

Mind the Sheep!

User Experience Evaluation &
Brain-Computer Interface Games



Hayrettin Gürkök



MIND THE SHEEP!

User Experience Evaluation &
Brain-Computer Interface Games

HAYRETTİN GÜRKÖK

Ph.D. Dissertation Committee

Chairman and Secretary

Prof. dr. ir. A. Mouthaan University of Twente, NL

Promotor

Prof. dr. ir. A. Nijholt University of Twente, NL

Assistant Promotor

Dr. M. Poel University of Twente, NL

Members

Dr. M. Congedo CNRS, FR

Prof. dr. ir. P. Desain Radboud University Nijmegen, NL

Prof. dr. D. Heylen University of Twente, NL

Prof. dr. C. Klimmt HMTM Hanover, DE

Prof. dr. M. Rauterberg Eindhoven University of Technology, NL

Prof. dr. F. van der Velde University of Twente, NL



CTIT Ph.D. Thesis Series

ISSN: 1381-3617, No. 12-228

Centre for Telematics and Information Technology

P.O. Box 217, 7500 AE

Enschede, The Netherlands



SIKS Dissertation Series No. 2012-27

The research reported in this thesis was carried out under the auspices of SIKS, the Dutch Research School for Information and Knowledge Systems.



The author gratefully acknowledges the support of the BrainGain Smart Mix Programme of the Netherlands Ministry of Economic Affairs and the Netherlands Ministry of Education, Culture and Science.



The research reported in this dissertation was carried out at the Human Media Interaction group of the University of Twente.

Printed and bound in The Netherlands by
Ipskamp Drukkers B.V.

L^AT_EX template classicthesis by André Miede
Cover design by Kardelen Hatun
Copyright © 2012 Hayrettin Gürkök

ISBN: 978-90-365-3395-9

MIND THE SHEEP!
USER EXPERIENCE EVALUATION &
BRAIN-COMPUTER INTERFACE GAMES

DISSERTATION

to obtain
the degree of doctor at the University of Twente,
on the authority of the rector magnificus,
prof. dr. H. Brinksma,
on account of the decision of the graduation committee
to be publicly defended
on Friday, 21 September 2012 at 12:45

by

HAYRETTİN GÜRKÖK

born on 3 May 1984

in Trabzon, Turkey

This thesis has been approved by:
Prof. dr. ir. A. Nijholt (promotor)
Dr. M. Poel (assistant promotor)

To my beloved parents...

You mean everything to me

PREFACE

In the spring of 2008, when I was writing my M.Sc. thesis on information retrieval and linguistics at Bilkent University, I was looking for a place where I could move from carrying out automated performance evaluation tests with information retrieval systems to conducting hands-on experiments with computer users. I contacted the Human Media Interaction (HMI) group at the University of Twente (UT) upon the suggestion of my dear friend, Kardelen Hatun, who was getting prepared for an interview at another group at the UT. Then, I was invited to Enschede by Prof. Anton Nijholt for an interview about a Ph.D. position on brain-computer interface games. After a mutually positive interview and with the support of the reference from my M.Sc. thesis supervisor, Dr. Murat Karamüftüoğlu, I was appointed to the position.

In my first year, with the help of my daily supervisor, Dr. Mannes Poel, I decided on the research direction that I wanted to take. Then, I started carrying out experiments, publishing papers and attending conferences. During my second year, I took the supervision of two M.Sc. students, now friends, Gido Hakvoort and Michel Obbink together with Mannes and Danny Plass-Oude Bos, my colleague from the Brainmedia group. Together we formed a nice working group and developed the experimental game that I used for my dissertation studies, *Mind the Sheep!*, which also gave its name to this dissertation. We demonstrated *Mind the Sheep!* at various venues and achieved good success with it, including the BrainGain Best Demonstration Award that we won in 2010. Until the end of my third year, when Anton, Mannes and my colleague Dr. Egon van den Broek convinced me to stop experimenting, I carried out various user experience experiments using *Mind the Sheep!* and published the results extensively. I also regularly collaborated with my colleagues from the Brainmedia group, Danny, Bram van de Laar, Dr. Christian Mühl, Dr. Boris Reuderink and Dr. Femke Nijboer. I devoted my fourth year solely to organising my studies into a story and put it into words. The outcome is the dissertation you are reading right now.

I have organised the dissertation in three parts and I suggest you to read the parts in the order that they appear. Part I contains 2 background chapters, and a rationale and overview chapter. Part II contains 3 chapters corresponding to three studies that contributed to the dissertation. Part III contains 2 chapters discussing and concluding the dissertation. Part IV is the appendix containing some supplementary material as well as the list of publications that I co-authored. I used British English not only due to personal choice but also as an homage to Lynn Packwood who spent so much labour in proofreading the

While reading the text you will see margin notes like this.

These notes will either draw your attention to important information in the text...

dissertation. I deliberately wrote the text using the plural first person narrative because otherwise I would have felt arrogant and discontent for ignoring the influence and assistance of several people in carrying out my research. In the next paragraph, I would like to extend my gratitude to them.

