged in a dream-like world. Shots were blurry and detached from reality.
2. **Reality** - Engaged in reality. Shots were focused and precise.

Within each of these groups there were 2 layers:

1. **Active**
2. **Passive**

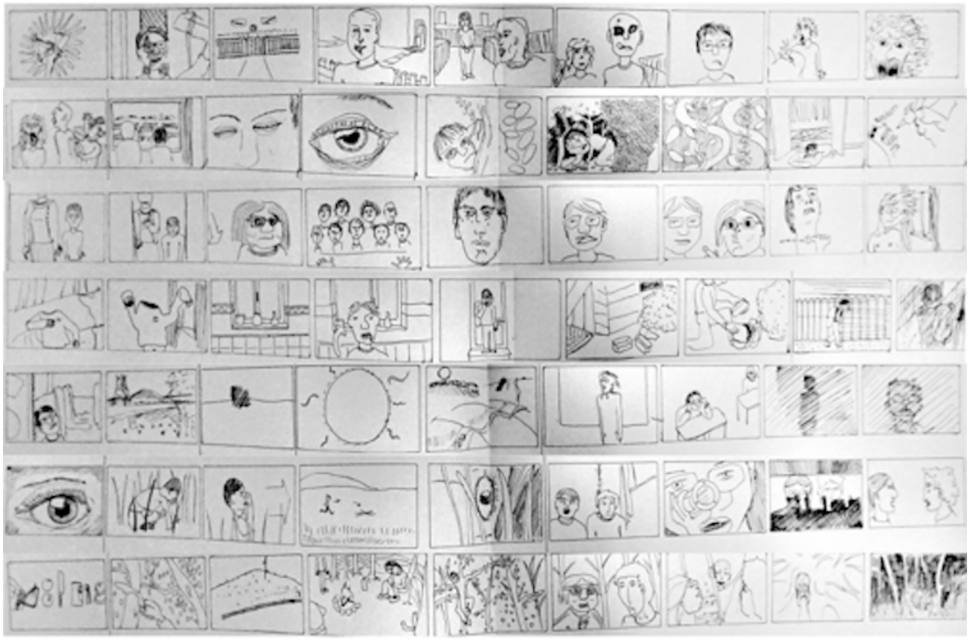


Figure 6.3: An original storyboard for the #Scanners film.

The 4 layers, therefore, were designed to have distinctly different feels or treatments:

1. Dream-Active
2. Dream-Passive
3. Reality-Active
4. Reality-Passive

While using #Scanners, blinks would force the experience to cut between either the Dream group or the Reality group, while variation in Attention and Meditation data would determine how the two layers within each group were shown. Within either the Dream or Reality group, the experience was designed such that 2 layers (2 of the 4 film clips) were playing at any one time. The clips were presented such that one clip was always present (opacity = 100%) whereas the second clip would vary on e.g. how higher their Attention was ($0\% < opacity < 100\%$). The pairing of these clips were classed as groups, and the two groups can roughly be described as dream-like and reality.

These controls were mapped onto the film's structure as shown in Figure 6.4. Blinking triggers transitions between the Reality and Dream layers. Each time the viewer blinks they move from Reality to Dream or vice versa. When the viewer is watching Reality, their level of measured Attention controls the mix of the two sub-layers. Paying high Attention mixes in more of what is termed the Reality-Active footage layer (Active because the viewer is presumed to be actively attending) whereas low Attention mixes in more of a second layer, the Reality-Passive footage. When the viewer is experiencing the Dream

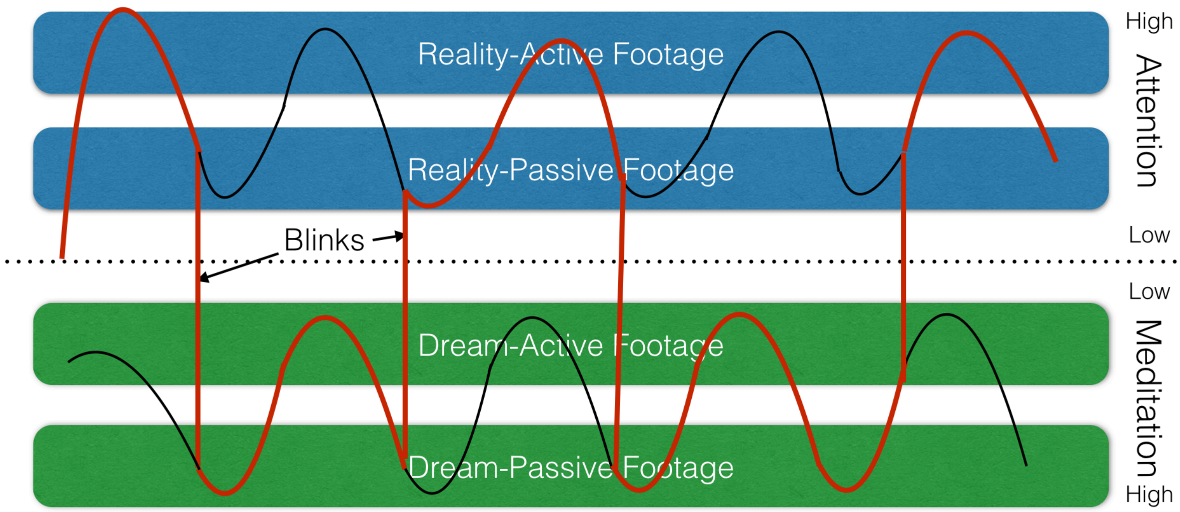


Figure 6.4: A #Scanners concept diagram, showing how a user moves between and within Dream and Reality layers.

footage, it is the measure of Meditation that controls the mix if the two layers which are corresponding called Dream-Active footage and Dream-Passive footage. Thus, in line with the inspirational sources for the film, Meditation is associated with the control of Dreams, and Concentration with the control of Reality, while Blinking triggers major transitions between them. The relationships would provide the explicit control mapping the BCI device to the experience, motivated by the discussion provided in TWAL: Two Way Affect Loops, above (**RQ2**).

Thematically, the film tackles the topic of bullying, being based on repressed memories of the artist, such that the system could mirror the internal process of recalling childhood memories, repressed memories and dream memories. Memories are not an actual record of what happened, they do not stay the same, and are malleable. Likewise the film is designed to be changed by viewing it, both consciously and unconsciously. Each layer of the film is told from a different point of consciousness, be it from a dream state, impartial matter of fact, day dreaming or high anxiety.

The script was then written over the course of 6 months before being shot at the artist's home town of Stoneyburn, West Lothian, Scotland. The film production consisted of a 10-day shoot with 6 crew members and around 30 actors. The artist and actors explored the themes of bullying and racism in discussions and by work-shopping scenes. In order to create the different layers, several scenes were shot with multiple cameras, whilst others were filmed asynchronously. Each layer had its own set of rules; for example, Dream Passive was shot using a tilt-shift lens in slow motion with a heightened colour palette. In contrast Reality Active was shot with a wide angle at normal speed in realistic colours.

Figure 6.5: A participant wearing the EEG device, experiencing #Scanners inside the caravan.

Likewise, layers within scenes were thematically separated, antagonism would be set to Reality Active, fantasy to Dream Active.

Preparing and editing the film for the system was very different to the usual editing process. When normally editing a film, an ordered sequence of clips is created that move from the start to the end of the film. Here, however, the artist created four synchronised sequences. Usually the relationship of temporally adjacent clips creates the meaning and flow of the film. However in this case, these attributes of the film are under control of the viewer so the practice of editing was to maximise the possibilities and create parallel potential meanings. This created a major editing challenge where the story had to read linearly both within and across the layers.

6.3.3 Study

Having made the film, we set about studying it using 'Performance-Led Research in the Wild' as described by Benford et al. [19]. As opposed to a rigorous scientific experimental methodology, this involves presenting the novel technological experience in an open public space with the aim of gaining rich insights into how people interacted with it. Consequently, we worked with the artist to stage the film at a high-profile public arts venue, and conducted a naturalistic study of public participants who acquired tickets and came along to try it out. This approach was taken to elicit insight into how the viewers experienced #Scanners **RQ3**.

6.3.3.1 Method and Approach

The film was presented at one of the UK's leading organisations for the development, support and exhibition of video, film and new and emerging media: the Foundation for Art and Creative Technology (FACT) in Liverpool. FACT attracts more than a million attendees and showcases more than 350 new media art from across the world, each year. The installation ran from the 14-19th of July 2015 between 10:30 and 17:30 each day. #Scanners was presented in an intimate 6 person capacity cinema, within a caravan (shown in Figure 6.5) to emulate a rich cinematic experience. The space had no windows, low lighting, plush seating, an eight foot projected image, and stereo speakers.

Participants were recruited in-situ, with advertising placed outside the caravan and through word of mouth, with a number of viewers viewing the experience on the recommendation of a friend. The experience was also advertised on the FACT website. In total, around 75 people had the opportunity to experience #Scanners as the main viewer, sometimes with up to 5 additional spectators. All viewers were told that data, and their unique version of #Scanners, would be recorded for later analysis and were given a chance to opt-out. 35 of these agreed to participate in our study and provided informed consent. Active viewers were fitted with the headset which took on average about 2-3 minutes to set up. Their film experience lasted approximately 16 minutes. Each participant's unique version of the film and brain data was recorded. The session concluded with a semi-structured debriefing interview, focused on people's experiences and feelings whilst using #Scanners, which typically lasted around 10 minutes. Video and brain data recordings were synchronised using network time. Unfortunately, log data for 13 of these participants was incomplete. As such our study focuses on the 24 who had agreed to be interviewed and for whom we have complete logs of BCI data.

6.3.3.2 Overall Response to #Scanners

We present, in the following section, a body of evidence into documenting how participants experienced #Scanners, in contribution towards answering **RQ3**. We begin by looking at the participants perceived experience on a macro level, identifying that overall, participants were generally positive about #Scanners:

> "Yeah it was quite crazy." - *P35*,
>
> "I felt like I could slow it down, speed it up and I could move on." - P33*
>
> "It'd be great to watch on drugs."- *P29*

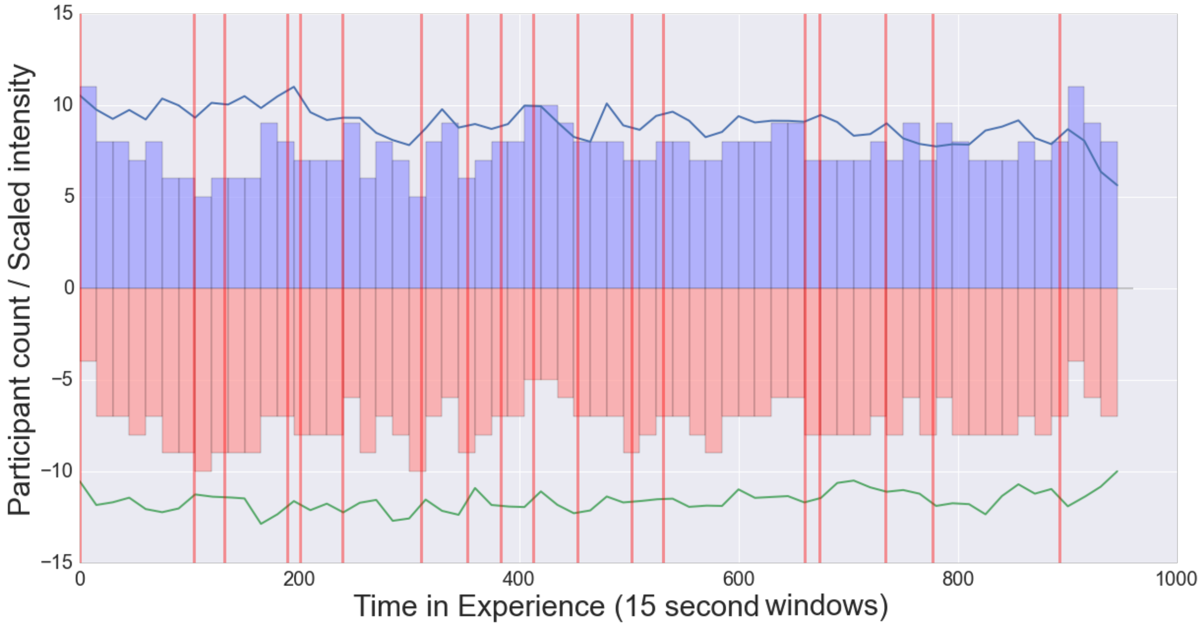


Figure 6.6: Visualisation of Attention (blue line), Meditation (green line) and state (Reality = purple bars, Dream = red bars).

The experience did feature some challenging subject matter, specifically around the subject of bullying -

> "I didn't enjoy the section where the child is bullied because I got bullied in school when I was a kid and I remember thinking... I didn't actually want to watch that." - *P33*.

Others noted a similar discomfort:

> "induced a sort of sense of unease." - *P15*

Considering the switch between reality and dream footage, one viewer said:

> "I found it, I don't know if scary is the word, but perhaps a little unnerving in places because usually when you see a film you see outside the character but because you could move to see from inside the characters perspective it was a bit like: oh, I don't want to pick that brick up." - *P1*

Figure 6.6 shows the collective experience of our 24 participants. The bars above zero (X-axis), indicate the number of viewers in the reality state at a given moment, whilst the bars below indicate the number of viewers in the dream state. The graph bars denote the number of viewers in each group at 15 second intervals.

The line plots indicate the scaled average intensity of attention (blue) and meditation (green) across all participants.
The vertical red lines indicate the timing of synchronised scene changes in the footage. The graph confirms that the control mechanism was reasonably balanced, showing that people did flip between dream and reality and that levels of attention and meditation varied in apparently sensible ways. It gives a first impression that the control mechanism was reasonably well behaved as we might have hoped.

While clearly not a controlled experiment, we have calculated some statistics to help characterise the overall nature of the experience in terms of how much people were deemed to be blinking, attending and meditating and how this shaped the overall viewing patterns of the film. Attention and Meditation values were provided directly from the EEG headset and range from 0 (low) to 100 (high). The average attention level was 53.23 (stdev 5.38), while the average meditation level was 57.75 (stdev 5.09), with a correlation ($r=0.07$) confirming the independence of the two measures.
We examined whether Attention and Meditation levels varied between scenes, using a repeated-measures ANOVA.
Only Attention varied significantly ($F(17,6)=4.125$, $p<0.0001$), indicating that people's Attention did vary from scene to scene, but their Meditation, or calmness, did not. Pairwise comparisons highlighted that Scenes 3 and 4 drew the highest levels of Attention, during which the primary character in the film experiences an intense and anxious nightmare. Conversely, scenes 17 and 18 drew the lowest levels of attention, which could be explained by decline in interest as time progresses. The focus of scenes 17 and 18, however, is on the main character resolving some of their problems towards a positive ending, involving a calm cycle ride through their neighbourhood. This may have simply commanded less attention than an intense nightmare, or perhaps viewers consciously withdrew attention from a less interesting scene.

There was a large variation in how participants blinked whilst experiencing #Scanners.
The normal blinking rate amongst adults is estimated at 10 blinks per minute [48], however, VDU use and watching TV is associated with lower rates of about 5 blinks per minute [169,97], which Ishimaru et al. used to detect when participants were watching TV. In line with such findings, the average number of blinks was 77.6 (stdev 35.5) during the 15:46 minute film, which is about of 5 blinks per minute.
The recorded average interval between blinks was 14.6 seconds (stdev 6.5s).

Beyond this lower average rate associated with watching TV, however, some participants clearly managed their blinking carefully.

P2 is one example of a participant who was clearly conscious of their blinking:

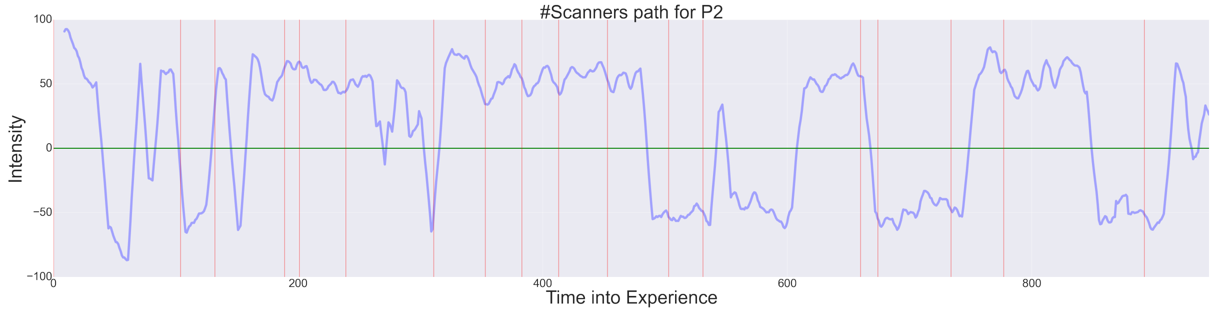


Figure 6.7: A Visualisation of *P2's* journey through #Scanners. Above the X-Axis is attention, below it is meditation. P2 blinked 28 times.

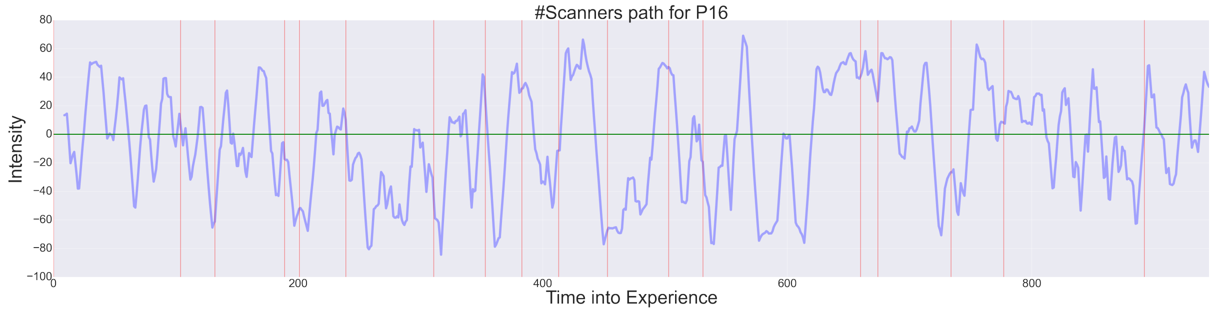


Figure 6.8: A visualisation of *P16's* journey through #Scanners. Above the X-Axis is attention, below it is meditation. P16 blinked 158 times.

> "I definitely had control over the way it was edited as I'd seen that version before I kind of knew I could change the footage that I hadn't seen and see the other version." - *P2*

P2 blinked only 28 times, or approximately two times per minute, including a break from blinking for more than two minutes between scenes for 7-11, as shown in Figure 6.7 (Note - a blink is indicated by the graph line crossing the X-Axis).