Anton; thanks for providing me this Ph.D. position, for educating me with your wisdom, for keeping a close watch on me despite your busy agenda, for caring for my all needs, and for answering my e-mails usually within 10 minutes (an estimate). Mannes; for giving me a hand when I fell, for being critical while reading my writings, and for putting me to the right direction by answering my questions or simply by leaving me on my own. Gido and Michel; for your invaluable contributions to Mind the Sheep! and to my experiments, and for your continuing friendship. Lynn; for proofreading all my papers and my dissertation, even the lines I am typing right now, for teaching me English, and for the small but much enjoyed talks we had every now and then. Christian, my paranymp; for showing me how to be thorough and critical while conducting research, giving me insight and support during my studies, and for the awesome time we spent in Genoa during eNTERFACE. Danny; for showing me how to do disciplined research and how to remain hopeful in difficult situations. Femke; for reading my whole dissertation and helping me to improve it, and for doing sincerely whatever you could do to support my academic career.

*...or they will
provide additional
information not
found in the text.*

Next, I would like to thank some other people who made the life more pleasant for me in the Netherlands. Kardelen, my schoolmate, housemate, paranymp, shopping advisor and many other hats you have been wearing for me; thank you for listening and understanding me, for looking after me, for putting up with me and for cheering me up with your adorable sense of humour. Oğuzcan; for advising me with your fully trustable rationality, for the quality time we have spent together and for your goodwill. My Turkish friends at the UT, Ramazan, Yakup, Muharrem, Akın, İmran; for creating a warm environment around me and for making me feel that I was not alone. Andrea, my office mate; for listening to me whenever I needed to talk and advising me to the best of your ability, for hosting me in amazing Padua, for improving my debating skills by opposing, more or less, everything and for the 'decent' Italian vocabulary you taught me. Andreea, my next-door office neighbour; for giving me a smile and persuading me to smile despite the problems, and for helping me without hesitation in anything I do. Bram; for your company in many cities/countries we've been to for some reason and for the thorough 'Beer 101' course you have taught (both the theory and practice). Core members of the lunch group II, Alejandro, Andreea, Christian, Daphne, Frans, Jorge, Maral, Randy; for your company which helped me bear the ultra-low edibility of the food in the canteen. Alice and Charlotte; for taking care

of all the troublesome stuff for me and for your efforts to make me feel more comfortable in the Netherlands by finding my first house and teaching me the Dutch language and culture. The Turks of Veendam, Ömer-Fatoş Maraş and Ali-Arzu, İbrahim-Filiz and Serkan Yılmaz; for being a family to me in the Netherlands and caring for me all the time.

Finally, I would like to thank to the people whose love and care reached to me despite the long distance between us. Grandfather; thank you for keeping track of me since the first day I started working and giving me confidence that you were always behind me. Sister; for being my best ally in this life and for indulging me with the fact that a gorgeous girl out there truly loves me. My dear parents; no word of any language is powerful enough to express my gratefulness to you.

July 2012, Enschede
Bize her yer Trabzon!

Hayrettin Gürkök

CONTENTS

I	INTRODUCTION	1
1	BRAIN-COMPUTER INTERFACES AND GAMES	3
1.1	Defining BCI	4
1.1.1	Acquiring Brain Activity (Imaging Modalities)	5
1.1.2	Interpreting Brain Activity (Neuromechanisms)	7
1.1.3	Interacting via Brain Activity	9
1.2	BCI Games	12
1.2.1	Motivations for Playing BCI Games	13
1.2.2	An Overview of BCI Games	16
2	USER EXPERIENCE EVALUATION AND GAMES	21
2.1	User Experience	21
2.2	Concepts for UX Evaluation of Games	25
2.2.1	Pre-Game User Experience	26
2.2.2	In-Game User Experience	27
2.2.3	Post-Game User Experience	31
2.2.4	Playability	33
2.3	Data Collection Methods for UX Evaluation of Games	34
2.3.1	Quantitative Methods	36
2.3.2	Qualitative Methods	37
3	EVALUATING USER EXPERIENCE OF BCI GAMES	41
3.1	State of the Art	41
3.1.1	Data Collection	41
3.1.2	Data Processing and Analysis	42
3.1.3	Results	43
3.2	Problem Statement	44
3.3	Methodology of the Thesis	46
3.3.1	Equalised Comparative Evaluation	47
3.3.2	Experimental Game: Mind the Sheep!	48
3.4	Overview of the Thesis	54
3.5	Contributions of the Thesis	54
II	STUDIES	57
4	EVALUATING SOCIAL INTERACTION AND CO-EXPERIENCE	59
4.1	Related Work	59
4.1.1	Social Interaction Evaluation	59
4.1.2	UX Evaluation with Multi-Player BCI Games	61
4.2	Methodology	61
4.2.1	Evaluating Social Interaction	61
4.2.2	ECE Details	62
4.2.3	Pilot Study: Optimising BCI Performance	63

4.3	Experiment	64
4.3.1	Setup	64
4.3.2	Participants	65
4.3.3	Results	66
4.4	Discussion	67
4.4.1	Speech	67
4.4.2	Instrumental Gestures	69
4.4.3	Utterances and Empathic Gestures	70
5	EVALUATING IMMERSION AND AFFECT	73
5.1	Related Work	73
5.2	Methodology	74
5.3	Experiment	74
5.3.1	Setup	74
5.3.2	Participants	75
5.3.3	Analysis	76
5.3.4	Results	76
5.4	Discussion	77
6	EVALUATING UX INDEPENDENT OF NOVELTY	81
6.1	Related Work	81
6.2	Methodology	82
6.2.1	ECE Details	82
6.2.2	Questionnaires	83
6.2.3	Evaluating UX With Respect to Expectations	84
6.2.4	Modality Switching Behaviour and UX	86
6.3	Experiment	86
6.3.1	Setup	86
6.3.2	Participants	87
6.3.3	Analysis	88
6.3.4	Results	89
6.4	Discussion	93
6.4.1	Questionnaire and Log Results	93
6.4.2	SUXES Results	97
6.4.3	Modality Switching Results	98
III	CONCLUSION	103
7	IMPLICATIONS OF THE STUDIES	105
7.1	Answering the Research Question	105
7.1.1	Challenging Control	105
7.1.2	Cognitive Involvement	106
7.1.3	Novelty	107
7.2	Reversing the Research Question	107
7.2.1	Challenging Control	107
7.2.2	Cognitive Involvement	108
7.2.3	Novelty	109