Others, however, blinked many times throughout the film, as shown by *P16* in Figure 6.8 who blinked 158 times, or 10 times per minute. *P2*, who had just watched their friend using #Scanners, said:

> "I think because I watched it before, there was some scenes I didn't want to see or I wanted to see the other side of the scene. So that I found really interesting, so I wasn't passive it was nice to have control over it."

This confirms that some participants deliberately refrained from blinking as a control tactic.

We decided to allocate participants into groups of high and low blinkers. The top third of blinkers were classified as high blinkers, whereas the bottom third were classed as low

Figure 6.9: Attention levels of High (red) and Low Blinkers(blue) by scene.

blinkers. We did this in order to investigate whether there were significant differences between the two. We saw no difference in overall Attention/Meditation levels, but we did see, on a scene-by-scene basis, that the low-blinking group tended to have higher levels of Attention (Shown in Figure 6.9).

6.3.3.3 Discovering Control

Having presented some general observations across the whole group of viewers we now turn to looking at the finer details of control as described by the viewers in the subsequent interviews (**RQ3**). Many who experienced #Scanners began without knowing how it could be controlled, and the amount that they discovered whilst watching the film varied; some quickly came to understand aspects of the control, even if they did not interpret its effect correctly. Aside from knowing they were wearing a brain scanner, however, a few participants remained unaware that they could have some control over #Scanners:

> "I didn't realize there was anything to control" - *P19*
>
> "I thought it was just how it was" - *P16*

Some, during the experience, realised that they had some control:

> "It did feel like something else was controlling. There was something more kind of [transient] in the edits, you know. It didn't flow in the same way that it would do if you were watching something else…" - *P15*

P28 discovered, during the experience, that blinking was having an effect:

> "I noticed that when I blinked, it changed between blue and red, and green and white. And I liked the blue and red so I tried to keep on that as long as I can." - *P28*

People didn't necessarily understand how they were controlling it, nor what impact their control was having:

> "[I tried to] alter my breathing …. I tried messing about with my hands in front of my eyes" - *P6*

Others believed they were altering the storyline, saying:

> "there were little bits where I could control whether people were being aggressive or not." - *P7*
>
> "I did [think I could control characters] near the beginning, especially when I thought: pick that brick up and hit them." - *P10*

Others believed they were influencing the temporal flow of the film, saying:

> "I thought, if I really, really focus on what's going on, it will travel quickly and I will get through this section that I don't really like, if that makes sense and it seemed to do that." - *P33*

Some participants knew at the start how the system worked, because they had watched a friend experience, or had spoken to a previous participant.

> "I enjoyed concentrating because I had the control of the concentrating" - *P5*

While some had control over their attention levels, it wasn't always easy for others to so:

> "my meditation, I tried to play with that but I wasn't sure if I was having much effect" - *P6*

Participants seemed happy, however, with this limited control:

> "I was happy with the amount of control, because I didn't know the parameters of how to affect it and trying to manually affect how your brain is reacting is really difficult" - *P1*

6.3.3.4 Exerting Control

For those that exerted a level of control, some found that this increased their engagement with the film:

> "more immersive definitely, I'm used to going along with a storyline and having no control over what's happening and feeling not-connected to the film whereas, that I felt more involved with it, more connected to the film, and to the characters as well" -*P2*

Others, like *P2* above, used this control to manage their exposure to difficult material; *P18* said:

> "The audio became really, really annoying and very abrasive. I was using the opportunity to just switch to a less abrasive... I mean, both were still abrasive but I was switching to the less abrasive at that point and checking in, every so often" - *P18*

Using control wasn't always easy due to its semi-autonomic nature:

> "yeah I tried to play with [blinking] but sometimes I blinked involuntary, so sometimes where I didn't want it to change it would change" - *P6*.

Similarly *P22* said:

> "I think sometimes like I was stopping myself from blinking and then my eyes will get dry."

Conversely, thinking about control meant that some participants found it hard to fully enjoy the film:

> “A lot of the time I found it difficult to remove myself from the thought of the fact that I was changing it and I was controlling it, and I kept thinking like why is my mind doing that pace like what’s going on in the film?” - *P23*

Similarly, *P31* said:

> “sometimes you notice that you have the control, and that flipped you out of flow. But sometimes you’ve really added to the dramatic effect.” - *P31*

P28 worried that even more control might reduced the enjoyment of the experience:

> “I think any more control maybe I think that, maybe more control would have taken away from the immersive elements of the film.” - *P28*

Based upon interview discussions, we categorised users as to whether they actively tried to Control (*N=11*) the system or not (*N=13*). As with High and Low blinkers, we did not see significant differences in Attention and Meditation across the whole film. Looking scene by scene (Figure 6.10) however, we do see that Controllers had almost consistently higher levels of Meditation. This is different to what we saw for High/Low Blinkers (Figure 6.9), where we saw the same effect but for Attention. This might imply that, instead of high Attention being associated with low-blinking, low-blinking was more of an autonomic response associated with high Attention. Conversely, those that tried to control their interaction managed to actively control their responses, including Meditation levels. The lack of difference in Attention between those that tried to control the system, and those that didn’t, supports the qualitative interview data suggesting that both groups had mixed success at affecting their own attention.

6.3.3.5 Releasing Control

Understanding and using control did not mean, however, that users retained control over the film throughout. Some enjoyed giving up the control, with *P31* saying

> “So you try not to blink. So that was, I think it did add to it, yeah feels good. I mean I could have let it go as well a little bit, but that was nice. Sometimes I just let go. It’s good.”

Some participants even forgot about their control for periods of time, with *P1* saying:

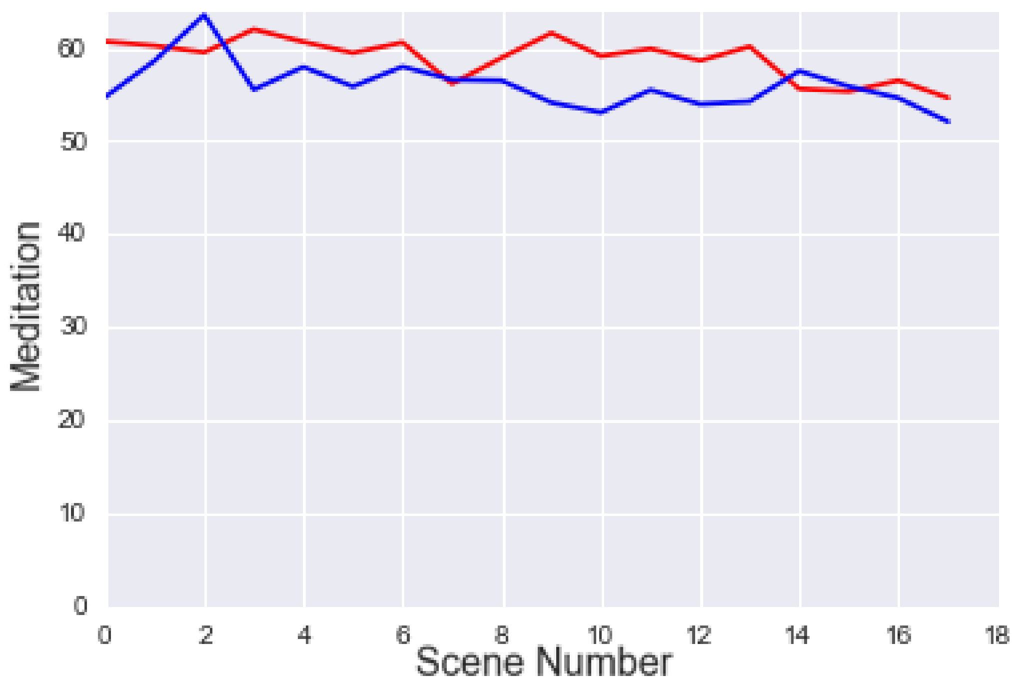


Figure 6.10: Meditation levels for those that self-identified as trying to Control (Red) and those who did not (Blue).

> "I completely forgot, I was concentrating on the film then a couple of minutes the pressure point there started aching and yeah I'm wearing a headset." - *P1*

In reflection, some said that it would have been better not to know exactly how to control the film. *P13* said:

> "I wish you hadn't told me before, its not as authentic if you know before" - *P13*

and *P14* said:

> "I would have liked to have done it, not knowing anything" - *P14*

When asked if *P10* (who did not figure out how to control the system) would have liked to have known in advance, they said:

> "No [because] then I'd be blinking like anything, and its our observation that changes what happens anyway. I'd have liked to be told that I would be told [how it worked] afterwards. Maybe I would have thrown myself into it more." - *P10*

6.3.3.6 Summary of Findings

Our findings reveals the diverse ways in which viewers experienced #Scanners and provides a significant body of evidence towards answering **RQ3**. No single experience of the film was the same and viewers engaged with the possibilities of BCI control in many different ways. Some actively employed control, adopting deliberate tactics such as not blinking, whilst others did not or were not aware that they could exert control. Some subsequently relinquished control as they became immersed in the film or as their autonomic responses reasserted themselves. Some learned about control for themselves, by experimenting or watching others, while others were told about it. Taken together, these findings present a complex, even bewildering picture. How are we to understand what is happening here? What lessons might filmmakers and other "designers" of interactive entertainment draw from our findings? This tells us that to answer **RQ3** is to identify that relationships between the viewer and interactive experience is an extremely complex and personal one, requiring a distinct understanding of how the experience can be designed to facilitate these relationships and not necessarily how to guide or control the experience. This is something we explore, in answering **RQ4**, through the following section. We set out to specify a structured design space for BHCI control to systematically relate various findings to one another.

6.4 Taxonomy of Passive Control

We now reflect on both the design and viewer experience of #Scanners in order to draw out more general lessons for HCI about the use of BCIs in interactive entertainment. Our discussion takes the form of a gradually building conceptual framework to explain the subtleties of control when using BCI as this - in various guises - emerged as the central issue from our experience. The aim of establishing a conceptual framework is to:

1. Define concepts to explain our findings
2. Ground them in HCI related literature
3. Reveal unusual strategies and tactics for designing future entertainment.

Our framework involves the definition and comparison of two key dimensions of control - *the extent to which control when using BCI can be considered to be voluntary and the extent to which the user is aware or trying to control the system.* This framework aims to address **RQ4** and provide a framework upon which interaction designers interested in utilising BHCI based control can develop future works using this interaction.

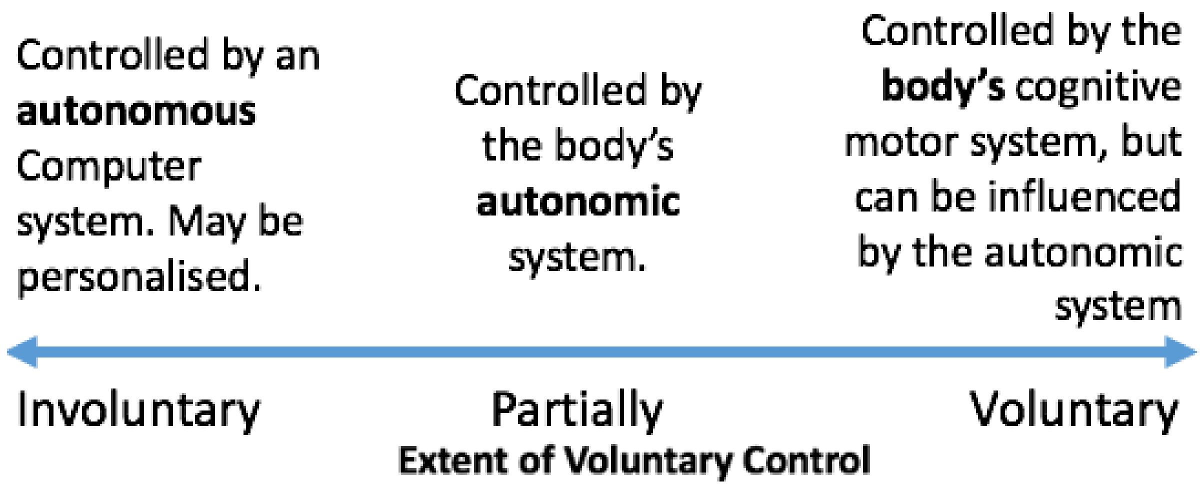


Figure 6.11: Extent of Voluntary Control as a dimension.

6.4.1 Extent of Voluntary Control

A key motivation for some BCI research, is that people are not fully or directly in control of their interactions, but that systems respond reactively to their behaviour [253], such as with the control of artificial limbs [155]. To be more precise, we would argue that their control of the system is not entirely voluntary. Thus, while we have seen examples that viewers can learn to voluntarily control their blinking in order to try and prolong or break away from scenes in the film, we have also seen how blinking is also an semi-autonomic bodily response to drying eyes, dust and other factors. In short, one cannot refrain from blinking forever. Our findings suggest that our interpretations of attention and meditation were perhaps subject to even more tenuous voluntary control. While viewers often wanted to control them - and some claimed that they could - control was exerted through indirect means such as trying to slow breathing in order to calm down and be more meditative. Whilst users were more able to influence their Attention levels, and we saw a significant difference in Attention levels between different scenes, Meditation levels did not vary significantly throughout the film.

Given these observations, it is useful to think of there being a dimension of **Extent of Voluntary Control**. At one end we find forms of control that are largely voluntary, such as choosing to move a mouse or press a key on a keyboard. At the other might be forms of control that are largely involuntary such as sensing the actions of the body's autonomic systems that continue to operate (e.g., the presence of any detectable brain activity at all). Our interpretations of BCI in terms of blinking, attention and meditation are notable for occupying a middle ground along the spectrum where control is partially voluntary. Similar to Marshall et al. breath-controlled bucking bronco [139], the user can choose to blink at certain points, but cannot avoid blinking at others. The user can try to relax, but may be affected by surprise or fear during the film. The user can choose

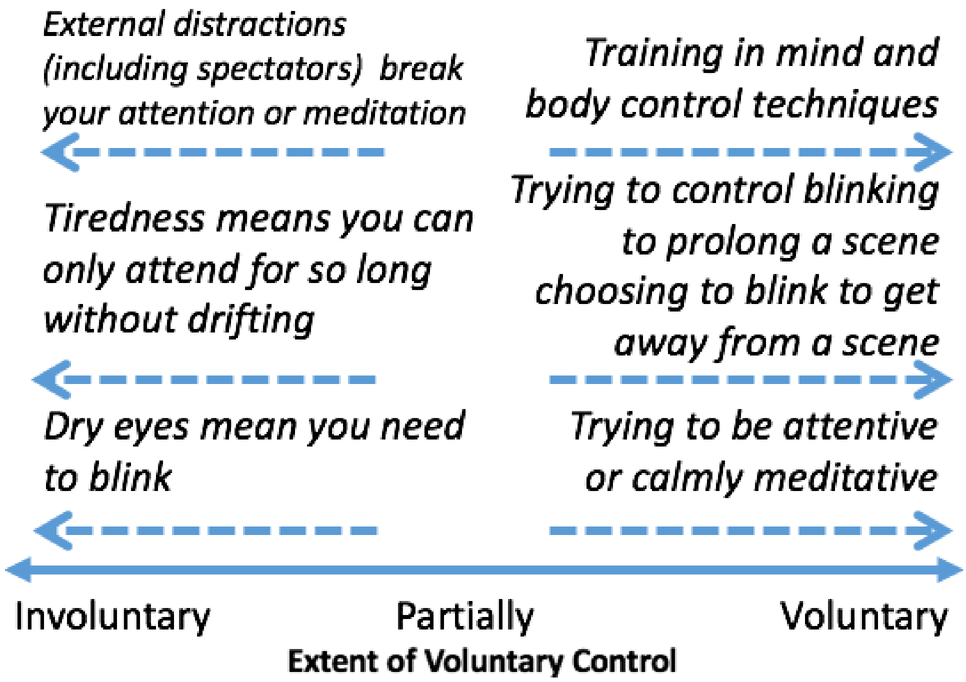


Figure 6.12: Triggers affecting voluntary control.

to be more attentive, but may struggle to maintain attention during less action-oriented scenes. The user's position on this spectrum can vary during an experience. Our findings revealed both deliberate and accidental triggers that might cause movement along this spectrum (see Figure 6.12).

6.4.2 Extent of Self-Awareness

Our second dimension, shown in Figure 6.13, concerns the extent to which one is self-aware of one's level of control over one's body, including thinking about controlling the system. This **Extent of Self-Awareness** can vary between being fully conscious of what one is doing, such as when manipulating a mouse or keyboard, to when our attention is focused elsewhere, such as, when we are deeply immersed in a state of flow when watching a film [43]; like riding a bike without thinking about how to ride it.

Our findings reveal that our particular treatments of BCI in terms of blinking, attention and meditation span various points along this dimension. Users can be consciously aware of trying to control their blinking or unaware of their everyday blinking behaviour. They can be deliberately trying to play close attention, whilst some became immersed in the film and forgot about trying to influence it. Moreover, we have seen how this level of consciousness may vary dynamically throughout an experience as a result of various internal and external triggers that are shown in Figure 6.14. We noted, for example, how changes in content such as a scene transition in a film might potentially move the user in

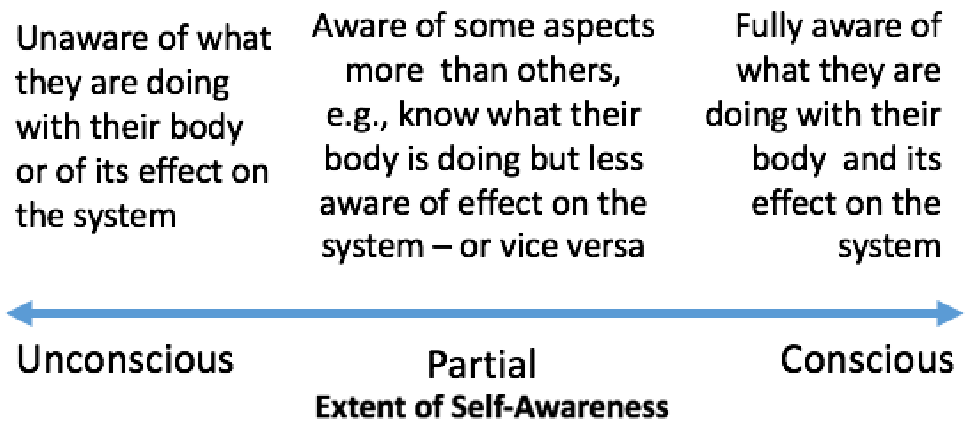


Figure 6.13: Extent of Self-Awareness of control as a dimension.