7.3	How to Evaluate the UX of BCI Games?	109
7.3.1	Data Collection Methods	109
7.3.2	UX Related Concepts	111
7.3.3	Participant Selection and Treatment	111
7.3.4	Comparing BCI to Other Modalities	112
8	LIMITATIONS OF THE THESIS AND FUTURE WORK	115
8.1	The Game and BCI Hardware	115
8.2	UX Evaluation	116
8.3	Participants	118
IV	APPENDIX	121
A	EXPERIMENTAL MATERIAL	123
A.1	Call for Participation in the Third Study	123
A.2	SUXES Questionnaires Used in the Third Study	123
A.2.1	Original Expectations Questionary	124
A.2.2	Original Perceptions Questionary	124
A.2.3	Modified Expectations Questionary	125
A.2.4	Modified Perceptions Questionary Merged with AttrakDiff2	126
B	TABLES	127
C	AUTHOR PUBLICATIONS CONTRIBUTING TO THE THESIS	133
D	AUTHOR PUBLICATIONS RELEVANT TO THE THESIS	135
	BIBLIOGRAPHY	137
	ABSTRACT	154
	SIKS DISSERTATION SERIES	157

LIST OF FIGURES

Figure 1	Three-component brain-computer interface application model	5
Figure 2	Proposed physiological computing framework	12
Figure 3	Data collection methods for UX evaluation	35
Figure 4	A screenshot of the game MTS!	49
Figure 5	A screenshot of the game MTS! while the SSVEP stimulation is on	50
Figure 6	A dog from the ASR version of the game MTS!	51
Figure 7	A screenshot from the multi-player version of the game MTS!	53
Figure 8	Two participants playing the game MTS!	65
Figure 9	Interpreting user perception with respect to user expectations	85
Figure 10	Performance indicators in Study 3	91
Figure 11	Factors affecting modality choice in Study 3	92
Figure 12	Active switcher performance in Study 3	93
Figure 13	Non-switcher performance in Study 3	94
Figure 14	Average selections durations in Study 3	95
Figure 15	Active switcher modality usage during MTS!-MM	99

LIST OF TABLES

Table 2	Properties of brain activity acquisition (measurement) methods used in HCI studies.	6
Table 3	Features and application domains for BCI interaction methods.	10
Table 4	UX related concepts categorised temporally with respect to interacting with a game.	25
Table 5	Aspects and related concepts evaluated in BCI game articles	43
Table 6	Summary of the studies conducted within thesis	54
Table 7	The results of the quantitative data analysis done for Study 1	67
Table 8	Social behaviour observed in Study 1	68

Table 9	IEQ dimension scores for Study 2	76
Table 10	SAM scores for Study 2	77
Table 11	Performance indicator values for Study 2	77
Table 12	Significant correlations between the IEQ dimension scores for Study 2	78
Table 13	Correlation coefficients for the SAM valence and dominance scores for Study 2	78
Table 14	NASA-TLX scores for Study 3	89
Table 15	GEQ scores for Study 3	90
Table 16	AttrakDiff2 scores for Study 3	90
Table 17	SUXES results for Study 3	92
Table 18	DOIs for the BCI game articles written between 2009-2011	130
Table 19	Recognition performance for stimulus frequencies to be used in MTS!-BCI.	131
Table 20	Recognition performance for dog names to be used in MTS!-ASR.	132

ACRONYMS

aBCI	Active BCI
ASR	Automatic speech recogniser
AV	Audio-visual
BCI	Brain-computer interface
CAR	Common average referencing
CCA	Canonical correlation analysis
ECE	Equalised comparative evaluation
EEG	Electroencephalography
ERD	Event-related descynchronisation
ERP	Event-related potential
ErrP	Error potential
ERS	Event-related scynchronisation
fMRI	Functional magnetic resonance imaging
GEQ	The Game Engagement Questionnaire
HCI	Human-computer interaction
IEQ	The Immersive Experience Questionnaire

IQR	Interquartile range
ITR	Information transfer rate
MEG	Magnetoencephalography
MM(I)	Multimodal (interaction)
MoA	Measure of adequacy
MoS	Measure of superiority
MRR	Mean reciprocal rank
MTS!	Mind the Sheep! (the game)
NASA-TLX	The NASA Task Load Index
NIRS	Near-infrared spectroscopy
P&C	Point-and-click
pBCI	Passive BCI
PC	Physiological computing
PX	Player experience
rBCI	Reactive BCI
RP	Readiness potential
SAM	The Self-Assessment Manikin
SS(V/A)EP	Steady-state (visually/auditorily) evoked potential
SNR	Signal-to-noise ratio
TS	Timed selection
UX	User experience
VE	Virtual environment
VR	Virtual reality
ZoE	Zone of expectations

Part I

INTRODUCTION

In agreement with the title of the thesis, in this first part, we will introduce brain-computer interface (BCI) games as well as user experience evaluation in games. This way, we will motivate the choices we made while conducting the studies reported in Part II. Then, we will argue for the importance of evaluating the user experience of BCI games and discuss the problems arising from the insufficient state-of-the-art evaluation trend. Finally, we will formulate our research question and describe our methodology to answer it.