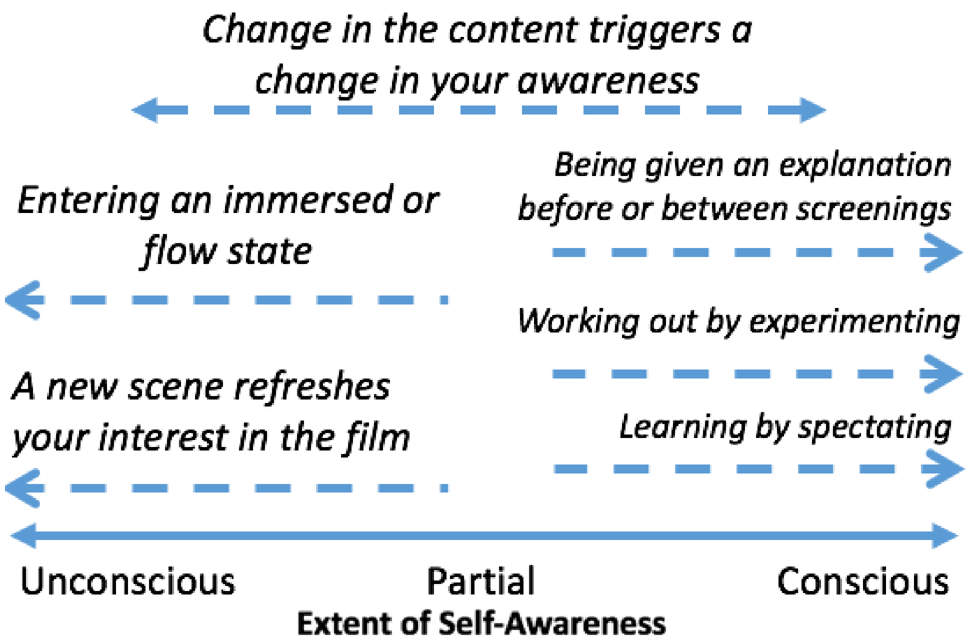


Figure 6.14: Triggers affecting Self-Awareness of Control.

either direction, re-engaging their attention with the film or causing them to reflect on whether the transition was caused by their blinking.

6.4.3 A Design Space for Entertaining BCIs

Combining these two dimensions reveals an important design space for the control of BCIs (and possibly other modalities too). Our experience revealed something of a tension in the use of BCIs where users move back and forth between voluntary and involuntary and between conscious and unconscious with different effects. Beyond helping explain our findings, we might also put this taxonomy to use as a design space for BCIs, especially for entertainment.

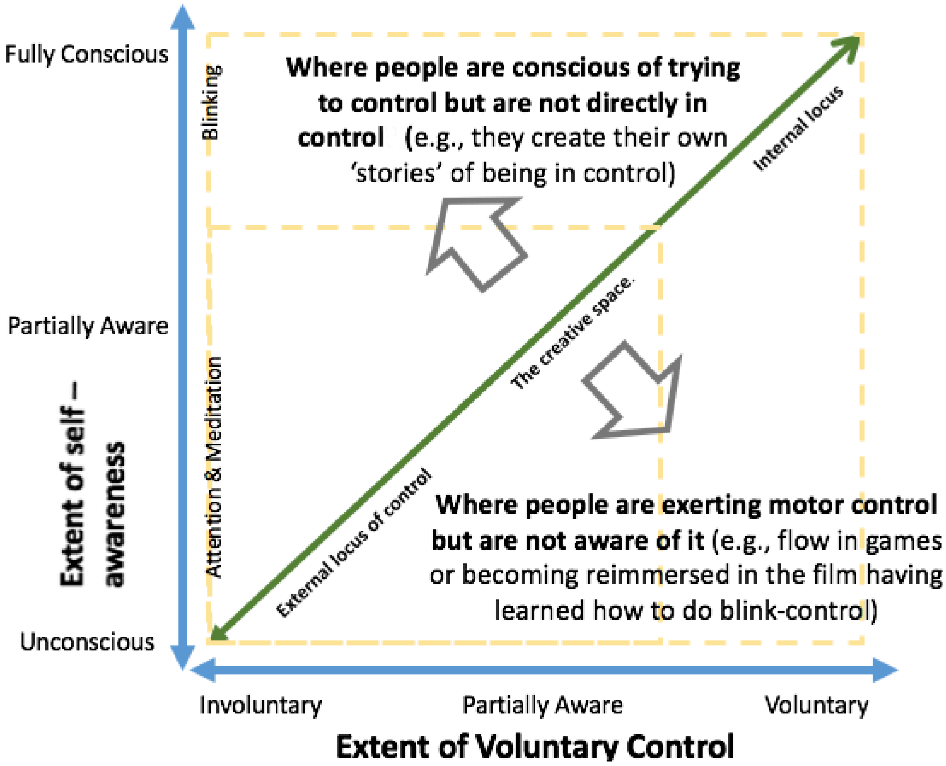


Figure 6.15: A design space for entertaining BCIs.

The green diagonal line in Figure 6.15 represents the traditional locus of control in HCI [202]. This moves between internal locus of control (e.g., direct manipulation) and external locus of control (e.g., autonomous, context-aware and ubicomp sensing systems).

We suggest that this space away from the central line offers a creative sweet-spot where designers can set up creative tensions and/or trigger users to move between different states: between immersion and self-reflection, and between being in control and surrendering it [139,20].

At the *top right*, the user is in control with the system responding, the human is therefore in control, a classic examples of direct manipulation.

The *top-left* represents a state of one becoming aware of involuntary control, something we observed in #Scanners where viewers consciously refrained from blinking to prolong/avoid a particular scene. At this point, the user is becoming aware of how aspects of the system works, and is refraining from particular, understood actions, to effect a particular state.

At the *bottom-right* a user is unaware that they are exerting control. This seems counterintuitive, as exertion of voluntary control is surely a conscious? We would relate this state to the concept of 'flow', where the demands of the task align with the skills of the user, creating an enjoyable, immersive environment in which the task is experienced [43].

At the *bottom-left* we see examples of ubiquitous/autonomous systems where a system will take control of the user's environment/physiology without their awareness. Examples of this may include climate control systems found in vehicles and buidings which automatically regulate the user's environment without their explicit awareness, yet affecting the bodies autonomic system as it reacts to changes in the temperature.

This is a liminal space - a space of in-betweenness and ambiguity - that can be particularly productive in creative fields and may encourage people to create their own interpretations or 'stories' of control as we saw in our study.

Our results already suggest some general strategies that involve *thinking off the line*: **Fully-conscious and involuntary** - where we explore physiological measures that people have even less understanding or voluntary control over, such as skin temperature, and **Voluntary but unconscious** - which encourage people into states, perhaps like experiencing flow [43], where people could use voluntary control, but do not need to consciously think about doing it.

6.4.4 Journeys through Control

We return to the idea that this is a dynamic picture - that participants can make various transitions around this space. In an attempt to characterise the many paths taken by those who experienced #Scanners, we developed the state diagram seen in Figure 6.16. The diagram provides a visualisation of the potential state and transfer of state a viewer may experience during #Scanners. Reflecting on the interview data led us to propose the following states.

State (**1**) indicates one of two entry points into the experience. In (**1a**), viewers have zero prior knowledge of the system's operation. There may be some awareness of the possibility of some control, since they are wearing the headset, but there is no explicit knowledge. Conversely, (**1b**) represents the state of knowing. Viewers in (**1b**) will possess varying degrees of knowledge of the system's operation but have yet to exert any elements of control.

State (**2**) represents the state of pre-discovery. For viewers transitioning from (**1a**), this will be the beginning of their discovery, they will begin to notice certain associations between their physiology and manipulation of the experience. Transition from (**1b**), will begin the process of confirming existing knowledge. Elements of immersion are possible in this state. Discovery is then witnessed in (**3**). This is the "*ah-ha*" moment where viewers figure out some/all elements of control associated with the film. We note certain individuals, who were in the upper right corner of (**3**), were exerting complete control

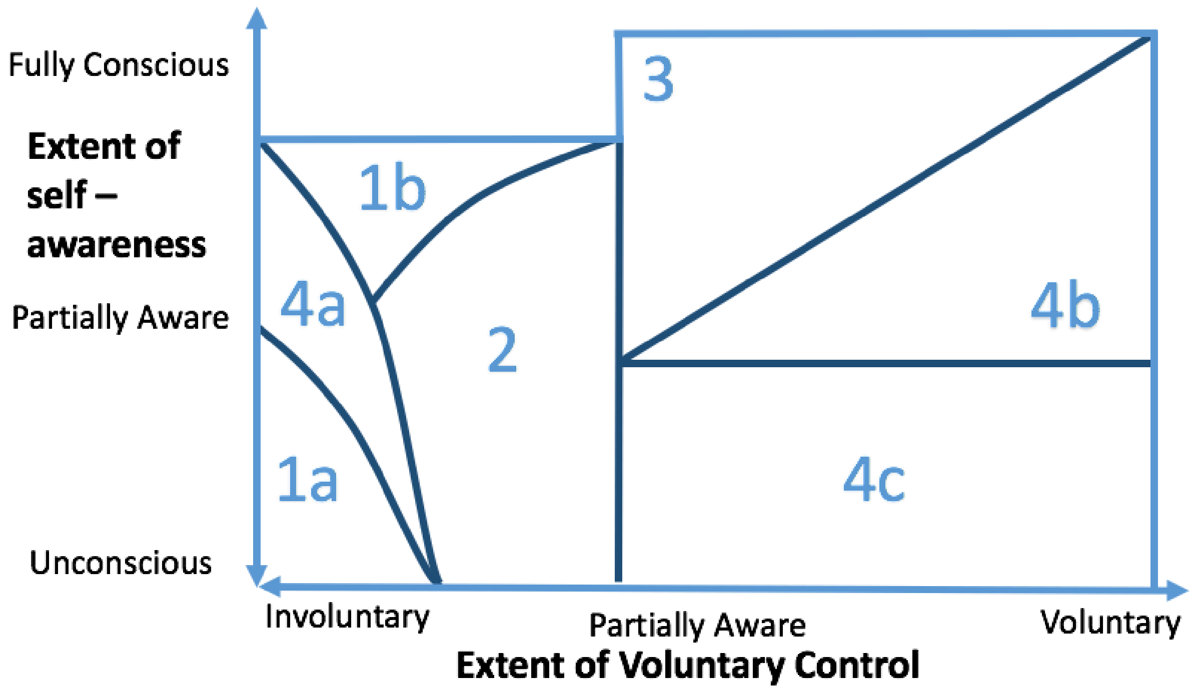


Figure 6.16: Major states participants travelled through whilst experiencing #Scanners.

over the film, but were not immersed in any way e.g. blinking at rapid rates to manipulate the experience but were not at all immersed in the film. Within this state, participants were typically thinking more about how they were controlling, rather they understand how it worked; people tried different ways to relax and tried focusing on different things, like controlling the storyline or the actors.

Post discovery occurs in (**4**). (**4a**) represents the viewer that did not discover more control (partial knowledge). They remain aware that something is affecting the experience but do not discover more. In (**4b**), some participants enter a stage of understanding with the system control, and begin to use it to control the system. Similarly, those that understand elements of control (e.g. blinking effects the cutting of scenes in some way), may simply slowly relinquish explicit control and fall back into immersion, where their knowledge of control has increased, but they no longer think about doing it (**4c**).

Table 6.1: Unique identified journeys whilst using #Scanners.

ID	Path	Participant	Total
J1	1a	16, 19, 29	3
J2	1b	21	1
J3	1a → 2 → 4a	10, 11, 15, 26	4
J4	1a → 2 → 4b	33, 35	2
J5	1a → 2	12	1
J6	1a → 2 → 3 → 4b	18	1

ID	Path	Participant	Total
J7	1a → 2 → 3 → 4b → 4c	1, 28, 30, 31	4
J8	1a → 2 → 3 → 4c	3, 8	2
J9	1a → 3 → 4b	22	1
J10	1b → 3 → 4b	2, 5, 6, 13	4
J11	1b → 2 → 3 → 4b	23,25	2
J12	1b → 2 → 3 → 4c	20	1
J13	1b → 2 → 3 → 4b → 4c	7	1

From detailed analysis of each viewer's interview and video data we were able to identify 13 unique paths through the experience. Table 6.1 allows us identify some of the interesting characteristics associated with the experience. Table 6.1 was primarily constructed based on the interview data we collected from participants after they had participanted in the experience. We rewatched each individual participants' video recordings and relayed their current state to the appropriate state detailed in Figure 6.16.

J1 and **J2**, for example, shows individuals that never moved beyond the initial state (**1**). For (**1a**) this experience would be analogous with watching a standard film (e.g. *P16*). Perhaps more interesting is *P21* who remained in (**1b**) i.e. they have prior knowledge of some of the systems control, but chose not to explore what exactly they could do with it.

J3 and **J4** was the most common journey taken by viewers who did not have prior knowledge of the system's control. The journey indicates that they discovered, based on the system's responses, behaviours that might create a change. Those that never quite work it out, then return to (**4a**), whilst those that do typically moved to (**4b**), where they begin to use control. Some, however, went as far as (**4c**), where in **J7** they then forget about control and enter a state of immersion, where they know what is controlling the system but stop thinking it.

For those that knew in advance how the system worked (**J10-J13**), the path was similar, but typically involved less time in the exploration state, and more time in the 4th states. Realising that some participants knew the operation of the system, however, we can begin to consider two types of transitions: *Intra-experience transitions* that happen during a given experience (i.e. screening of the film) and *Inter-experience transitions* that happen in between experiences, for example as a result of receiving an explanation of how the experience works or perhaps as a result of being a spectator to someone else's experience. In inter-experience contexts, viewers will likely trace a path around our design space as they engage in possibly repeat experiences. In this regards, P33 said:

> “I want to have another go to see what I can do, because I think I was quite passive. I was very aware of my emotional responses watching it. I was kind of..quite…I guess I was monitoring my emotional responses quite…and allowing them to be quite strong because I kind of had some vague idea that, that might, you know, provide more information for the feedback thing”. -*P33*

6.5 Conclusion

In this work, we successfully applied BHCI to provide a **Novel Interaction** to the viewer of an interactive film. By creating an interactive film, where the scenes seen by viewers varied depending on their levels of meditation and attention, we have explored a novel design area for including BHCI in multimedia experiences (**RQ2**). The uncertainty and non-explicit form of control has lead to an interesting new creative space in which experience designers can begin to explore (**RQ4**). The experience was explored with members of general public as an installation at a week-long national arts exhibition.
Most notable, amongst our findings, was that while the BHCI based adaptation made the experience more immersive for many viewers, thinking about control often brought them out of the experience (**RQ3**).
This led us to propose a two-dimensional taxonomy of control, considering both the understanding of the control, and how much users think about control(**RQ4**). A traditional belief in HCI is that Direct Manipulation (being able to control exactly what you want to control) sits at the top of both these dimensions.

We examined, however, how users deviate from line, and enjoyed the experience more by either not knowing exactly how it worked, or by giving up control and becoming re-immersed in the experience(**RQ4**). We conclude that these deviations from the line between knowledge and conscious control over interaction are most interesting design opportunities to explore within future BHCI adaptive multimedia experiences.

6.5.1 Contribution to HCI craft knowledge

6.5.1.1 Framework of Control

The primary output of this study is the specification of an interaction taxonomy that details the manner in which user’s discover various stages of control in a system exhibiting non-direct forms of interaction. Through reflective analysis of participants’ experience of #Scanners, we have produced a taxonomy of control that specifies how and when user’s may choose to exert this control as they discover elements of the systems operation. We believe this taxonomy represents a significant contribution to the HCI community.

First, the taxonomy provides interaction and experience designers with a foundation from which they can design, develop and implement new interaction experiences that involve indirect and passive forms of interaction. Given the novelty of this type of interaction, the specification of this taxonomy can provide significant aid to these designers, as there is little existing reference upon which they can base their future work. Further, given the commercialisation of psychophysiological sensors (including BCIs), it is likely that the relevance of the taxonomy will be increasingly more significant in the near future.

We believe the taxonomy will also provide value during the evaluation stages of a project, especially when concluding or exploring the results of a scientific HCI based user study. The taxonomy could be used as a basis for explaining or identifying actions or behaviours exhibited by participants, in a manner similar to which the taxonomy was developed.

The ultimate aim in producing this taxonomy and it's contribution to HCI is to provide a useful classification of the stages of control that a participant may exhibit during stages of interaction with a passively controlled interaction.

6.5.1.2 Verification of a HCI research methodology

In this study we contributed to the body of knowledge documenting the use of the Performance Led Research in the Wild methodology[19]. The methodology is well regarded, and has a number of existing documented applications in the HCI literature. We further contribute to this documentation through the application of this methodology in this study.

Chapter 7

Discussion

In this section we reflect upon the studies presented in this thesis and discuss the findings and contributions to the field of HCI. We also discuss the broader impact of these discoveries and how they may be applied to future works.

7.1 Applied BHCI for Evaluation

In our two initial studies, we conducted a form of HCI evaluation using a BHCI centred approach, in accordance with exploring our thematic interests in **methods** of applying BHCI and evaluating different forms of **input control**. In doing so, we hoped to utilise BHCI to provide a quantified, objective measure of how much mental work was performed by a participant whilst completing the provided study task. We used these objective, brain-based measures as the primary form of evaluation in each study, with the hope of observing differences between task conditions.

In our TAP study, we evaluated the effect of using different verbal protocols had upon participants against a baseline of simply competing the mathematical study task. The aim of completing this work therefore was to a) Demonstrate the application of fNIRS for Evaluation; b) Investigate whether TAPs do significantly impact a user's MWL and c) Detail how a researcher may go about performing this kind of evaluation, using fNIRS.

In evaluating TAP, we addressed the two initial aims above. We demonstrated that fNIRS is well suited to conducting BHCI based evaluation and that TAPs do not significantly interfere with the user's MWL. Although we are not the first to apply fNIRS for evaluation in HCI (Peck et al. evaluated different forms of visualisations using fNIRS - [173]), we are among the first to use this approach and contribute to the early body of literature documenting this application. We do however address a gap in the literature regarding our understanding of the impact of using TAPs as a tool in a research study. The results of the study indicated it did not, in-so-long as the verbalisations related to the task at hand. It was found that non-task related verbalisations did significantly effect the MWL of participants. Prior to this finding, we could not be certain of the impact a TAP may have upon influencing the results of the study. We have demonstrated, through BHCI based evaluation however, that they can be safely integrated into HCI studies, without affecting the MWL of a participant. Not only does this result reinforce the value and integrity of TAPs, it also serves as an example of the successful application of BHCI based Evaluation.