The intelligent computers of today are able to perceive their environment using sensor technologies and to respond with the aid of advanced decision making algorithms. They welcome us into an elevator or a photo booth, and they accompany us in our pockets or on our clothes. Considering the amount of interaction we enter into with these pervasive machines, we need natural, intuitive user interfaces which understand or anticipate our intentions and react to make our lives easier. Thus, we should be able to interact with computers in the same way that we do with humans. In other words, human-computer interaction (HCI) should carry the characteristics of human-human interaction.

*Human-like
human-computer
interaction*

Human-human interaction relies on concurrent use and perception of behavioural signals (cues) such as speaking, moving, gazing and gesturing which convey various messages (communicative intentions). We show our approval by a thumbs-up perhaps accompanied with speech, a wink or a nod. In order to describe an object, we talk about it, at the same time moving our hands to explain its different features such as its size or shape. While we are sending our signals, our conversation partner receives them through his multiple senses; he listens to us and watches our gestures. For a human-like interaction, the interfaces of the modern HCI offer multiple sensing (input) and response (output) modalities. Within this interaction style, called multimodal interaction (MMI), computers hear us via the microphone, see through the camera and even feel through haptic devices. In return, they may give feedback in the form of an embodied conversational agent (Cassell et al., 2000) or through a tactile glove (Burdea, 2000).

Although computers can mimic some human senses, there are situations in which they need to possess better sensing abilities than humans. There are times that we, consciously or not, conceal our mental or emotional states. Some people are just not comfortable with expressing themselves overtly or they deliberately suppress their behavioural cues as in the case of bluffing. Moreover, in the absence of a human conversation partner, the cues may become subtle or may even vanish. In expressing our intentions, we are also not always explicit. This is perhaps because we are so tired that we do not want to move or our hands are occupied so that we can not use them or we are physically disabled. Still, we expect computers to understand our implicit emotions, difficulties and intentions.

*Beyond human-like
HCI*

Computers cannot read our minds but brain-computer interfaces (BCIs) can infer our psychological states and intentions by interpreting

our brain signals. The inferences made by BCI are rather limited due to, on the one hand, our limited knowledge of brain dynamics and, on the other hand, the limited capability of the brain activity acquisition tools and processing methods. Therefore, so far, the primary BCI users have been the severely disabled individuals for whom a BCI is the only option to restore their mobility and communication. With this motivation, the classical BCI applications have been brain-controlled wheelchairs (Leeb et al., 2007), spelling devices (Sellers and Donchin, 2006) and smart home environments (Holzner et al., 2009). For an extensive overview of classical BCIs we refer the reader to the review by Wolpaw et al. (2002).

Why BCI and
MMI?

With the emerging portable and usable brain signal acquisition hardware (Liao et al., 2012) BCI has started to be considered as an HCI modality for non-disabled users as well. In comparison to existing traditional (such as mouse and keyboard) or alternative novel (such as automatic speech recogniser (ASR) or Wii Remote) HCI modalities, BCI is a slow and/or unreliable (i.e. imperfect) modality. Actually, there is often a tradeoff between the speed and reliability of BCI because, to be able to perform reliable recognition, BCI requires accumulation of data but eventually this decreases its response time. Therefore, BCI applications for HCI mostly rely on slow paced interaction and/or MMI in which the weaknesses of BCI can be compensated by other modalities.

1.1 DEFINING BCI

The last half-decade has witnessed a debate over the widely-accepted definition of BCI by Wolpaw et al. (2002): “A BCI is a communication system in which messages or commands that an individual sends to the external world do not pass through the brain’s normal output pathways of peripheral nerves and muscles.”. We claim that this definition is not an incorrect one but an incomplete one. As we mentioned before, BCI can be helpful in situations where the user does not have an explicit command or message to send but implicit psychological states to be understood. A survey conducted by Nijboer et al. (2012) during the 4th International BCI Meeting held in Asilomar in 2010 revealed that research and development with BCI have evolved to answer this need of users. When 143 stakeholders were asked whether they would call a system a BCI or not, more than 60% of the respondents indicated that they would consider a fatigue monitor or an emotionally adaptive avatar as example BCI applications. Considering this controversy, we opt for a higher level and more inclusive definition. We represent a BCI application as a cycle with three procedural components which outputs supporting actions according to human intention or psychological state derived through

BCI application, a
functional definition

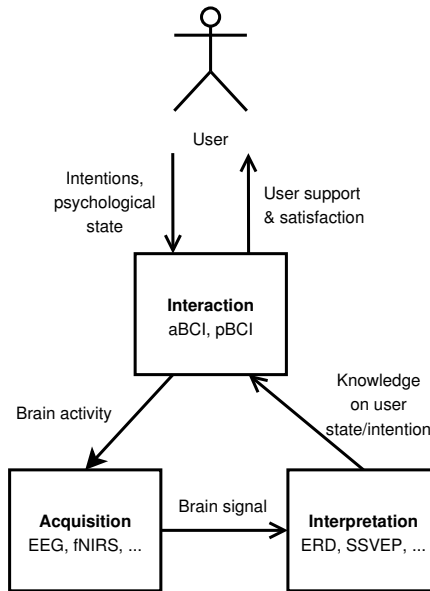


Figure 1: Three-component BCI application model. The user generates brain activity due to their intentions or psychological states. Brain activity is acquired and quantified as a signal. Then, the signal is interpreted to obtain knowledge on user state or intention. Finally, this knowledge is employed in satisfying the user's need.