We followed a similar approach for our LEAP study, focussing on evaluating different forms of **input control** for completing a jigsaw puzzle. In doing so, we explored the **methods** of applying BHCI in this domain, and strove towards more ecologically valid settings. We attempted to evaluate the impact of an augmented reality input device using a consumer grade EEG device. Ultimately we were unable to identify variations in MWL via the EEG, despite observing differences using an alternative measure (NASA-TLX). However, in completing the study, we did contribute to a gap in the existing body of knowledge surrounding these different forms of **input control**, the evaluation and relative comparison of which had not been performed before. We also outlined a suitable task for conducting this work, which was previously undocumented.

The allure of BHCI based evaluation is clear; having an objective, quantifiable metric for evaluating a product or interface in an unbiased manner is a valuable tool in the arsenal of a HCI research. Current approaches to HCI evaluation are typically subjective and reflective processes, relying on the skill of the researcher to elicit insight into the focus of the study. In addition to being applied in a HCI context, as we have advocated here, the approach could benefit a number of other fields.

Marketing is an example of such fields that have embraced the application of BCI technology for evaluating the impact of a marketing campaign upon the intended audience. Coined as 'Neuromarketing' a number of academic and commercial enterprises are actively exploring the application of BCI for the purposes of evaluating advertising material and campaigns.

In [228], Vecchiato et al. utilised EEG (as well as other physiological measures) to observe 15 participants as they watched TV based commercial advertisements. They identified increased cortical activity in the theta band in participants who 'remembered' an advertisement versus those that 'forgot' an advertisement. Similarly, in [230], viewer reactions to political speeches given by the Italian prime minister were investigated in 13 healthy subjects, with results indicating that cortical activity varied according to the viewers political bias. The authors propose that the work could be use to strengthen the quality of future speeches in an attempt to attract and pursue “swing” voters. Vecchiato et al. have conducted a number of other Neuromarketing related works [227,229,231,232,9]. A review of various imaging technologies (EEG and MEG) are detailed in [227].

An interesting consideration, and point of discussion especially as this approach continues to see adoption is the privacy implications of the approach, especially when this data is being used for economic gain. The ethics and moral implications of any new technology need to be considered, and this is especially true for BHCI based evaluation since the data that is obtained from conducting these types of studies is highly personal to the individual. Murphy et al. coin the term “Neuroethics” and call for the consideration and protection for the vulnerable in society and for a general code of ethics by which academics and companies using Neuromarketing should abide [154]. Wilson et al. explores the impact upon the free will of consumers as Neuromarketing techniques and insights develop [243]. These examples of works exploring the privacy and moral implications within Neuromarketing provide the BHCI researcher with the foresight for considering the implications of their works to this regard. It is important that we as a community respect the privacy and rights of others, and explore the ethical and moral implications of our innovations. Whilst Neuromarketing has the potential to affect communities en-mass, through the observations derived from a small group, BHCI will (potentially) impact and be utilised by all members of a community, raising the significance of this questioning.

7.2 Novel Passive Interactions

Perhaps the primary contribution of our final study - #Scanners, is the specification of a framework that details how individuals experience these types of interactions. Here, we consider the utility of this framework - how might it be useful in the broader contexts of Multimedia, HCI and the future of BHCI. Below, we document the potential application areas which we believe can benefit from the insight we obtain from our taxonomy of control.

In providing this documentation we highlight the important contribution that the framework provides. The specification of the framework itself aims at addressing a gap in our current understanding of how to design these interactions and how they are experienced by individuals. Prior to conducting this work, the specification of these forms of interactions did not This gap formed the basis of research questions **RQ3** and **RQ4**. The discussion below applies the specification of the framework to demonstrates

7.2.1 Applying the Framework to Novel Interactions

One natural area of application for our framework is in the development of experiences similar to #Scanners - Bio-Responsive multimedia. Below we document aspects of this media that could benefit from our framework.

Design of Interactive Content

The taxonomy might be used for generating new design ideas. Here we follow Höök and Lowgren's notion of strong concept [93], a form of "intermediate design knowledge" that embodies:

- a core design idea
- bridges between specific instances and generalized theory
- concerns interactivity
- and can help generate new designs

With this in mind, we suggest three ways in which our framework might enable the design of future experiences.

> Can the structure of interactive media respond to the transitions between control, specified by the taxonomy?

In #Scanners, we documented how one film-maker scripted, shot and edited a bio-responsive film. Even this single example reveals various creative strategies, but different types of films might be suitable to different types of physiological control. Some of our participants enjoyed and preferred being below the line of control, if not unaware of it. Using measures where the form of control was ambiguous and often involuntary facilitated this lack of control. There are many design opportunities to consider for using other measures of biological response, such as heart rate, breathing, or skin temperature changes to subtly adjust aspects of the film. Suspense thrillers, for example, might monitor users for a difference between being bored and enthralled in order to shorten or extend the scene, where stress is detected from EEG and blink rates [78]. Other participants in our study utilised control, and benefited from being above the line of control. Participants actively tried to remain in scenes, or used blinking to try and avoid bits that they did not like. Similarly, designers may ask how to respond to a user that voluntarily closes their eyes for sustained periods of time. During horror films, this could be used as an active controller to skip scenes, or passively to make the film *sound* scarier. Höök et al., for example, have shown that EEG data and blink rates (also measured by #Scanners) vary according to stress levels [78].

The taxonomy contributes to detailing how experience designers can develop these forms of brain based novel interactions. We noted in our problem statement, that this form of interaction is truly novel, with few works focussing on passive forms of brain based interaction for entertainment. This lack of documented work creates a natural knowledge gap since. By providing the taxnomy to the HCI community, we set an initial foundation off of which researchers and experience designers who are interested in developing this type of interaction may establish a basis for the desing of their experience. This framework can equally be applied as a post-analysis study tool for understanding and rationalising how pariticpants responded to a given experience.

Repeat Screening/Staging of Experiences

Our recognition of inter-experience transitions suggests that more attention needs to be paid to the screening of experiences. Of particular importance here is developing strategies for moving people between different modes of engagement as they re-encounter the experience anew. Should their first experience be a "naive" one before they then find out how it learns? Should they move between spectating and driving as argued by Reeve et al. in their proposals for designing spectator interfaces [188]? What might be the best orders for combining all of these? The notion of varying repeat experiences in this is unusual in film where film-makers are not usually directly concerned whether we enjoy it differently on subsequent viewings. It is perhaps more common to consider the longevity of enjoying games, but even then the focus is more on progressing through levels than on systematically varying the experience each time.

Collective Experiences

In its current implementation #Scanners is controlled by an individual user, but films are often watched in groups, whether at the cinema or at home. Within our installation, even, the #Scanners experience can be - and was - witnessed by many viewers concurrently. This led participants to wonder what was being revealed about them

> "At first I was afraid of what people would think of what I was doing. It was the same as that - what are the people thinking I'm thinking about the stuff on the screen" - *P13*.

There is a big opportunity, therefore, to investigate how multimedia might respond to the biological responses of a collective audience. Experiences might monitor the average response across an audience, where Kirke et al., for example, have recently explored the use of audience arousal to vary a film experience with multiple possible endings [110]. Conversely, different people could control different aspects of the experience, where Leslie and Mullen, for example, provided separate control over music streams to each participant [128].

Many natural film-watching experiences, of course, are social, whether at the Cinema or at home with friends. In it's current implementation, #Scanners is controlled by an individual user. In our study, however, a number of participants were aware of other spectators in the caravan. *P10* wondered if other people would affect how well #Scanners worked for them, saying:

> "I wasn't sure if I should be reacting to that (kids making noise) around me or if I should be focused" - *P10*

Furthermore, participants wondered what the #Scanners system would reveal about them, saying

> "At first I was afraid of what people would think of what I was doing. It was the same as that - what are the people thinking I'm thinking about the stuff on the screen" - *P13*

In this regard, *P13* went on to say:

> "the film is about an Indian lad being bullied and I've got 4 Indian people sitting in and I'm starting to think is this a set up." - *P13*

This sort of concern might be more important, again, for different types of films; users might be especially concerned with what a system reveals about their levels of attention whilst watching a film containing scenes of a sexual nature. Similarly, users might worry what a system could reveal about their own romantic relationship as they watch relationships unfold in a romantic drama. Beyond considering what a film reveals about an experience driven by a solo user, there is a large opportunity to investigate how multimedia might respond to the biological responses of a collective audience. Kirke et al., for example, have recently explored the use of audience arousal to vary a film experience with multiple possible endings [110].

Alternatively, different aspects of the multimedia could also be controlled by different viewers, with sound being affected by one person, and colour by another.

Control could also be shared between participants, with control over change being given to the person who is concentrating the most. An alternative collective approach would be to externalise the bio-data of each viewer [195] within the viewing space, so that they can see if they are experiencing the film in a similar or different way to others.

Further, if different participants had different screens, the differences in their experiences could be a source of conversation and discussion about the meaning of the film, allowing people to recommend a different perspective on the scene to others.

7.2.2 Applying the Framework to BHCI

We now consider the utility of our framework - how might it be useful in the context of BHCI in general.

The first possibility is as a "sensitising concept", to assist in the design or analysis stages of future studies. We might for example, design future studies to explore in greater detail the relationships between voluntary and involuntary control or between conscious and unconscious control, or to explore some of the specific transitions that we have identified in greater depth. Our concepts might also prove useful for analysis data captured from other BHCI controlled entertainment experiences. With much other work looking at interactions via other physiological data, our framework might equally apply to other forms of broadly physiological control such as breathing and heart rate, but might also be expanded by them.

Another possibility is exploring the application of the framework to other works utilising a interaction loop similar to the Two Way Affect Loop (TWAL), described earlier. One example of such work is ExoBuilding, a physiologically driven form of adaptive architecture presented by Schnädelbach et al [194]. Through ExoBuilding, participants would be placed in a tent like structure whilst wearing a variety of physiological sensors. During the experience the lighting and actuation of the tent would dynamically adjust according to the participants' physiology - creating an interaction loop similar to what we describe with TWAL. In their discussion of the work, the authors wrote:

> "*The prototyping process also highlighted that there are different ranges of control that one might expect over one's own physiology. For example, breathing is typically controlled autonomically, but can also be controlled voluntarily (e.g. breathing exercises). EDA and heart rate are in a separate category as*

> *control is much more indirect. With experience, one might know what to do to affect the signal (running to raise heart rate or pinching oneself to raise EDA). However, it is already much harder to lower the signals or to prevent them from rising (e.g. training to avoid detection through a lie detector). Signals such as peripheral skin temperature are perhaps even much harder to control. These different ranges of control and the ways that they are brought to the attention of building inhabitants are clearly important in the design of such environments but also in the study of them.*" - **Schnädelbach et al.** [194].

This demonstrates that certain tactics were identified by the authors of this early work, with some discussion on the discovery and exertion of control being discussed. Applying our work to ExoBuilding may allow us to further detail aspects of the framework that were not elicited through #Scanners and provide a more complete modelling of this form of control. Equally, the application of the framework to the data obtained from the work of Schnädelbach et al. might allow us to categorise or explain some of the study findings in greater detail. This would allow us to observe the generality of the framework also, and it's application to different task environments and to different types of physiological data.

We can also reflect to the earlier studies presented in this thesis and see if there are aspects of the framework that apply to these studies. In both studies (TAP and LEAP), the application of BCI technology was inconsequential to the control of the task. However, we must still consider the effect of Reactivity - the psychological effects of believing that you are being observed, in this context. We can model for example the extent of the participants self-awareness during these studies. It is possible that, for example, in the LEAP study, some participants may have believed that somehow the EEG device was having an affect upon the jigsaw pieces they saw on screen. Despite not being reported directly, it cannot be discounted as a possible interfering variable in the results of the study. Similarly the task of developing a form of control could be designed and modelled using the studies framework allowing game pieces to be moved by the participants physiology - this is something we could model and design for using the framework.

Through #Scanners, we have documented yet another existing gap in our understanding of passive, brain based interactions. A significant body of existing BCI research is focussed on applying signal processing techniques to the data obtained from the device, in order to perform some kind of sense-making. In presenting #Scanners, we have presented a unique, **Novel Interaction** that focusses on understanding how users engage in passive brain based interaction. The focus of this work is upon understanding the user, how they discover, exert and relinquish control. The existing body of work does not focus on the human aspects of this work and this is an important distinction to make - understanding how users experience these interactions is vital, complimentary research in developing these experiences. We can have the suitable sense-making of the BCI signal, but without the knowledge of how to map it to the experience in an effective and engaging manner for the user, the experience breaks down. The taxonomy that we have developed from studying #Scanners will provide a foundation for other researchers to explore this space.

7.3 BCI Technology and BHCI

Through conducting the studies presented in this thesis, there have been opportunities to apply a variety of BCI technology to BHCI centred research. Each form of BCI technology comes with it's own set of considerations when applied in this context. The choice of technology plays an important factor in the design, development and results of a study, and can significantly affect the generality or validity of it's findings. This is something we intended to explore through answering **RQ1**, which attempts to address the gap in our understanding of the requirements of a BCI device for facilitating Natural Interactions. Using the experience we have gained from using a variety of devices, we discuss the current landscape of BCI technology in it's application within BHCI and detail the potential impact of future innovation in this regard.

There has been a common theme of discussion when a particular device or technology is chosen for use in a BHCI study - compromise. In our TAP study for example, fNIRS was chosen thanks to it's reputation for providing high quality data and being less sensitive to human derived artefacts, relative to other BCI technologies - such as EEG and MRI. Whilst reasonably comfortable for short-medium term use, a number of participants reported discomfort from wearing the device, with some noting slight headaches as a result of prolonged use. Equally, the device is somewhat invasive, requiring a large wired band to be secured to the participant's forehead. Additionally, for data collection purposes, a large, dedicated data collection machine is needed - significantly impacting the ecology of the study environment. The preference for high-quality data comes at the expense of the potential validity of the framing of the results in the context of the 'normal' environment in which a task might be conducted. This may prevent us from generalising the results more broadly, since we cannot discount the impact of the device upon the study results. We detailed, that our understandings over the requirements of a BCI were not well understood.

Similarly, a compromise was required in our LEAP study. The primary objective for the BCI technology used in the study was to minimally impact upon the participant during an extended period of observation. At the time, there were no fNIRS products that fulfilled this requirement. We identified consumer-grade EEG as a suitable alternative to

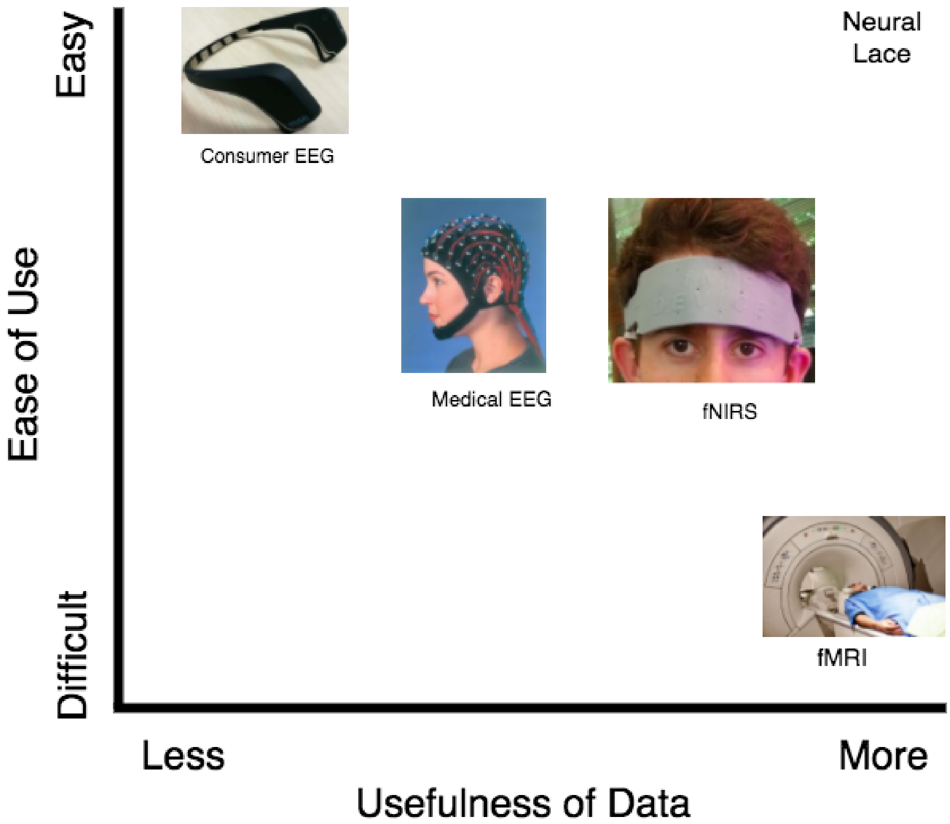


Figure 7.1: The landscape of technology choices for application within BHCI.

fNIRS. With a number of products on the marketplace providing a non-invasive, portable and comfortable operation at an affordable price point - this made consumer-grade EEG a clear candidate for application in this study. However, the immediate usefulness of the data proved to be an issue, and we were unable to quantify the same metric - Mental Workload in a reliable way using our device. This is likely a result of the quality of the components used in the device itself and the fundamental differences in the monitoring technology (optical versus electrical).

With these experiences in mind, we attempt to characterise the current landscape of available BCI technology in the context of BHCI based application. Reflecting upon our studies, we identified two classes of attributes which were primary factors in choosing a technology:

1. **Usefulness of Data** - How immediately useful is the data that we obtain from the device? Does it require significant pre-processing before it can be used? How much does human derived motion effect the quality of the data? Is it a reliable measure?

2. **Ease of Use** - How easy is it to integrate the technology into standard HCI style studies? Does it significantly impact upon the ecology of the task? Is it comfortable to wear for extended periods? Is the device portable/wireless?

Figure 7.2: The NirSport fNIRS device by NirX. Image Credit - NirX.

Using these properties as axes on a graph, in Figure 7.1 we plot the current landscape of BCI technology. This graph provides an interesting space for discussing and evaluating the current state of BCI technology in it's application of BHCI.

We see for example the effect of a competitive marketplace on EEG in this 2D space. Prior to the commercialisation, EEG was used almost exclusively in clinical or research settings, requiring high quality and accurate data with little consideration for the form factor in which the product was delivered. However, consumer grade devices are required to consider the comfort and aesthetics of the device. Consequently, in order to provide an affordable competitive product, the component quality of the devices will of a lower quality in comparison to the medical grade version. We see the effect of this in our 2D plane, with consumer grade EEG moving laterally in line with these changes.