brain activity (see Figure 1). The interaction block manages the high level interaction between the user and the BCI. It is responsible for evoking or instructing the user to generate the brain activity required for the BCI application to operate. In return, it provides feedback and/or service to support and satisfy the user with respect to their intentions and psychological states. The acquisition block acquires and quantifies the user's brain activity. The interpretation block interprets the digital signals generated by the acquisition block and outputs a prediction on user intention or state based on the neuromechanisms stemming from the neurological functioning of the brain. Next, we will describe each component in detail.

1.1.1.1 *Acquiring Brain Activity (Imaging Modalities)*

The first experiments on acquiring (measuring) human brain activity date back to the 1920s. Berger (1929) was the first to publish the results of electroencephalography (EEG) experiments on humans (translated version available by Gloor (1969)). EEG is a technique for acquiring the electrical activity of the brain from the scalp by use of electrodes.

	EEG	MEG	NIRS	fMRI
Measured activity	Electrical	Magnetic	Hemo- dynamic	Hemo- dynamic
Temporal resolution	Good	Good	Low	Low
Spatial resolution	Low	Low	Low	Good
Portability	High	Low	High	Low
Cost	Low	High	Low	High

Table 2: Properties of brain activity acquisition (measurement) methods used in HCI studies.

Since Berger’s first experiments, not only have EEG recordings become prevalent but other acquisition techniques relying on electrical, magnetic and hemodynamic (blood movement) responses of the brain have also emerged.

Brain activity acquisition methods¹ can be categorised according to the manner of deployment as being invasive or non-invasive. We will limit our discussion to non-invasive methods because invasive methods are not used in HCI studies. Among non-invasive methods (see Table 2), MEG (magnetoencephalography) and fMRI (functional magnetic resonance imaging) are carried out with immobile machines and require good shielding from the environment so they are bound to controlled laboratory environments. On the other hand, EEG and NIRS (near-infrared spectroscopy) are carried out with portable, easily deployable and relatively inexpensive devices. Wireless implementations are also feasible, making them even more convenient to use. Therefore, they are more suitable for HCI research.

*EEG for HCI:
portable and
inexpensive*

*EEG for HCI:
fast-responding*

EEG and MEG measure the activity of the fast dendritic currents in a large population of brain cells. Thus, the recordings of the measurements have low latency (i.e. high temporal resolution). fMRI and NIRS measure the blood oxygenation in the brain, which is a much slower correlate of the brain activity. Therefore they offer lower temporal resolution. On the other hand, fMRI has relatively higher spatial resolution as it can sample the activity of deep brain structures. Spatial resolution is a concern that has more priority in neuroscientific studies than it has in BCI applications. Neuroscientific research tries to identify sources in the brain which are reliable indicators of certain brain activities. Therefore, with better spatial resolution, more sources can be investigated and identified. On the other hand, BCI applications measure brain activity at the specialised locations which are already

¹ Also known as imaging modalities in neuroscience but we will not refer to them as modalities so as not to cause any conflict with the HCI definition of modality.

identified by neuroscientific studies. Therefore, spatial resolution is not equally critical for these applications.

Taken together, we can conclude that EEG is a preferable acquisition method for HCI as it is portable, inexpensive and responds fast. For a detailed description of the acquisition methods, the reader should refer to Lebedev and Nicolelis (2006), Kübler and Müller (2007), and van Gerven et al. (2009).

1.1.2 Interpreting Brain Activity (Neuromechanisms)

Once the brain activity is acquired as a signal, the next step is to interpret its content. In doing this, we benefit from neuromechanisms which signify certain changes in the signal with respect to an event. The event can be a voluntary action such as moving a hand or looking at something as well as an involuntary reaction to a stimulus or an error. In this section we will briefly cover the most commonly employed neuromechanisms.

The brain maintains an ongoing (rhythmic) activity in the absence of an external or internal intervention. These rhythms are identified by the frequency and brain location they occur at. Two closely related neuromechanisms, *event related desynchronisation* (ERD) and *event related synchronisation* (ERS) (Pfurtscheller and Lopes da Silva, 1999) (often referred to together as ERD/ERS) are the suppression and enhancement of the rhythmic brain activities respectively in relation to an event. By observing the signal amplitude in certain frequencies measured at specific parts of the brain we can infer the underlying brain activity. As an example, the Rolandic μ rhythm oscillates between 9-13 Hz in the sensorimotor area. It desynchronises during execution, preparation, observation or imagination of motor actions. So, by analysing the amplitude of the signal recorded from the sensorimotor area between 9-13 Hz, it is possible to understand when a person executes or imagines executing a motor action, such as a hand, foot or tongue movement (Pfurtscheller et al., 2006). If in an application certain motor actions are matched to some commands, then one can control the application without any device or even actual movement. Scherer et al. (2007) used motor imagery to navigate in a virtual environment (VE) and execute certain commands in Google Earth. Another example is the alpha rhythm oscillating between 8-13 Hz in the posterior region. It is blocked or attenuated by attention, especially visual, and mental effort so it has been associated with physical relaxation and relative mental inactivity (Deuschl and Eisen, 1999). Plass-Oude Bos et al. (2010) used parietal alpha power in the game World of Warcraft to switch the player avatar between an elf and a bear according to the player's level of relaxedness.

ERPs are known as
event related fields
in magnetic activity
measurement

Another family of neuromechanisms is the *event related potentials* (ERPs). ERPs are short-lived amplitude deflections in the brain signal, time-locked to a particular event. That is, they are expected at a fixed positive or negative latency with respect to an event. Thus, by observing the amplitude at this fixed latency, we can infer a person's reaction or intention. Various ERPs have been employed in BCI applications; we will introduce the most commonly used ones below. ERPs are identified by the triggering event, direction of deflection, observed location and latency. For the purpose of this paper, we will only emphasise the triggering event for each ERP and describe example applications. We encourage the reader to refer to Luck (2005) and Fabiani et al. (2007) for a complete overview. A commonly used potential of the brain, *P300*, occurs when we attend to a random series of stimuli that contains an infrequently presented stimulus or set of stimuli (Farwell and Donchin, 1988). Edlinger et al. (2009) used *P300* to select and control items in a virtual apartment while Campbell et al. (2010) used it to select and dial contacts on a real mobile phone. Intentions can also be inferred through the *readiness potential* (RP, also known as the *Bereitschaftspotential*) which precedes voluntary motor movements (Shibasaki and Hallett, 2006). Krepki et al. (2007) used lateralised RPs to predict the actual or imaginary finger movements of users and translate them into commands in a Pac-Man game. Another widely exploited set of potentials are the *error potentials* (ErrPs) which are reactions of the brain to errors (Ferrez and del R. Millán, 2007). Förster et al. (2010) used ErrPs to train their hand gesture recognition system based on the errors occurring during interaction.

When we attend to a stimulus repeating with a certain frequency, the amplitude of the signal measured in the brain area processing the stimulation is enhanced at the frequency of the stimulation. This enhancement is known as the *steady-state evoked potential* (SSEP) and is another frequently used neuromechanism (Regan, 1977). By presenting multiple stimuli with distinct repetition frequencies, we can detect which of the stimuli a person was paying attention to. If each of these stimuli is associated with a message, then we can understand the person's intention. Martinez et al. (2007) used four checkerboards each flickering with a unique frequency and associated with a direction (up, down, left and right) for navigating a car on the computer screen. As in this work, when the stimulation is a visual one, the resulting response is called a *steady-state visually evoked potential* (SSVEP) and is observed over the occipital (visual) cortex. In the literature there are also studies with auditory (Herdman et al., 2002) and vibratory (Muller-Putz et al., 2006) stimulation.

Apart from the above-mentioned standard neuromechanisms, there are power changes reported to occur at specific frequencies distributed across the scalp in correlation with emotions (Chanel et al., 2009) and

certain mental activities such as mental object rotation (Nikolaev and Anokhin, 1998) or problem solving (Fink et al., 2009). These correlates could, for instance, be used to detect a user's mental or emotional state for assisting the user. We would like to finally note that for some events there are more than one representative neuromechanisms, such as ERD and RP signifying motor execution or imagery, so combined use of these can yield a better recognition capability (Fatourechi et al., 2007).

While using externally evoked neuromechanisms (i.e. those which are evoked through external stimulation) such as ERPs and SSEPs, attention should be paid to the features of the stimuli, the stimulation device and the environment. The stimulation parameters might significantly affect not only the strength or the presence of the brain response but also the comfort and experience of the user. For SSVEP based BCIs, Bieger and Garcia Molina (2010) wrote an excellent report on the influence of stimulation parameters (such as the environment, the stimulation device, and the flicker frequency, colour and shape of the stimulus) on recognition performance and user comfort. Also for P300 stimulation, effects of factors such as screen size (Li et al., 2011), and colour of and distance between stimuli (Salvaris and Sepulveda, 2009) on recognition accuracy have been reported.

Attention points for externally evoked neuromechanisms

Externally evoked neuromechanisms offer some advantages to BCI developers as well. Firstly, they are easier to detect since they are time-locked to stimulation for which the onset and offset can be observed. Therefore, typical signal analysis for this class of neuromechanisms is restricted to an established time interval and brain location. Secondly, they have a high signal-to-noise ratio (SNR) due to the signal averaging procedure (Dawson, 1954) performed over several trials. With signal averaging, spontaneous responses cancel each other out so that only the response of interest survives (Mouraux and Iannetti, 2008). We note that the greater the number of trials, the higher the correct detection probability. On the other hand, the greater the number of trials, the slower the perceived speed of the interface.

Advantages of externally evoked neuromechanisms in BCIs

If we compare the two most prominent externally evoked neuromechanisms, P300 and SSEP, we see that SSEP provides extra convenience for BCI development. As it is phase-locked to stimulation, SSEP analysis is limited to a specific frequency (i.e. that of the stimulation and its harmonics). So SSEP analysis is performed on a specific brain location, time interval and frequency.