We have yet to see the same commercialisation efforts in fNIRS that have been witnessed in EEG. There are some manufacturers in this space however that are beginning to consider the 'Ease' of which fNIRS can be applied to BHCI. The NirX Sport, shown in Figure 7.2, is one example of a more portable, high quality fNRIS device currently available on the market but is specifically for research purposes, and still not of the form factor we observe in EEG and remains prohibitively expensive (~$50,000).

Similarly, another competitor in this space Artinis has developed the Portalite, a single sensor, wireless and minimally invasive fNIRS, pictured in Figure 7.3. The properties of this device are particularly appealing to BHCI researchers due to it being minimally invasive and fairly portable, but remains expensive to purchase.

Figure 7.3: The Portalite device by Artinis. Image Credit - Artinis.

Figure 7.1 also provides us with some motivation for speculating on the potential future forms of BCI technology. There is an indication that prominent technology companies are beginning to enter the BCI space. At their 2017 F8 conference, Facebook announced that it's research and development group, Building8, had begun developing BCI's capable of allowing users 'to type with their brain'. Similarly, technology entrepreneur and innovator, Elon Musk has also announced a venture into the field with his company - Neuralink. The company is believed to be implantable BCI capable of providing ''ultra high bandwidth brain-machine interfaces to connect humans and computers.". With these large companies beginning to enter the marketplace, it will be interested to observe the movement and additions of technology plotted on the graph in Figure 7.1. We suspect that an ideal will someday be reached, with the balance between usefulness of data and ease of use being catered for by some currently unknown BCI technology.

Whilst a number of the findings we document through our TAP and LEAP studies, regarding suitability of various BCI's may appear to be intuitive, none of these findings are formally documented in the existing literature, especially for tasks of extended periods. It was for this reason (in accordance with our literature review), that we developed **RQ1** - to address this current gap in our understanding of BCI technology. In documenting the feedback from participants in this manner we provide the community with guidance on how to chose a suitable BCI device in accordance to the requirements of their study. This is further facilitated by Figure 7.1 presented above, and should hopefully encourage other BHCI researchers to documented how successfully a particular technology was applied to a particular study task and environment.

7.4 Future Work

7.4.1 Application of Machine Learning for Data Analysis

In our LEAP study, we were unable to identify significant differences in Mental Workload (MWL) between conditions using the EEG data alone. The reasons for this could either be the quality of the data we obtain from the device is not sufficient to identify these changes or, our method of analysing the data is too simple to uncover these variations. It's possible therefore that further analysis of the EEG data using more sophisticated data analysis techniques would help identify these differences in MWL. If possible, this would provide BHCI researchers with an ideal scenario whereby they can use affordable, non-invasive and portable EEG devices whilst obtaining a reliable quantified measure for observing the participant.

To identify these variations we will investigate the application of Machine Learning (ML) as a tool in our analysis of BHCI study data. As we documented in our Literature Review, a significant amount of work in the field of Neuroimaging, is already investigating the application of ML in this capacity. A lot of this work however is focussed on traditional applications including aiding diagnosis stages of medical procedures and facilitating actuated movements of prosthetic limbs. Additionally, given the context of the existing research area, medical grade devices are typically the ones investigated in this context.

In our investigation however, we will focus directly on the application of these approaches in the quantification of the participants mental state, using devices that appear in the top right quadrant of our technology chart (Figure 7.1). This quadrant of the chart provides us with the ideal combination of both ease of use of the BCI technology (non-invasive and portable) whilst equally providing useful data for the purposes of BHCI centred research.

There are some existing examples of BHCI work utilising ML for quantifying participants MWL in the literature. Peck et al. for example utilised a Support Vector Machine (SVM) to classify fNIRS data as an indicator of user preference in a movie recommendation system [171]. Similarly, Afergan used an SVM to develop PHYLTR, a realtime classification framework capable of quantifying fNIRS derived brain data [4]. We will look to expand upon these works and investigate the general application of these ML techniques for BCI technology.

7.4.2 Cognitively Queryable Video (CQV)

One aspect of BHCI centred research we are currently exploring is the fusion of brain derived data as a queryable metric for the automatic production of recorded video.

Best explained through an example; consider a long (multi-hour) user study where a user interacts with a new version of a search interface and asked to perform a number of information retrieval based tasks using the new interface. The aim of the researchers conducting the study is to identify the 'uncomfortable' parts of the user's interaction with the Search User Interface (SUI) - the moments of friction, where the users experiences a negative interaction. A screen recording of the user's interaction with the SUI is captured as a part of the study, as is the user's 'cognitive status' using some commercially available and non-invasive brain monitoring device, in a BHCI centred approach we have described throughout this thesis.

With CQV, we propose that the collected brain data can be used as a queryable index against which the video can be automatically edited and queried. Continuing with the example above, assume through the data obtained from the brain monitoring device that we are able to obtain a measure of frustration from the user as they complete the study task. Using this index of frustration, a researcher can automatically (with some simple scripting), reduce a multi-hour recording to the "moments of interest" where frustration peaks - an indicator that something in the SUI perhaps hasn't behaved in a manner the user was expecting, or was perhaps fundamentally broken.

With this idea in mind, we saw an opportunity to trial such a system. On a recent, 4 hour (approx) drive from Manchester, England to Glasgow, Scotland - we decided to test the proposed CGV approach. We collected the following data during the duration of the drive:

1. **Neurosky EEG** - EEG data was recorded from the driver for the entire duration of the journey. The Neurosky Mindwave Mobile (as used in #Scanners), provided a safe, non-invasive, wireless and comfortable measure of EEG for the extended duration of the recording. Data from the device was recorded to a laptop, controlled by the passenger.

2. **Front Facing Video Capture** - A web-cam, secured to the dashboard of the vehicle, capturing a view of the road ahead, at 5 FPS. This was installed to identify what was happening on the road itself and attribute the on-road activity to variations in the driver's brain data. For example, a car suddenly pulling out into the drivers lane may have a significant effect on the driver's attention and frustration levels.

Figure 7.4: A frame from the video recording during the 4 hour drive.

3. **Driver Facing Video Capture** - A web-cam, secured to the dashboard of the vehicle, capturing a view of the driver, at 5 fps. This was installed to capture the driver's in-car experience, again with the aim of attributing events to variations in the driver's brain data. For example, if the driver stalls or is unable to 'find' the correct gear, the drivers attention/frustration/meditation levels may be affected accordingly.

4. **Mobile Phone based Accelerometer and GPS** - A continual recording of the acceleration/braking (250Hz) and GPS location (1HZ) of the vehicle was recorded. The data was ultimately not used in this trial usage, but future applications of this approach in a driving context could use this data for visualisation or as a means of statistic attribution e.g. Frustration is strongly linked to heavy braking.

Data collection was monitored by the passenger to ensure that good contact quality was maintained between the EEG headset and the driver's forehead. The webcams were attached to, and powered by, the in-car laptop. The Neurosky was battery powered and data was transmitted via Bluetooth.

Having collected the 4 hours of brain and video data, we wrote a simple Python script that would allow us to query the brain data and edit the video accordingly.

Using the CGV approach in this study allowed us to reduce an otherwise unusable 4 hours of video, down into a collection of 'cuts' where the driver's brain activity was within a prescribed limit. Below we provide links to some the CGV output generated from the data collected in this trial study. We provide some supporting **descriptive** observations of what is featured in the generated video. It is important for us to note that these are just observations, and not statistically backed **explanations** of what we're showing through this approach. Due to the nature and timing of the study, we have not yet performed a full statistical investigation to confirm (or disprove) these observations. To do this, a detailed labelling of the original video would need to be performed in order to identify what the CGV produced video contained and whether it was linked to the brain data in a statistically robust and significant manner.

- **Attention >= 98(/100)** - In this investigation we were interested in identify what caused the driver's attention (as measured by the Neurosky) to spike into it's top 2 percentile range. We note that the video produced using CGV for this query seems to exclusively be comprised of moments where the driver was performing an overtaking manoeuvre. This is an intuitive result as we would expect (rather - hope) that the driver is paying particular attention to the road and their surroundings when performing an overtake manoeuvre.
  - Available at https://www.youtube.com/watch?v=nQLJZb-sxdU
- **Attention <= 10(/100)** - For this query, we were keen to understand when the driver's attention was low (bottom 10 percentile), and attempt to understand why this might be (based on the generated video content). The generated video contains a lot of 'inside lane' driving with little to no cars nearby - a clear road. Equally, we note that a number of the scenes in the generated video feature the driver talking to the passenger - indicating that the device is possibly sensitive to verbal artefacts (this needs further investigation).
  - Available at https://www.youtube.com/watch?v=rG8lqPflu9M
- **Meditation >= 97(/100)**
  - Available at https://www.youtube.com/watch?v=elqGK3eqKnM
- **Meditation <= 10(/100)**
  - Available at https://www.youtube.com/watch?v=eNsZuP308ME

We are continuing to investigate this work and the data obtained during this study. We are keen to perform the necessary labelling and statistical analysis of the video to be able to gain a deeper appreciation for what is precisely being generated through this CGV approach.

7.4.3 Integrating BHCI into VR

The recent trend and parallel development/adoption of Virtual Reality, Brain Sensing Measures and associated technology such as Augmented Reality by large corporations, and the rise in the interests in the consumer market have set a positive tone for research in these disciplines. An important Human Factors area that is a catalyst to broad VR applications is the measure of perception, mental workload, and immersion amongst other issues, which are determining factors in the experience of using virtual environments. Traditional approaches in studying these issues use well-developed subjective measures via questionnaires. Through BHCI, we could potentially provide an objective approach in resolving many subjective uncertainties amongst other prospects. With this proposed future work we discuss the integration of these two emerging fields in order to provide a continuous, objective, physiological measure of an individual's VR experience for the purposes of enhancing user experience and improving user performance. The aim of this future work is to merge two complementary fields of work, and investigate the implications which could potentially open up avenues of research which were traditionally difficult due to the limitations of equipment, or the lack of a quantified approach.

Relating quantitative data from brain-monitoring devices into feedback about a VR experience is one of our ultimate goals in conducting this research. VR is an inherently complex, multidimensional experience that affects multiple regions of the human brain. In work conducted prior to that presented in this thesis, we investigated the challenges of using brain-monitoring technologies to evaluate IIR tasks: that tasks have different stages, that behaviour quickly diverges after the first interaction (and thus is hard to compare), and that brain measurements vary dramatically over time [177]. Thankfully, there is a significant corpus of work exploring this problem space in the field of Psychology and we are able to adapt approaches described in the literature to suit our own needs. As we have detailed in Related Work section and the work we present in the main body of this thesis, the HCI community provides a template for this interaction of techniques. The benefits of this integration can be significant however. Existing approaches such as Multidimensional ratings of cognitive load (NASA-TLX, SWAT, and Workload Profile) can be, by itself, a difficult task to accomplish on top of the experimental activities. Experimental subjects are required to reflect on their MWL as a means of capturing subjective experience. Even though this is done immediately after the experiment, the need to reflect and recall the experience can make the self-evaluation inaccurate, as well as impacting on the nature of the task. Brain based measures can help in real-time performance monitoring and the observations of cognitive load at the time of the activity, and can act as an objective evaluation of the user's subjective evaluation. As a result, the inaccuracy in the correlation between the MWL and the performance that often plagued such experiments may be better monitored. This will also result in improved ecological validity, as the user will not be required to leave the immersion of a particular experience in order to complete the subjective questionnaire. The use of BHCI approach will also allow for the fine grained analysis of experiences on a second-by-second basis – a significant benefit over existing subjective approaches, whose results are indicative of the experience in general. Through this continuous, brain derived measure, researchers will be able to pin-point good and bad moments of an experience and adapt it as necessary. Achieving this with existing approaches would require extensive post experience interview sessions and these would still not provide the granularity that we can achieve with a brain based measure.

Below are some potential outputs of this future integration of technologies:

Task-based training – We believe the fusion of these two technologies will provide a powerful training system for physical tasks that require coordination or skill development in some way. An example we are actively exploring is the development of work or skills requiring the coordination of the hands and fingers, and even the entire physical body with a trained mental capacity for control of tools (e.g., sword work in martial arts, cultural dance, and etc). The practice requires extensive training and refinement to an individual's breathing and swing technique. VR will be able to provide real-time corrective feedback with regards to these two aspects of the skill, while the BCI input will aid in the adaptation of content delivery with regards to difficulty and training elements according to the individual's current workload levels. The key benefit here is the maximal performance of the individuals learning experience without reaching a state of overload on the user's part

The design of good experience – There are a large number of VR applications available in the marketplace today, and VR entertainment usage is on the rise with the porting of VR to mobile phones and headsets (e.g., Samsung Gear VR), but little to no work has been conducted in evaluating their effectiveness in terms of engaging the user on a cognitive level. We're especially interested in experiences that evoke particular emotions (e.g. horror or hero based experiences), workload (e.g. Flight simulator experiences) or training (e.g. Brain training) in users since these induce specific cognitive reactions which will be quantified in the BCI based recording. Conducting such experiments will allow us to compare between experiences and identify their effectiveness.

But the integration of these technologies are not without their challenges. One clear practical limitation in this proposed approach is the physical 'real-estate' available on a user's head. Regions of interest within the brain vary depending on research interests, but typically, researchers are interested in evaluating Working Memory, emotions, decision-making and Mental Workload. These activities primarily occur within the Pre-Frontal Cortex, an area of the brain located directly behind the forehead [11]. For this reason, a number of brain sensing devices sit directly on the forehead, as shown in Figure 1. The issue here lies in the physical placement of the brain measure whilst simultaneously wearing VR headsets, which occupy the same region of the forehead.

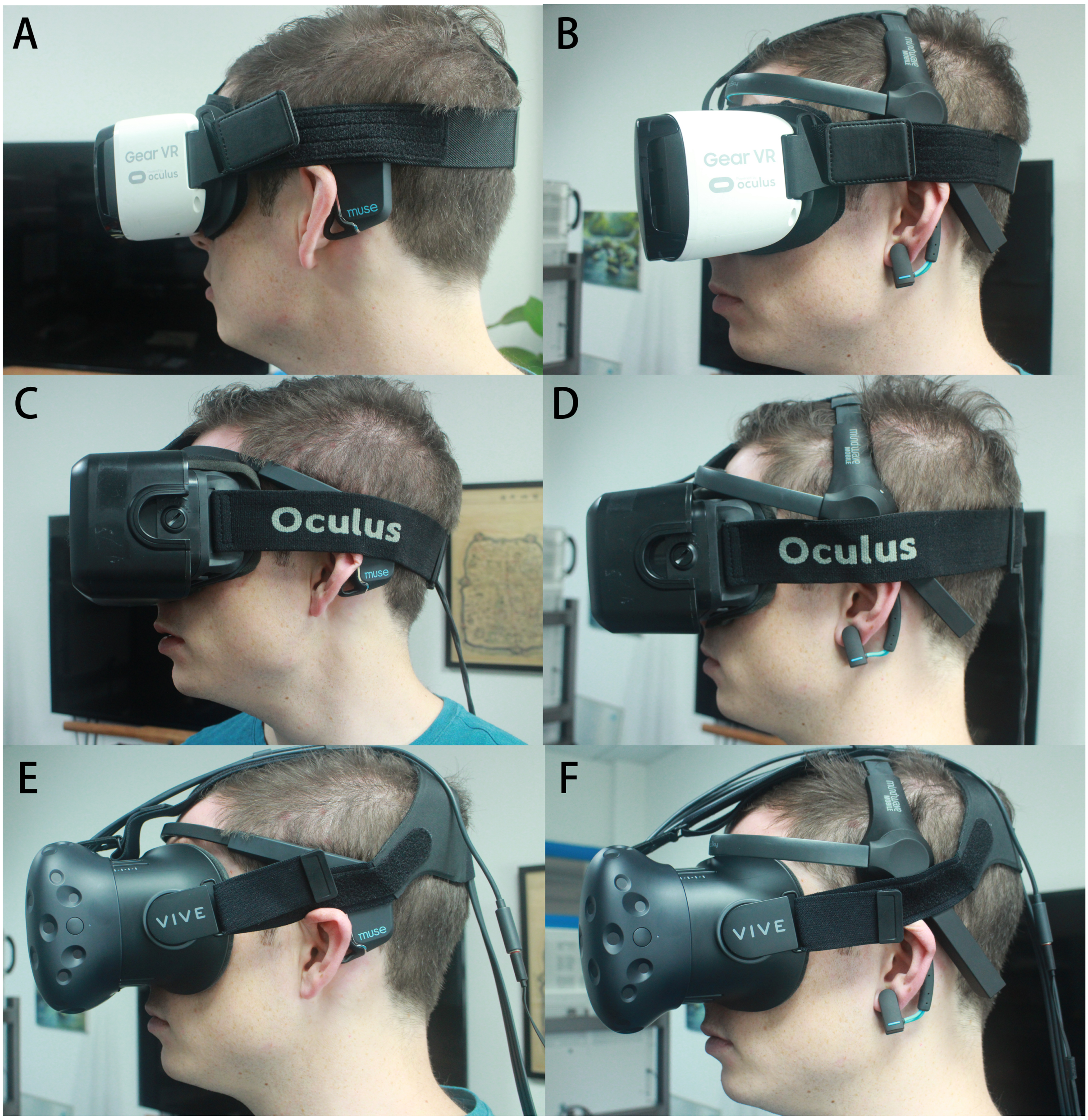


Figure 7.5: Demonstration of the practicality of 2 types of BCI devices with 3 types of VR goggles

Figure 7.5 demonstrates a participant wearing both VR goggles and a portable, EEG based BCI device, simultaneously. This two-sided design issue raises an interesting opportunity for innovation and product differentiation in the VR Headset market place. Since we see that a number of existing devices occupying the wearer's forehead, it comes at no additional cost to integrate a physiological sensor into the VR Headset itself. Doing so would expose this new VR system to the broad range of benefits, including the evaluation approach described here. Such a product, to the best of our knowledge, is not currently available in the marketplace. The integration of the brain measurement into a VR device would also help control or mitigate the issues of electrical interference upon EEG based measures, originated from the electrical components in the VR device.

Chapter 8

Conclusion

8.1 Summary of Work

Through the work presented in this thesis we have documented our exploration into the research field of Brain based Human Computer Interaction (BHCI). Specifically, we sought to investigate the application of BHCI to facilitating lightweight forms of natural user interaction. To achieve this, we explored the application of two BCI technologies (fNIRS and EEG), and evaluated their suitability for developing these forms of interactions.

In our initial study, we explored the application of fNIRS, an optical brain-sensing technology, that has shown to be well suited for use in HCI settings by Solovey et.al [210]. Developing on the initial work set out by Solovey et.al, we designed, developed and executed a study exploring the use of verbal protocols within HCI studies and their cognitive impact upon the participant. In doing so, we sought to seek insight into the first of four research questions:

> **RQ1**. What are the characteristics of a suitable BCI technology for supporting natural forms of interaction?