Extra advantages offered by SSEP for BCI development

1.1.3 Interacting via Brain Activity

Interpreting the brain activity based on neuromechanisms allows us to arrive at knowledge about a user's intention, mental processing or emotional state. We differentiate between BCIs with respect to their

TYPE OF BCI	INTERACTION WITH BCI	USED FOR	EXAMPLE APPLICATIONS
aBCI	Intended	Direct control	Motor imagery-based navigation, SSVEP-based selection, P300 speller, neurofeedback
pBCI	Unintended	Indirect control	Adaptive systems, error handling via Er-rPs

Table 3: Features and application domains for BCI interaction methods.

ways of utilising this knowledge in an application according to the user’s needs. Zander et al. (2010) identifies three types of BCIs: active (aBCI), reactive (rBCI) and passive (pBCI). In this work, we adapt this categorisation but reduce it to aBCI and pBCI only (see Table 3). We regard rBCI as a special case of aBCI since they only differ in generation of brain activity (self-initiated in aBCI and stimulus-induced in rBCI) which does not influence our discussion.

In *aBCI*, the user intends to interact with the BCI application in order to control it directly. Example aBCI applications include VE navigation based on imaginary movements or SSVEP. In *pBCI*, the user’s primary aim is not to interact with the BCI application or possibly he does not have an aim at all. The BCI system monitors the user passively in order to adapt the task or the environment for improving and enriching the HCI or the quality of life. This might be by monitoring the attention level, emotional state or mental load of the user. pBCIs rely on brain signals generated during natural interaction of the user with his environment so they do not require any additional effort (such as attention to stimulation). Therefore, they can also function in parallel to aBCIs or other modalities.

We would like to stress that the interaction methods can be used to describe BCI applications, but not the BCI neuromechanisms since a neuromechanism can be utilised in different ways in applications. For example a P300 speller (Serby et al., 2005) would be an rBCI since the user interacts with the BCI system for spelling words. On the other hand, a P300 workload monitor (Allison and Polich, 2008) would be a pBCI as the user has a primary task to devote attention to other than responding to the workload monitor. Having made this distinction, we would like to draw the reader’s attention to yet another important detail. Depending on the context and the goal of the user, different interaction

*Interaction methods
do not describe
neuromechanisms,
but BCI applications*

methods can be utilised to operate the very same BCI application. In the aforementioned BCI game, Alpha-World of Warcraft (Plass-Oude Bos et al., 2010), the player avatar changes between the elf and bear shapes according to the relaxedness of the player. During the game players might intentionally try to regulate their relaxedness for better performance or they might simply enjoy seeing the game reflect their natural state. In the former scenario the game would be an aBCI while in the latter a pBCI. Moreover, a blend of these two is highly probable during the game. Therefore BCI interaction methods are applicable to the applications but dependent on the user and the context.

Having discussed the interaction methods and made the distinction between aBCI and pBCI, we believe this is the right place to situate BCI within physiological computing (PC) systems. A PC system is “a category of technology where electrophysiological data recorded directly from the central nervous system or muscle activity are used to interface with a computing device” (Fairclough, 2011). Per this definition, an EEG-based BCI application should be considered as a PC system. Indeed Fairclough (2011) counts BCI as a PC category along with muscle interfacing, biofeedback, biocybernetic adaptation and ambulatory monitoring (Figure 2a). In this categorisation, BCI is reduced to aBCI due to its classical definition. However, going back to the survey by Nijboer et al. (2012), of the 143 BCI stakeholders a majority (65.7%) viewed pBCI as BCI as well. According to Zander et al. (2010), two PC categories, biocybernetic adaptation (e.g. adaptive systems to avoid cognitive overload) and ambulatory monitoring (e.g. operator monitoring to improve safety) are also example pBCI applications. Therefore, we suggest that the PC framework requires an update so that these two categories encompass pBCI, and the category BCI is replaced with aBCI (see Figure 2b).

We think further that the PC framework would still be unsatisfactory with the updates we suggested since both muscle interfacing and aBCI are about operating things. They let us command applications or systems directly. So, they should be combined in one category (see Figure 2c). Here, we would like to note that we differentiate between muscle interfaces used for direct control and using muscular activity for biofeedback, biocybernetic adaptation or ambulatory monitoring. This is actually exactly what we did with aBCI and pBCI. We now want to draw attention to the inconsistency regarding aBCI between the proposed PC framework and our classification of BCI interaction methods (see Table 3). In the latter, biofeedback is mentioned as aBCI (under the name neurofeedback) while in the former it is not. This is simply because in our classification it was not necessary to finely differentiate the high level goal in controlling things. However, the PC framework separates controlling things for the end goal of regulating oneself from doing so for solely operating a machine. Therefore, our

A BCI application can utilise more than one neuromechanisms

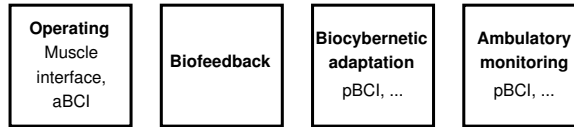
BCI and physiological computing, an integrative framework



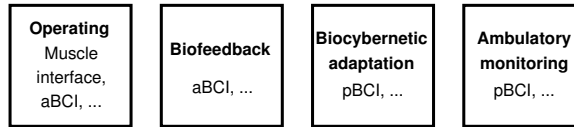
(a) The original PC framework by Fairclough (2011)



(b) Update 1: aBCI replaced BCI, pBCI introduced



(c) Update 2: Muscle interface and aBCI merged



(d) The final PC framework: Biofeedback includes aBCI

Figure 2: Updating PC framework to represent aBCI and pBCI

final proposed PC framework would include aBCI under both operating and biofeedback (see Figure 2d).