Through the application of fNIRS (in combination with other MWL based measures) we were able to demonstrate - quantitatively, that a widely used HCI evaluation protocol - "Think Aloud Protocol", did not significantly interfere with the participant's MWL under study conditions, as long as the verbalisations were related to the study task. However, in conducting the study, we also identified that our fNIRS device did not provide us with the ideal properties for supporting forms of natural interactions. Specifically, we identified that the fNIRS was uncomfortable after extended periods of use and many participants noted some discomfort/tension as a result of wearing the device. We, as researchers, also identified that the device was fairly non-portable, and the requirement of wired connectivity would significantly impede upon potential natural interaction experiences. It was for these reasons, that we decided to explore an alternative BCI technology.

In our second study, we continued to investigate research question **RQ1**. We applied a consumer grade, lightweight, wireless EEG brain-sensing device to evaluate the effect of using 3D input modalities upon a participants cognition, relative to more traditional/familiar forms of input. To perform this work, we developed a study using a Jigsaw puzzle as the study task, which participants would solve using: 1) AR input, 2) Mouse input and 3)Physical pieces. The results indicated that the EEG sensor was not as sensitive in detecting variations of MWL, but did provide the desired properties for developing natural forms of interactions. On this basis, we reasoned that MWL was one of many potential metrics which we could obtain from EEG, and decided that the lightweight, portable properties of the device warranted further exploration in this space.

Our final study would build off the work performed in answering **RQ1** through conducting the previous two studies, and we would apply this to answering the remaining three research questions:

> **RQ2**. How can BHCI be used to develop natural forms of indirect control?
>
> **RQ3**. How are these natural forms of indirect control experienced by the users?
>
> **RQ4**. What design considerations must we make when developing indirect natural interactions using BHCI?

To investigate these research questions we partnered with an artist/film-producer to develop a new form of cinematography utilising BHCI and natural forms of interactions to produce a novel new form of cinema. #Scanners was a project investigating the effects of a two-way affect loop (TWAL) upon the viewer. TWAL is an cinematic experience where the film is affecting the viewer and the viewer's physiology is affecting the composition of the experience in real-time (**RQ2**). We studied the effect of using TWAL through a research-in-the-wild methodology. The free-form approach of the methodology undertaken in this study allowed us to identify how the stages of experience were experienced by the viewers. Specifically, we learnt how people discovered forms of control (if they did at all), how they chose to exert control after discovery, the extent to which they consciously exerted control over the experience. From these observations, we were able to develop a taxonomy of control which we contribute so that other experience designers interested in developing a BHCI based form of natural control may use (**RQ3**).

Finally, we reflect on the 3 studies to provide answers to our final research question **RQ4**. Through our initial 2 studies we identified that the current form factor, not the fundamental technology itself, was the reason fNIRS is not currently well suited to developing these types of experience. Consumer grade EEG has benefited from a number of companies competing in the marketplace to provide more affordable, compact and innovative BCI devices. fNIRS has yet to benefit from this form of competition, but may in the future. Fundamentally therefore, we identified that the properties of a device being lightweight, portable and comfortable to wear for extended periods were critically important in considering BCI technology for this type of application. We also learned that this applied form of BHCI provides an interesting and engaging creative space in which interesting, novel new experiences can be developed (**RQ4**).

8.2 Key Contributions

In conducting this body of work we believe that we have provided a significant contribution to the field of HCI. Below we detail these contributions.

C1 **- Taxonomy of Control for BHCI based Natural Interaction**

Through a research-in-the-wild based research methodology, we were able to identify how viewers discovered and exerted control over an interactive cinematic experience controlled via BHCI. This led us to propose a two-dimensional taxonomy of control, considering both the understanding of the control, and how much users think about control. This taxonomy plots how and when viewers discovered control and how they chose to exert or relinquish control (consciously or otherwise) as the experience progressed. The taxonomy provides a concrete foundation off which interaction designers can develop, utilising the 'creative spaces' identified by the taxonomy to provide engaging experiences.

C2 **- Demonstrated the application of BHCI for Evaluation**

A significant body of existing work into BCI and towards applications of BHCI have focussed on the modelling of Mental Work and Emotion, but few have provided concrete applications of BHCI driven experiences. Through the three studies presented in this thesis, we believe that a broad array of BHCI techniques, technologies and takeaways have been contributed to the community. We believe there is significant importance in contributing examples of applied BHCI in the continuation of it's future development and adoption by other researchers.

C3 **- Verification of Think Aloud Protocols via BHCI based Evaluation**

Think Aloud Protocols are a recognised research tool designed to elicit insight into participants thought processes and decision-making as they complete a task. The protocol has been utilised in thousands of studies across a number of research fields, including HCI and Human Factors. Identifying the validity and cognitive impact of such an important and widely used protocol is therefore a significant contribution to the community. A collection of psychophysiological evaluations of the protocol exist in the literature. For example, Hertzum et al. measured eye movement under different levels of verbalisations and identified differences between 'classic' and 'relaxed' thinking aloud [84]. However, to the best of our knowledge, this is the first, direct-brain based evaluation of TAP's that has been documented in the literature. Our findings indicate that, the act of verbalising during a task does not significantly affect MWL accept when the verbalisations do not relate to the task at hand e.g. repeating 'blah blah blah' whilst completing a task.

References

[1] Abujelala, M., Abellanoza, C., Sharma, A., and Makedon, F. Brain-ee: Brain enjoyment evaluation using commercial eeg headband. *Proceedings of the 9th acm international conference on pervasive technologies related to assistive environments*, ACM (2016), 33.

[2] Adamczyk, P.D. and Bailey, B.P. If not now, when?: The effects of interruption at different moments within task execution. *Proceedings of the sigchi conference on human factors in computing systems*, ACM (2004), 271–278.

[3] Adhikarla, V.K., Sodnik, J., Szolgay, P., and Jakus, G. Exploring direct 3d interaction for full horizontal parallax light field displays using leap motion controller. *Sensors 15*, 4 (2015), 8642–8663.

[4] Afergan, D., Hincks, S.W., Shibata, T., and Jacob, R.J. Phylter: A system for modulating notifications in wearables using physiological sensing. *International conference on augmented cognition*, Springer (2015), 167–177.

[5] Afergan, D., Peck, E.M., Solovey, E.T., et al. Dynamic difficulty using brain metrics of workload. *Proceedings of the 32nd annual acm conference on human factors in computing systems*, ACM (2014), 3797–3806.

[6] Afergan, D., Shibata, T., Hincks, S.W., et al. Brain-based target expansion. *Proceedings of the 27th annual acm symposium on user interface software and technology*, ACM (2014), 583–593.

[7] Anderson, E.W., Potter, K., Matzen, L., Shepherd, J., Preston, G., and Silva, C. A user study of visualization effectiveness using EEG and cognitive load. *Computer Graphics Forum 30*, 3 (2011), 791–800.

[8] Andrillon, T., Nir, Y., Cirelli, C., Tononi, G., and Fried, I. Single-neuron activity and eye movements during human rem sleep and awake vision. *Nature communications 6*, (2015).

[9] Astolfi, L., Vecchiato, G., De Vico Fallani, F., et al. The track of brain activity during the observation of tv commercials with the high-resolution eeg technology. *Computational intelligence and neuroscience 2009*, (2009).

[10] Backs, R.W., Ryan, A.M., and Wilson, G.F. Psychophysiological measures of workload during continuous manual performance. *Human Factors: The Journal of the Human Factors and Ergonomics Society 36*, 3 (1994), 514–531.

[11] Baddeley, A. Working memory. *Science 255*, 5044 (1992), 556–559.

[12] Baddeley, A. The episodic buffer: A new component of working memory? *Trends in cognitive*

sciences 4, 11 (2000), 417–423.

[13] Baddeley, A.D. Is working memory still working? *European psychologist 7*, 2 (2002), 85–97.

[14] Baddeley, A.D. and Hitch, G. Working memory. *The psychology of learning and motivation 8*, (1974), 47–89.

[15] Bassily, D., Georgoulas, C., Guettler, J., Linner, T., and Bock, T. Intuitive and adaptive robotic arm manipulation using the leap motion controller. *ISR/robotik 2014; 41st international symposium on robotics; proceedings of*, VDE (2014), 1–7.

[16] Beatty, J. and Lucero-Wagoner, B. The pupillary system. *Handbook of psychophysiology 2*, (2000), 142–162.

[17] Becker, A.B., Warm, J.S., Dember, W.N., and Hancock, P. Effects of jet engine noise and performance feedback on perceived workload in a monitoring task. *the International journal of aviation psychology 5*, 1 (1995), 49–62.

[18] Benbadis, S.R. and Rielo, D. EEG artifacts. *eMedicine Neurology, Mar*, (2010).

[19] Benford, S., Greenhalgh, C., Crabtree, A., et al. Performance-led research in the wild. *ACM Transactions on Computer-Human Interaction (TOCHI) 20*, 3 (2013), 14.

[20] Benford, S., Greenhalgh, C., Giannachi, G., Walker, B., Marshall, J., and Rodden, T. Uncomfortable interactions. *Proceedings of the sigchi conference on human factors in computing systems*, ACM (2012), 2005–2014.

[21] Binder, J.R., Frost, J.A., Hammeke, T.A., Cox, R.W., Rao, S.M., and Prieto, T. Human brain language areas identified by functional magnetic resonance imaging. *Journal of Neuroscience 17*, 1 (1997), 353–362.

[22] Birbaumer, N. Breaking the silence: Brain–computer interfaces (bci) for communication and motor control. *Psychophysiology 43*, 6 (2006), 517–532.

[23] Birbaumer, N., Ghanayim, N., Hinterberger, T., et al. A spelling device for the paralysed. *Nature 398*, 6725 (1999), 297–298.

[24] Birbaumer, N., Murguialday, A.R., Wildgruber, M., and Cohen, L.G. Brain-computer-interface (bci) in paralysis. In *The european image of god and man.* Brill, 2010, 483–492.

[25] Bos, D.P.-O., Reuderink, B., Laar, B. van de, et al. In *Brain-computer interfaces.* Springer, UK, 2010, 149–178.

[26] Bracegirdle, A., Mitrovic, S., and Mathews, M. Investigating the usability of the leap motion controller: Gesture-based interaction with a 3d virtual environment. (2014).

[27] Brodmann, K. *Vergleichende lokalisationslehre der grosshirnrinde in ihren prinzipien dargestellt auf grund des zellenbaues.* Barth, 1909.

[28] Brouwer, A.-M., Van Wouwe, N., Mühl, C., Erp, J. van, and Toet, A. *Perceiving blocks of emotional pictures and sounds: Effects on physiological variables.* 2013.

[29] BRUNNER, D., VASKO, R., DETKA, C., MONAHAN, J., REYNOLDS III, C., and KUPFER, D.

Muscle artifacts in the sleep eeg: Automated detection and effect on all-night eeg power spectra. *Journal of sleep research 5*, 3 (1996), 155–164.

[30] Brünken, R., Steinbacher, S., Plass, J.L., and Leutner, D. Assessment of cognitive load in multimedia learning using dual-task methodology. *Experimental psychology 49*, 2 (2002), 109.

[31] Cabeza, R. and Nyberg, L. Imaging cognition ii: An empirical review of 275 pet and fMRI studies. *Journal of cognitive neuroscience 12*, 1 (2000), 1–47.

[32] Campbell, Z., Zakzanis, K.K., Jovanovski, D., Joordens, S., Mraz, R., and Graham, S.J. Utilizing virtual reality to improve the ecological validity of clinical neuropsychology: An fmri case study elucidating the neural basis of planning by comparing the tower of london with a three-dimensional navigation task. *Applied Neuropsychology 16*, 4 (2009), 295–306.

[33] Carter, T., Seah, S.A., Long, B., Drinkwater, B., and Subramanian, S. UltraHaptics: Multi-point mid-air haptic feedback for touch surfaces. *Proceedings of the 26th annual acm symposium on user interface software and technology*, ACM (2013), 505–514.

[34] Castellanos, C. Biomorphic aggregator. 2007.

[35] Chambayil, B., Singla, R., and Jha, R. EEG eye blink classification using neural network. *Proceedings of the world congress on engineering*, (2010), 2–5.

[36] Chance, B., Anday, E., Nioka, S., et al. A novel method for fast imaging of brain function, non-invasively, with light. *Optics express 2*, 10 (1998), 411–423.

[37] Chandler, P. and Sweller, J. Cognitive load theory and the format of instruction. *Cognition and instruction 8*, 4 (1991), 293–332.

[38] Charters, E. The use of think-aloud methods in qualitative research an introduction to think-aloud methods. *Brock Education*, (2013).

[39] Chen, D. and Vertegaal, R. Using mental load for managing interruptions in physiologically attentive user interfaces. *CHI'04 extended abstracts on human factors in computing systems*, ACM (2004), 1513–1516.

[40] Cherng, F.-Y., Lin, W.-C., King, J.-T., and Lee, Y.-C. An eeg-based approach for evaluating graphic icons from the perspective of semantic distance. *Proceedings of the 2016 chi conference on human factors in computing systems*, ACM (2016), 4378–4389.

[41] Cobb, W. The normal adult eeg. In *Electroencephalography.* Macmillan, New York, 1963, 232–249.

[42] Cronin, D. Usability of micro-vs. macro-gestures in camera-based gesture interaction. 2014.

[43] Csikszentmihalyi, M. and Csikzentmihaly, M. *Flow: The psychology of optimal experience.* Harper-Perennial New York, 1991.

[44] Curtis, C.E. and D'Esposito, M. Persistent activity in the prefrontal cortex during working memory. *Trends in cognitive sciences 7*, 9 (2003), 415–423.

[45] Dasari, D., Shou, G., and Ding, L. ICA-derived eeg correlates to mental fatigue, effort, and workload

in a realistically simulated air traffic control task. *Frontiers in Neuroscience 11*, (2017), 297.

[46] Devices LLC. *FNIR imager & cobi studio manual.* Biopac, 2013.

[47] Doud, A.J., Lucas, J.P., Pisansky, M.T., and He, B. Continuous three-dimensional control of a virtual helicopter using a motor imagery based brain-computer interface. *PloS one 6*, 10 (2011), e26322.

[48] Doughty, M.J. Further assessment of gender-and blink pattern-related differences in the spontaneous eyeblink activity in primary gaze in young adult humans. *Optometry & Vision Science 79*, 7 (2002), 439–447.

[49] Dumas, J. and Redish, J. *A practical guide to usability testing.* Intellect Books, 1999.

[50] Economo, C.F. von and Koskinas, G.N. *Die cytoarchitektonik der hirnrinde des erwachsenen menschen.* J. Springer, 1925.

[51] Ehrler, F., Mittempergher, C., Duret, C., Daehne, P., and Lovis, C. Exploratory usability comparison of three interaction paradigms for touchless spatio-temporal manipulations of 3D images using leap motion. *Swiss Medical Informatics 31*, (2015).

[52] Engle, R.W., Kane, M.J., and Tuholski, S.W. Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence, and functions of the prefrontal cortex. (1999).

[53] Ericsson, K. and Simon, H. *Protocol analysis: Verbal reports as data.* MIT Press, 1984.

[54] Esposito, F., Aragri, A., Piccoli, T., Tedeschi, G., Goebel, R., and Di Salle, F. Distributed analysis of simultaneous eeg-fMRI time-series: Modeling and interpretation issues. *Magnetic resonance imaging 27*, 8 (2009), 1120–1130.

[55] Fadiga, L., Craighero, L., and D'Ausilio, A. Brocas area in language, action, and music. *Annals of the New York Academy of Sciences 1169*, 1 (2009), 448–458.

[56] Falcao, C., Lemos, A.C., and Soares, M. Evaluation of natural user interface: A usability study based on the leap motion device. *Procedia Manufacturing 3*, (2015), 5490–5495.

[57] Fere, C. Note on changes in electrical resistance under the effect of sensory stimulation and emotion. *Comptes rendus des seances de la societé de biologie 5*, (1888), 217–219.

[58] Fink, A., Grabner, R., Neuper, C., and Neubauer, A. EEG alpha band dissociation with increasing task demands. *Cognitive brain research 24*, 2 (2005), 252–259.

[59] Fischer, J. Understanding receptivity to interruptions in mobile human-computer interaction. 2011.

[60] Fletcher, P., Shallice, T., Frith, C., Frackowiak, R., and Dolan, R. The functional roles of prefrontal cortex in episodic memory. ii. retrieval. *Brain 121*, 7 (1998), 1249–1256.

[61] Folgieri, R. and Zampolini, R. BCI promises in emotional involvement in music and games. *Computers in Entertainment (CIE) 11*, 4 (2015), 4.

[62] Gabrieli, J.D., Poldrack, R.A., and Desmond, J.E. The role of left prefrontal cortex in language and memory. *Proceedings of the national Academy of Sciences 95*, 3 (1998), 906–913.

[63] Gallese, V., Fadiga, L., Fogassi, L., and Rizzolatti, G. Action recognition in the premotor cortex.

Brain 119, 2 (1996), 593–609.

[64] Gao, X., Xu, D., Cheng, M., and Gao, S. A bci-based environmental controller for the motion-disabled. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on 11*, 2 (2003), 137–140.

[65] Garcia-Molina, G., Tsoneva, T., and Nijholt, A. Emotional brain–computer interfaces. *International journal of autonomous and adaptive communications systems 6*, 1 (2013), 9–25.

[66] Gerjets, P., Walter, C., Rosenstiel, W., Bogdan, M., and Zander, T.O. Cognitive state monitoring and the design of adaptive instruction in digital environments: Lessons learned from cognitive workload assessment using a passive brain-computer interface approach. *Frontiers in neuroscience 8*, (2014).

[67] Ginns, P. Integrating information: A meta-analysis of the spatial contiguity and temporal contiguity effects. *Learning and Instruction 16*, 6 (2006), 511–525.

[68] Girouard, A. Adaptive brain-computer interface. *CHI'09 extended abstracts on human factors in computing systems*, ACM (2009), 3097–3100.

[69] Girouard, A., Hirshfield, L.M., Solovey, E., and Jacob, R.J. Using functional near-infrared spectroscopy in hci: Toward evaluation methods and adaptive interfaces. *Proc. chi 2008 workshop on brain-computer interfaces for hci and games*, (2008).

[70] Girouard, A., Solovey, E.T., and Jacob, R.J. Designing a passive brain computer interface using real time classification of functional near-infrared spectroscopy. *International Journal of Autonomous and Adaptive Communications Systems 6*, 1 (2013), 26–44.

[71] Girouard, A., Solovey, E.T., Hirshfield, L.M., et al. From brain signals to adaptive interfaces: Using fNIRS in hci. In *Brain-computer interfaces*. Springer, 2010, 221–237.

[72] Glover, G.H., Li, T.-Q., and Ress, D. Image-based method for retrospective correction of physiological motion effects in fMRI: RETROICOR. *Magnetic resonance in medicine 44*, 1 (2000), 162–167.

[73] Goodman, J., Brewster, S., and Gray, P. Using field experiments to evaluate mobile guides. *Proceedings of hci in mobile guides, workshop at mobile hci*, (2004).

[74] Grafton, S.T., Fadiga, L., Arbib, M.A., and Rizzolatti, G. Premotor cortex activation during observation and naming of familiar tools. *Neuroimage 6*, 4 (1997), 231–236.

[75] Grierson, M. and Kiefer, C. Better brain interfacing for the masses: Progress in event-related potential detection using commercial brain computer interfaces. *CHI '11 extended abstracts on human factors in computing systems*, ACM (2011), 1681–1686.

[76] Guna, J., Jakus, G., Pogačnik, M., Tomažič, S., and Sodnik, J. An analysis of the precision and reliability of the leap motion sensor and its suitability for static and dynamic tracking. *Sensors 14*, 2 (2014), 3702–3720.

[77] Gürkök, H., Plass-Oude Bos, D., Laar, B., Nijboer, F., and Nijholt, A. User experience evaluation in bci: Filling the gap. *International Journal of Bioelectromagnetism 13*, 1 (2011), 54–55.

[78] Haak, M., Bos, S., Panic, S., and Rothkrantz, L. Detecting stress using eye blinks and brain activity from eeg signals. *Proceeding of the 1st driver car interaction and interface (DCII 2008)*, (2009).

[79] Hari, R., Forss, N., Avikainen, S., Kirveskari, E., Salenius, S., and Rizzolatti, G. Activation of human

primary motor cortex during action observation: A neuromagnetic study. *Proceedings of the National Academy of Sciences 95*, 25 (1998), 15061–15065.

[80] Hart, S.G. and Staveland, L.E. Development of nasa-tlx (task load index): Results of empirical and theoretical research. *Human mental workload*, (1988).

[81] Haskins, C. Making the leap from traditional input devices: An evaluation of leap motion. (2016).

[82] Hasson, U., Landesman, O., Knappmeyer, B., Vallines, I., Rubin, N., and Heeger, D.J. Neurocinematics: The neuroscience of film. *Projections 2*, 1 (2008), 1–26.

[83] Haxby, J.V., Horwitz, B., Ungerleider, L.G., Maisog, J.M., Pietrini, P., and Grady, C.L. The functional organization of human extrastriate cortex: A pet-rCBF study of selective attention to faces and locations. *Journal of Neuroscience 14*, 11 (1994), 6336–6353.

[84] Hertzum, M., Hansen, K.D., and Andersen, H.H.K. Scrutinising usability evaluation: Does thinking aloud affect behaviour and mental workload? *Behav. Inf. Technol. 28*, 2 (2009), 165–181.

[85] Hill, D. and Keynes, R. Opacity changes in stimulated nerve. *The Journal of physiology 108*, 3 (1949), 278.

[86] Hillard, B., El-Baz, A.S., Sears, L., Tasman, A., and Sokhadze, E.M. Neurofeedback Training Aimed to Improve Focused Attention and Alertness in Children With ADHD A Study of Relative Power of EEG Rhythms Using Custom-Made Software Application. *Clinical EEG and neuroscience 44*, 3 (2013), 193–202.

[87] Hirshfield, L.M., Solovey, E.T., Girouard, A., et al. Brain measurement for usability testing and adaptive interfaces: An example of uncovering syntactic workload with functional near infrared spectroscopy. *Proceedings of the sigchi conference on human factors in computing systems*, ACM (2009), 2185–2194.

[88] Hoffmann, S. and Falkenstein, M. The correction of eye blink artefacts in the eeg: A comparison of two prominent methods. *PLoS One 3*, 8 (2008), e3004.

[89] Hoffmann, U., Vesin, J.-M., Ebrahimi, T., and Diserens, K. An efficient p300-based brain–computer interface for disabled subjects. *Journal of Neuroscience methods 167*, 1 (2008), 115–125.

[90] Hong, K.-S., Naseer, N., and Kim, Y.-H. Classification of prefrontal and motor cortex signals for three-class fNIRS–BCI. *Neuroscience letters 587*, (2015), 87–92.

[91] Hoonakker, P., Carayon, P., Gurses, A.P., et al. Measuring workload of icu nurses with a questionnaire survey: The nasa task load index (tlx). *IIE transactions on healthcare systems engineering 1*, 2 (2011), 131–143.

[92] Höök, K. Affective loop experiences–What are they? In *Persuasive technology.* Springer, 2008, 1–12.

[93] Höök, K. and Löwgren, J. Strong concepts: Intermediate-level knowledge in interaction design research. *ACM Transactions on Computer-Human Interaction (TOCHI) 19*, 3 (2012), 23.

[94] Hughes, P. and others. Comparing the utility and usability of the microsoft kinect and leap motion sensor devices in the context of their application for gesture control of biomedical images. European

Congress of Radiology 2015 (2015).

[95] Iqbal, S.T., Adamczyk, P.D., Zheng, X.S., and Bailey, B.P. Towards an index of opportunity: Understanding changes in mental workload during task execution. *Proceedings of the sigchi conference on human factors in computing systems*, ACM (2005), 311–320.

[96] Irazoqui-Pastor, P., Mody, I., and Judy, J.W. In-vivo eeg recording using a wireless implantable neural transceiver. *Neural engineering, 2003. conference proceedings. first international ieee embs conference on*, IEEE (2003), 622–625.

[97] Ishimaru, S., Kunze, K., Kise, K., et al. In the blink of an eye: Combining head motion and eye blink frequency for activity recognition with google glass. *Proceedings of the 5th augmented human international conference*, ACM (2014), 15.

[98] Izzetoglu, M., Bunce, S.C., Izzetoglu, K., Onaral, B., and Pourrezaei, K. Functional brain imaging using near-infrared technology. *IEEE Engineering in Medicine and Biology Magazine 26*, 4 (2007), 38.

[99] James, G.A., He, G., and Liu, Y. A full-size mri-compatible keyboard response system. *NeuroImage 25*, 1 (2005), 328–331.

[100] Javed, W., Kim, K., Ghani, S., and Elmqvist, N. Evaluating physical/virtual occlusion management techniques for horizontal displays. In *Human-computer interaction INTERACT 2011.* Springer Berlin Heidelberg, 2011, 391–408.

[101] Jensen, O. and Tesche, C.D. Frontal theta activity in humans increases with memory load in a working memory task. *European journal of Neuroscience 15*, 8 (2002), 1395–1399.

[102] Jobsis, F.F. Noninvasive, infrared monitoring of cerebral and myocardial oxygen sufficiency and circulatory parameters. *Science 198*, 4323 (1977), 1264–1267.

[103] Joyce, C.A., Gorodnitsky, I.F., and Kutas, M. Automatic removal of eye movement and blink artifacts from eeg data using blind component separation. *Psychophysiology 41*, 2 (2004), 313–325.

[104] Jung, C. N psychophysical relations of the associative experiment. *Journal of Abnormal Psychology 7*, (1907), 247–255.

[105] Kahneman, D. *Attention and effort.* Citeseer, 1973.

[106] Kakei, S., Hoffman, D.S., and Strick, P.L. Muscle and movement representations in the primary motor cortex. *Science 285*, 5436 (1999), 2136–2139.

[107] Kane, M.J. and Engle, R.W. The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: An individual-differences perspective. *Psychonomic bulletin & review 9*, 4 (2002), 637–671.

[108] Keil, A., Mussweiler, T., and Epstude, K. Alpha-band activity reflects reduction of mental effort in a comparison task: A source space analysis. *Brain Research 1121*, 1 (2006), 117–127.

[109] Khalid, M.B., Rao, N.I., Rizwan-i-Haque, I., Munir, S., and Tahir, F. Towards a brain computer interface using wavelet transform with averaged and time segmented adapted wavelets. *Computer, control and communication, 2009. ic4 2009. 2nd international conference on*, IEEE (2009), 1–4.

[110] Kirke, A., Williams, D., Mir, E., Bluglass, A., Whyte, C., and Eccleston, P.A. Technical report on

a short live-action film whose story with soundtrack is selected in real-time based on audience arousal during performance. 2013.

[111] Kirschner, P.A. Cognitive load theory: Implications of cognitive load theory on the design of learning. *Learning and instruction 12*, 1 (2002), 1–10.

[112] Klimesch, W. EEG alpha and theta oscillations reflect cognitive and memory performance: A review and analysis. *Brain research reviews 29*, 2 (1999), 169–195.

[113] Krauledat, M., Grzeska, K., Sagebaum, M., et al. Playing pinball with non-invasive bci. *Advances in neural information processing systems*, (2009), 1641–1648.

[114] Laar, B., Nijboer, F., Gürkök, H., Plass-Oude Bos, D., and Nijholt, A. User experience evaluation in bci: Bridge the gap. *International Journal of Bioelectromagnetism 13*, 3 (2011), 157–158.

[115] Laar, B. van de, Gürkök, H., Bos, D.P.-O., Poel, M., and Nijholt, A. Experiencing bci control in a popular computer game. *IEEE transactions on computational intelligence and AI in games 5*, 2 (2013), 176–184.

[116] LaFleur, K., Cassady, K., Doud, A., Shades, K., Rogin, E., and He, B. Quadcopter control in three-dimensional space using a noninvasive motor imagery-based brain–computer interface. *Journal of neural engineering 10*, 4 (2013), 046003.

[117] Lalor, E.C., Kelly, S.P., Finucane, C., et al. Steady-state vep-based brain-computer interface control in an immersive 3D gaming environment. *EURASIP journal on applied signal processing 2005*, (2005), 3156–3164.

[118] Landis, J.R. and Koch, G.G. The measurement of observer agreement for categorical data. *Biometrics*, (1977), 159–174.

[119] Lau, H. and Rosenthal, D. Empirical support for higher-order theories of conscious awareness. *Trends in cognitive sciences 15*, 8 (2011), 365–373.

[120] Laufs, H., Krakow, K., Sterzer, P., et al. Electroencephalographic signatures of attentional and cognitive default modes in spontaneous brain activity fluctuations at rest. *Proceedings of the national academy of sciences 100*, 19 (2003), 11053–11058.

[121] Leanne, M. and Robert, J. Using brain measurement to evaluate reality based interactions. *Challenges in the Evaluation of Usability and User Experience in Reality Based Interaction 5*, (2009), 19–20.

[122] Lee, J.C. and Tan, D.S. Using a low-cost electroencephalograph for task classification in hci research. *Proceedings of the 19th annual acm symposium on user interface software and technology*, ACM (2006), 81–90.

[123] Lee, J.D. and See, K.A. Trust in automation: Designing for appropriate reliance. *Human Factors: The Journal of the Human Factors and Ergonomics Society 46*, 1 (2004), 50–80.

[124] Lee, K. Evaluation of attention and relaxation levels of archers in shooting process using brain wave signal analysis algorithms. *Sci Sensitivity 12*, (2009), 341–350.

[125] Leeb, R., Friedman, D., Müller-Putz, G.R., Scherer, R., Slater, M., and Pfurtscheller, G. Self-paced (asynchronous) bci control of a wheelchair in virtual environments: A case study with a tetraplegic.

Computational intelligence and neuroscience 2007, (2007).

[126] Leeb, R., Lancelle, M., Kaiser, V., Fellner, D.W., and Pfurtscheller, G. Thinking penguin: Multimodal brain–Computer interface control of a vr game. *IEEE Transactions on Computational Intelligence and AI in Games 5*, 2 (2013), 117–128.

[127] León-Carrión, J. and León-Domínguez, U. Functional near-infrared spectroscopy (fNIRS): Principles and neuroscientific applications. *inNeuroimaging—Methods, P. Bright, Ed*, (2012), 47–74.

[128] Leslie, G. and Mullen, T. Moodmixer: Eeg-based collaborative sonification. *Proceedings of the international conference on new interfaces for musical expression*, Citeseer (2011), 296–299.

[129] Limb, C.J. and Braun, A.R. Neural substrates of spontaneous musical performance: An fMRI study of jazz improvisation. *PLoS one 3*, 2 (2008), e1679.

[130] Logothetis, N.K. The neural basis of the blood–oxygen–level–dependent functional magnetic resonance imaging signal. *Philosophical Transactions of the Royal Society of London B: Biological Sciences 357*, 1424 (2002), 1003–1037.

[131] Luo, A. and Sullivan, T.J. A user-friendly ssvep-based brain–computer interface using a time-domain classifier. *Journal of Neural Engineering 7*, 2 (2010), 026010.

[132] Luria, A.R. *Higher cortical functions in man.* Springer Science & Business Media, 2012.

[133] Lysaght, R.J., Hill, S.G., Dick, A., Plamondon, B.D., and Linton, P.M. *Operator workload: Comprehensive review and evaluation of operator workload methodologies.* DTIC Document, 1989.

[134] MacKenzie, I.S., Kauppinen, T., and Silfverberg, M. Accuracy measures for evaluating computer pointing devices. *Proceedings of the sigchi conference on human factors in computing systems*, ACM (2001), 9–16.

[135] MacKenzie, S. User studies and usability evaluations: From research to products. *Proceedings of graphics interface 2015*, Canadian Human-Computer Communications Society (2015), 1–8.

[136] Macleod, M. *Performance measurement and ecological validity.* Chapter, 1996.

[137] Maior, H.A., Pike, M., Sharples, S., and Wilson, M.L. Examining the reliability of using fNIRS in realistic hci settings for spatial and verbal tasks. *Proceedings of chi*, (2015), 3807–3816.

[138] Marshall, J., Rowland, D., Rennick Egglestone, S., Benford, S., Walker, B., and McAuley, D. Breath control of amusement rides. *Proceedings of the sigchi conference on human factors in computing systems*, ACM (2011), 73–82.

[139] Marshall, J., Rowland, D., Rennick Egglestone, S., Benford, S., Walker, B., and McAuley, D. Breath control of amusement rides. *Proceedings of the sigchi conference on human factors in computing systems*, ACM (2011), 73–82.

[140] Masuko, S. and Hoshino, J. A fitness game reflecting heart rate. *Proceedings of the 2006 acm sigchi international conference on advances in computer entertainment technology*, ACM (2006).

[141] Mathan, S., Whitlow, S., Dorneich, M., Ververs, P., and Davis, G. Neurophysiological estimation of interruptibility: Demonstrating feasibility in a field context. *In proceedings of the 4th international*

conference of the augmented cognition society, (2007).

[142] Matthews, P. and Jezzard, P. Functional magnetic resonance imaging. *Journal of Neurology, Neurosurgery & Psychiatry 75*, 1 (2004), 6–12.

[143] McDonald, S. and Petrie, H. The effect of global instructions on think-aloud testing. *CHI*, ACM (2013), 2941–2944.

[144] McDonald, S., Edwards, H.M., and Zhao, T. Exploring think-alouds in usability testing: An international survey. *IEEE Transactions on Professional Communication 55*, 1 (2012), 2–19.

[145] McMenamin, B.W., Shackman, A.J., Maxwell, J.S., et al. Validation of ica-based myogenic artifact correction for scalp and source-localized eeg. *Neuroimage 49*, 3 (2010), 2416–2432.

[146] Megaw, T. The definition and measurement of mental workload. *Evaluation of human work, Eds. Esmond N. Corlett, and John R. Wilson*, (2005), 525–551.

[147] Miles, J. and Shevlin, M. *Applying regression and correlation: A guide for students and researchers.* Sage, 2001.

[148] Miller, E.K., Erickson, C.A., and Desimone, R. Neural mechanisms of visual working memory in prefrontal cortex of the macaque. *Journal of neuroscience 16*, 16 (1996), 5154–5167.

[149] Miller, E.K., Freedman, D.J., and Wallis, J.D. The prefrontal cortex: Categories, concepts and cognition. *Philosophical Transactions of the Royal Society B: Biological Sciences 357*, 1424 (2002), 1123–1136.

[150] Miller, G.A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological review 63*, 2 (1956), 81.

[151] Mueller, S. PEBL: The psychology experiment building language (version 0.10).[Computer experiment programming language]. *Retrieved Nov*, (2012).

[152] Mulder, G. The heart of mental effort. (1980).

[153] Murch, W. *In the blink of an eye: A perspective on film editing.* Silman-James Press, 2001.

[154] Murphy, E.R., Illes, J., and Reiner, P.B. Neuroethics of neuromarketing. *Journal of Consumer Behaviour 7*, 4-5 (2008), 293–302.

[155] Müller-Putz, G.R. and Pfurtscheller, G. Control of an electrical prosthesis with an ssvep-based bci. *Biomedical Engineering, IEEE Transactions on 55*, 1 (2008), 361–364.

[156] Nabiyouni, M., Laha, B., and Bowman, D.A. Poster: Designing effective travel techniques with bare-hand interaction. *3D user interfaces (3DUI), 2014 ieee symposium on*, IEEE (2014), 139–140.

[157] Nacke, L.E. Wiimote vs. controller: Electroencephalographic measurement of affective gameplay interaction. *Proceedings of the international academic conference on the future of game design and technology*, ACM (2010), 159–166.

[158] Nagashima, Y. Bio-sensing systems and bio-feedback systems for interactive media arts. *Proceedings of the 2003 conference on new interfaces for musical expression*, National University of Singapore (2003),

48–53.

[159] Nenonen, V., Lindblad, A., Häkkinen, V., Laitinen, T., Jouhtio, M., and Hämäläinen, P. Using heart rate to control an interactive game. *Proceedings of the sigchi conference on human factors in computing systems*, ACM (2007), 853–856.

[160] Nicola Bizzotto, M., Alessandro Costanzo, M., and Leonardo Bizzotto, M. Leap motion gesture control with osirix in the operating room to control imaging: First experiences during live surgery. *Surgical innovation 1*, (2014), 2.

[161] Niedermeyer, E. and Silva, F.L. da. *Electroencephalography: Basic principles, clinical applications, and related fields*. Lippincott Williams & Wilkins, 2005.

[162] Nielsen, C.M., Overgaard, M., Pedersen, M.B., Stage, J., and Stenild, S. It's worth the hassle!: The added value of evaluating the usability of mobile systems in the field. *Proceedings of the 4th nordic conference on human-computer interaction: Changing roles*, ACM (2006), 272–280.

[163] Nishimoto, S., Vu, A.T., Naselaris, T., Benjamini, Y., Yu, B., and Gallant, J.L. Reconstructing visual experiences from brain activity evoked by natural movies. *Current Biology 21*, 19 (2011), 1641–1646.

[164] Noachtar, S. and Rémi, J. The role of eeg in epilepsy: A critical review. *Epilepsy & Behavior 15*, 1 (2009), 22–33.

[165] Ogawa, S., Lee, T.-M., Kay, A.R., and Tank, D.W. Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proceedings of the National Academy of Sciences 87*, 24 (1990), 9868–9872.

[166] Okumoto, S., Zhao, F., and Sawada, H. Tactoglove presenting tactile sensations for intuitive gestural interaction. *Industrial electronics (isie), 2012 ieee international symposium on*, (2012), 1680–1685.

[167] Olmsted-Hawala, E.L., Murphy, E.D., Hawala, S., and Ashenfelter, K.T. Think-aloud protocols: A comparison of three think-aloud protocols for use in testing data-dissemination web sites for usability. *Proceedings of the sigchi conference on human factors in computing systems*, ACM (2010), 2381–2390.

[168] Paas, F.G. and Van Merriënboer, J.J. The efficiency of instructional conditions: An approach to combine mental effort and performance measures. *Human Factors: The Journal of the Human Factors and Ergonomics Society 35*, 4 (1993), 737–743.

[169] Patel, S., Henderson, R., Bradley, L., Galloway, B., and Hunter, L. Effect of visual display unit use on blink rate and tear stability. *Optometry & Vision Science 68*, 11 (1991), 888–892.

[170] Paulesu, E., Frith, C.D., and Frackowiak, R.S. The neural correlates of the verbal component of working memory. *Nature 362*, 6418 (1993), 342.

[171] Peck, E.M., Afergan, D., and Jacob, R.J. Investigation of fNIRS brain sensing as input to information filtering systems. *Proceedings of the 4th augmented human international conference*, ACM (2013), 142–149.

[172] Peck, E.M., Afergan, D., Yuksel, B.F., Lalooses, F., and Jacob, R.J. Using fNIRS to measure mental workload in the real world. *Advances in physiological computing*, (2014), 117–139.

[173] Peck, E.M.M., Yuksel, B.F., Ottley, A., Jacob, R.J., and Chang, R. Using fNIRS brain sensing to

evaluate information visualization interfaces. *Proceedings of the sigchi conference on human factors in computing systems*, ACM (2013), 473–482.

[174] Penfield, W. and Roberts, L. *Speech and brain mechanisms.* Princeton University Press, 2014.

[175] Pfurtscheller, G., Stancak, A., and Neuper, C. Event-related synchronization (ers) in the alpha band—an electrophysiological correlate of cortical idling: A review. *International journal of psychophysiology 24*, 1 (1996), 39–46.

[176] Piccione, F., Giorgi, F., Tonin, P., et al. P300-based brain computer interface: Reliability and performance in healthy and paralysed participants. *Clinical neurophysiology 117*, 3 (2006), 531–537.

[177] Pike, M., Wilson, M.L., Divoli, A., and Medelyan, A. CUES: Cognitive Usability Evaluation System. *EuroHCIR2012*, (2012), 51–54.

[178] Pike, M.F., Maior, H.A., Porcheron, M., Sharples, S.C., and Wilson, M.L. Measuring the effect of think aloud protocols on workload using fNIRS. *Proceedings of the 32nd annual acm conference on human factors in computing systems*, ACM (2014), 3807–3816.

[179] Pineda, J.A., Silverman, D.S., Vankov, A., and Hestenes, J. Learning to control brain rhythms: Making a brain-computer interface possible. *IEEE transactions on neural systems and rehabilitation engineering 11*, 2 (2003), 181–184.

[180] Pires, G., Torres, M., Casaleiro, N., Nunes, U., and Castelo-Branco, M. Playing tetris with non-invasive bci. *Serious games and applications for health (segah), 2011 ieee 1st international conference on*, IEEE (2011), 1–6.

[181] Pizzagalli, D.A. and others. Electroencephalography and high-density electrophysiological source localization. *Handbook of psychophysiology 3*, (2007), 56–84.

[182] Pollard, T., Reimer, M., San, D., et al. Template for writing a phd thesis in markdown. 2016. https://doi.org/10.5281/zenodo.58490.

[183] Pollatos, O., Schandry, R., Auer, D.P., and Kaufmann, C. Brain structures mediating cardiovascular arousal and interoceptive awareness. *Brain research 1141*, (2007), 178–187.

[184] Price, C.J. The anatomy of language: A review of 100 fMRI studies published in 2009. *Annals of the New York Academy of Sciences 1191*, 1 (2010), 62–88.

[185] Qiyuan, J., Richer, F., Wagoner, B.L., and Beatty, J. The pupil and stimulus probability. *Psychophysiology 22*, 5 (1985), 530–534.

[186] Raghavachari, S., Kahana, M.J., Rizzuto, D.S., et al. Gating of human theta oscillations by a working memory task. *Journal of Neuroscience 21*, 9 (2001), 3175–3183.

[187] Raichle, M.E., MacLeod, A.M., Snyder, A.Z., Powers, W.J., Gusnard, D.A., and Shulman, G.L. A default mode of brain function. *Proceedings of the National Academy of Sciences 98*, 2 (2001), 676–682.

[188] Reeves, S. Designing the spectator experience stuart reeves, steve benford, claire o'Malley, mike fraser. (2005).

[189] Reid, G.B. and Nygren, T.E. The subjective workload assessment technique: A scaling procedure

for measuring mental workload. *Advances in psychology 52*, (1988), 185–218.

[190] Reuderink, B., Nijholt, A., and Poel, M. Affective pacman: A frustrating game for brain-computer interface experiments. *International conference on intelligent technologies for interactive entertainment*, Springer (2009), 221–227.

[191] Rizzolatti, G., Fadiga, L., Gallese, V., and Fogassi, L. Premotor cortex and the recognition of motor actions. *Cognitive brain research 3*, 2 (1996), 131–141.

[192] Rowe, D.W., Sibert, J., and Irwin, D. Heart rate variability: Indicator of user state as an aid to human-computer interaction. *Proceedings of the sigchi conference on human factors in computing systems*, ACM Press/Addison-Wesley Publishing Co. (1998), 480–487.

[193] Ryu, K. and Myung, R. Evaluation of mental workload with a combined measure based on physiological indices during a dual task of tracking and mental arithmetic. *International Journal of Industrial Ergonomics 35*, 11 (2005), 991–1009.

[194] Schnädelbach, H., Irune, A., Kirk, D., Glover, K., and Brundell, P. ExoBuilding: Physiologically driven adaptive architecture. *ACM Transactions on Computer-Human Interaction (TOCHI) 19*, 4 (2012), 25.

[195] Schnädelbach, H., Rennick Egglestone, S., Reeves, S., Benford, S., Walker, B., and Wright, M. Performing thrill: Designing telemetry systems and spectator interfaces for amusement rides. *Proceedings of the sigchi conference on human factors in computing systems*, ACM (2008), 1167–1176.

[196] Scholkmann, F., Kleiser, S., Metz, A.J., et al. A review on continuous wave functional near-infrared spectroscopy and imaging instrumentation and methodology. *Neuroimage 85*, (2014), 6–27.

[197] Scicali, A. and Bischof, H.-P. Usability study of leap motion controller. *Proceedings of the international conference on modeling, simulation and visualization methods (msv)*, The Steering Committee of The World Congress in Computer Science, Computer Engineering; Applied Computing (WorldComp) (2015), 39.

[198] Seixas, M., Cardoso, J., and Dias, M.T.G. One hand or two hands? 2D selection tasks with the leap motion device. (2015).

[199] Sereno, M.I., Dale, A., Reppas, J., Kwong, K., and others. Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science 268*, 5212 (1995), 889.

[200] Shively, R.J., Bortolussi, M.R., Battiste, V., Hart, S.G., Pepitone, D.D., and Matsumoto, J.H. Inflight evaluation of pilot workload measures for rotorcraft research. (1987).

[201] Shneiderman, B. The future of interactive systems and the emergence of direct manipulation†. *Behaviour & Information Technology 1*, 3 (1982), 237–256.

[202] Shneiderman, B. *Designing the user interface: Strategies for effective human-computer interaction.* Addison-Wesley Reading, MA, 1986.

[203] Shneiderman, B. *Designing the user interface: Strategies for effective human-computer interaction.* Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1986.

[204] Shneiderman, B. Direct manipulation for comprehensible, predictable and controllable user inter-

faces. *Proceedings of the 2Nd international conference on intelligent user interfaces*, ACM (1997), 33–39.

[205] Silva[1], E.S., Abreu[1], J.A.O. de, Almeida[1], J.H.P. de, Teichrieb, V., and Ramalho, G.L. A preliminary evaluation of the leap motion sensor as controller of new digital musical instruments. (2013).

[206] Smith, A.M., Mancini, M.C., and Nie, S. Second window for in vivo imaging. *Nature nanotechnology 4*, 11 (2009), 710.

[207] Smith, E.E. and Jonides, J. Storage and executive processes in the frontal lobes. *Science 283*, 5408 (1999), 1657–1661.

[208] Smith, M. and Gevins, A. Neurophysiologic monitoring of mental workload and fatigue during operation of a flight simulator. *Proceedings of SPIE - The International Society for Optical Engineering 5797*, 2005, 116–126. https://www.scopus.com/inward/record.uri?eid=2-s2.0-27544480351&doi=10.1117%2f12.602181&partnerID=40&md5=c3f2f8811c6986bfef144e125d3ad30b.

[209] Solovey, E.T. and Jacob, R.J. Meaningful human-computer interaction using fNIRS brain sensing..

[210] Solovey, E.T., Girouard, A., Chauncey, K., et al. Using fNIRS brain sensing in realistic hci settings: Experiments and guidelines. *Proceedings of the 22Nd annual acm symposium on user interface software and technology*, ACM (2009), 157–166.

[211] Stach, T., Graham, T.C.N., Yim, J., and Rhodes, R.E. Heart rate control of exercise video games. *Proceedings of graphics interface 2009*, Canadian Information Processing Society (2009), 125–132.

[212] Stanton, N.A. and Young, M.S. Automotive automation: Investigating the impact on drivers' mental workload. *International Journal of Cognitive Ergonomics 1*, 4 (1997), 325–336.

[213] Stanton, N.A., Young, M., and McCaulder, B. Drive-by-wire: The case of driver workload and reclaiming control with adaptive cruise control. *Safety science 27*, 2 (1997), 149–159.

[214] Steimle, J., Khalilbeigi, M., Mühlhäuser, M., and Hollan, J.D. Physical and digital media usage patterns on interactive tabletop surfaces. *ACM international conference on interactive tabletops and surfaces*, ACM (2010), 167–176.

[215] Stern, J.A., Walrath, L.C., and Goldstein, R. The endogenous eyeblink. *Psychophysiology 21*, 1 (1984), 22–33.

[216] Sweller, J. Cognitive load during problem solving: Effects on learning. *Cognitive science 12*, 2 (1988), 257–285.

[217] Sweller, J. Cognitive load theory, learning difficulty, and instructional design. *Learning and instruction 4*, 4 (1994), 295–312.

[218] Sweller, J., Van Merrienboer, J.J., and Paas, F.G. Cognitive architecture and instructional design. *Educational psychology review 10*, 3 (1998), 251–296.

[219] Tan, D. and Nijholt, A. Brain-computer interfaces and human-computer interaction. In *Brain-computer interfaces*. Springer, UK, 2010.

[220] Tatum, W.O. Ellen r. grass lecture: Extraordinary eeg. *The Neurodiagnostic Journal 54*, 1 (2014),

3–21.

[221] Tiganj, Z., Mboup, M., Pouzat, C., and Belkoura, L. An algebraic method for eye blink artifacts detection in single channel eeg recordings. *17th international conference on biomagnetism advances in biomagnetism–Biomag2010*, Springer (2010), 175–178.

[222] Toronov, V., Webb, A., Choi, J.H., et al. Investigation of human brain hemodynamics by simultaneous near-infrared spectroscopy and functional magnetic resonance imaging. *Medical physics 28*, (2001), 521.

[223] Tremoulet, P., Barton, J., Craven, P., et al. Augmented cognition for tactical tomahawk weapons control system operators. *Foundations of augmented cognition*, (2006), 313–318.

[224] Tsang, P.S. and Velazquez, V.L. Diagnosticity and multidimensional subjective workload ratings. *Ergonomics 39*, 3 (1996), 358–381.

[225] Ulmer, S. and Jansen, O. *FMRI*. Springer, 2010.

[226] Valli, A. The design of natural interaction. *Multimedia Tools and Applications 38*, 3 (2008), 295–305.

[227] Vecchiato, G., Astolfi, L., De Vico Fallani, F., et al. On the use of eeg or meg brain imaging tools in neuromarketing research. *Computational intelligence and neuroscience 2011*, (2011), 3.

[228] Vecchiato, G., Astolfi, L., Fallani, F.D.V., et al. Changes in brain activity during the observation of tv commercials by using eeg, gsr and hr measurements. *Brain topography 23*, 2 (2010), 165–179.

[229] Vecchiato, G., Astolfi, L., Fallani, F.D.V., et al. The study of brain activity during the observation of commercial advertsing by using high resolution eeg techniques. *Engineering in medicine and biology society, 2009. embc 2009. annual international conference of the ieee*, IEEE (2009), 57–60.

[230] Vecchiato, G., Astolfi, L., Tabarrini, A., et al. EEG analysis of the brain activity during the observation of commercial, political, or public service announcements. *Computational intelligence and neuroscience 2010*, (2010), 6.

[231] Vecchiato, G., Kong, W., Giulio Maglione, A., and Wei, D. Understanding the impact of tv commercials. *IEEE pulse 3*, 3 (2012), 42.

[232] Vecchiato, G., Toppi, J., Astolfi, L., et al. Spectral eeg frontal asymmetries correlate with the experienced pleasantness of tv commercial advertisements. *Medical & biological engineering & computing 49*, 5 (2011), 579–583.

[233] Villringer, A. and Chance, B. Non-invasive optical spectroscopy and imaging of human brain function. *Trends in neurosciences 20*, 10 (1997), 435–442.

[234] Villringer, A., Planck, J., Hock, C., Schleinkofer, L., and Dirnagl, U. Near infrared spectroscopy (NIRS): A new tool to study hemodynamic changes during activation of brain function in human adults. *Neuroscience letters 154*, 1 (1993), 101–104.

[235] Weichert, F., Bachmann, D., Rudak, B., and Fisseler, D. Analysis of the accuracy and robustness of the leap motion controller. *Sensors 13*, 5 (2013), 6380–6393.

[236] Weinger, M.B., Reddy, S.B., and Slagle, J.M. Multiple measures of anesthesia workload during

teaching and nonteaching cases. *Anesthesia & Analgesia 98*, 5 (2004), 1419–1425.

[237] Wickens, C.D. Multiple resources and mental workload. *The Journal of the Human Factors and Ergonomics Society 50*, 3 (2008), 449–455.

[238] Wickens, C.D., Gordon, S.E., and Liu, Y. *An introduction to human factors engineering.* Pearson Prentice Hall Upper Saddle River, 2004.

[239] Wickens, C.D., Hollands, J.G., Banbury, S., and Parasuraman, R. *Engineering psychology & human performance.* Psychology Press, 2015.

[240] Wierwille, W.W. and Casali, J.G. A validated rating scale for global mental workload measurement applications. *Proceedings of the human factors and ergonomics society annual meeting*, Sage Publications (1983), 129–133.

[241] Wilson, J.R. and Sharples, S. *Evaluation of human work.* CRC Press, 2015.

[242] Wilson, M.L. Evaluating the cognitive impact of search user interface design decisions. *EuroHCIR*, Citeseer (2011), 27–30.

[243] Wilson, R., Gaines, J., and Hill, R.P. Neuromarketing and consumer free will. *Journal of consumer affairs 42*, 3 (2008), 389–410.

[244] Wobrock, D., Frey, J., Graeff, D., Rivière, J.-B. de la, Castet, J., and Lotte, F. Continuous Mental Effort Evaluation during 3D Object Manipulation Tasks based on Brain and Physiological Signals. *INTERACT*, (2015).

[245] Wolpaw, J.R. and McFarland, D.J. Control of a two-dimensional movement signal by a noninvasive brain-computer interface in humans. *Proceedings of the National Academy of Sciences of the United States of America 101*, 51 (2004), 17849–17854.

[246] Woodruff, A., Augustin, S., and Foucault, B. Sabbath day home automation: "It's like mixing technology and religion". *Proceedings of the sigchi conference on human factors in computing systems*, ACM (2007), 527–536.

[247] Ye, J.C., Tak, S., Jang, K.E., Jung, J., and Jang, J. NIRS-SPM: Statistical parametric mapping for near-infrared spectroscopy. *Neuroimage 44*, 2 (2009), 428–447.

[248] Yiyuan, Z., Tangwen, Y., Dayong, D., and Shan, F. Using nasa-tlx to evaluate the flight deck design in design phase of aircraft. *Procedia Engineering 17*, (2011), 77–83.

[249] Young, G., Zavelina, L., and Hooper, V. Assessment of workload using nasa task load index in perianesthesia nursing. *Journal of PeriAnesthesia Nursing 23*, 2 (2008), 102–110.

[250] Yuksel, B.F., Afergan, D., Peck, E.M., et al. BRAAHMS: A novel adaptive musical interface based on users' cognitive state. *Proc. nime*, (2015).

[251] Yuksel, B.F., Oleson, K.B., Harrison, L., et al. Learn piano with bach: An adaptive learning interface that adjusts task difficulty based on brain state. *Proceedings of the 2016 chi conference on human factors in computing systems*, ACM (2016), 5372–5384.

[252] Zander, T. and Kothe, C. Towards passive brain-computer interfaces: Applying brain-computer

interface technology to human-machine systems in general. *Journal of Neural Engineering 8*, 2 (2011).

[253] Zander, T.O. and Kothe, C. Towards passive brain–computer interfaces: Applying brain–computer interface technology to human–machine systems in general. *Journal of neural engineering 8*, 2 (2011), 025005.

[254] Zander, T.O., Kothe, C., Jatzev, S., and Gaertner, M. Enhancing human-computer interaction with input from active and passive brain-computer interfaces. In *Brain-computer interfaces*. Springer, 2010, 181–199.

[255] Zheng, B., Jiang, X., Tien, G., Meneghetti, A., Panton, O.N.M., and Atkins, M.S. Workload assessment of surgeons: Correlation between nasa tlx and blinks. *Surgical endoscopy 26*, 10 (2012), 2746–2750.