Since BCI is a PC system, it is highly probable that our discussions within this work are applicable to some other PC systems. But comparing BCI to such systems is beyond our purpose therefore, without loss of generality, our discussions will be limited to BCI.

1.2 BCI GAMES

Up to here, we have explained that BCI offers unique sensing capabilities for HCI which can hardly be replicated by any other modality. BCI can infer our intentions or the states we are in, even if we covertly express them or do not express them at all. On the other hand, we have also explained that BCI is an imperfect controller, unable to replace most of the available HCI modalities. Despite its shortcomings, one particular HCI community has shown continuous interest in employing BCI in applications: the gaming community. In the next subsection, we will explain why this should not actually surprise us.

1.2.1 Motivations for Playing BCI Games

To understand why one would show interest in playing BCI games, we should first look at the reasons behind playing computer games in general. Johnson and Wiles (2003) claim that the reason people play games is simply to experience positive affect. Hassenzahl et al. (2010) further demonstrated that positive affect was related to need fulfillment. They showed that positive affect was significantly correlated with psychological needs such as competence, stimulation and relatedness (Sheldon et al., 2001). Therefore, people would play games which tend to fulfill their psychological needs. Indeed, we see a correspondence between some psychological needs (Sheldon et al., 2001) and some game playing motivations (Rouse, 2005), such as competence and challenge or relatedness and socialisation. Through its unique characteristics, BCI can enable a game to satisfy some of the well-known player motivations. In this subsection, we will discuss three examples of such player motivations.

1.2.1.1 Challenge

Why do people enjoy playing the violin? Is that because it is easy or difficult to do so (Overbeeke et al., 2003)? How is it that they play a single measure of a piece tens of times to play it perfectly?

Kubovy (1999) suggests that when we achieve a goal or when we feel that we are doing something well, we experience positive affect. We become motivated to do the things that challenge us (White, 1959). Challenge is one of the elements of flow, which is the optimal experience for any activity and described as “so gratifying that people are willing to do it for its own sake, with little concern for what they will get out of it, even when it is difficult, or dangerous” (Csíkszentmihályi, 1990). Many researchers have shown the link between flow and games (Cowley et al., 2008). Based on the flow theory, Sweetser and Wyeth (2005) proposed a model describing which elements a game should have in order to provide flow. Their model suggests that a game should offer challenges matching player skills and both must exceed a certain threshold. Similarly, Carroll and Thomas (1988) suggest that “*examples of fun indeed must have sufficient complexity or they fall flat (jokes that are too obvious, games that are not challenging)*”. Moreover, “*things are fun when we expect them to be of moderate complexity (interesting but tractable) and then in fact find them to be so (i.e., not too difficult or too easy)*”.

Nijholt et al. (2009) recommend using BCI as a challenge mechanism in games. People playing a BCI game need to invest certain effort to find a way to generate the desired brain signals. Because, there is not always a standard mental activity that generates a brain signal. For example, there are more than one ways to desynchronise the μ

It is not the interface that is challenging; it is the signal generation that is challenging

rhythm by imagining movements. People might imagine themselves performing the action with interior view or they might imagine seeing themselves or someone else performing with an exterior view. Each of these ways of imagining leads to different – but still related – brain activity (Neuper et al., 2005) and, consequently, control ability. Therefore, in a game, players need to repeat their actions until they find the right mental activity to drive the game. Such purposeful repetition also brings fun (Blythe and Hassenzahl, 2003). Furthermore, humans do not possess a sense that can confirm mental activity. Let us consider the following example. When a player presses the left arrow key to steer their spaceship to the left, their touch and vision confirms that they pressed the correct key. When they say "Fire" to fire a bomb from the spaceship, their hearing confirms that they pronounced the correct word – though they may feel somewhat uncertain due to their accent. On the other hand, when they imagine moving their left hand to steer the spaceship to the left, they can not confirm that they are doing it right. Such uncertainty can motivate people to keep trying, until they resolve the uncertainty (Kagan, 1972).

BCI control is a skill

Sweetser and Wyeth (2005) draw attention to the point that the challenge posed by a game should be dynamic. That is, *"the level of challenge should increase as the player progresses through the game and increases their skill level"*. In the context of BCI games, this brings the question whether mental activity generation can be regarded as a skill and, if so, whether this skill can be increased. According to Wolpaw et al. (2002), people can manipulate and learn to improve their voluntary mental actions as well as involuntary reactions as they keep interacting with a BCI that provides accurate feedback.

1.2.1.2 Fantasy

Games let players do things that they cannot do – at least safely or without being criticised – in real life, such as fly or smash cars. However, in a virtual world, it is not trivial to provide the very same or at least a seemingly realistic sensation resulting from doing something in the real world. Such a sensation is known as presence and defined as *"the perception in which even though part or all of an individual's current experience is generated by and/or filtered through human-made technology, part or all of the individual's perception fails to accurately acknowledge the role of the technology in the experience"* (International Society for Presence Research, 2000). Riva (2009) claims that rather than our perception, it is our chain of actions that create the presence. He explains that a user *"is more present in a perceptually poor virtual environment ... where he/she can act in many different ways than in a real-like virtual environment where he/she cannot do anything."* Actually, 'to act' is not our ultimate goal. We aim 'to be' in the virtual world and to act is one way of satisfying

'To be' in the game world

our aim. So, we are more present in a virtual world in which we can represent ourselves more. At this point, the means with which the player drives the game becomes crucial.

Traditional game controllers, such as a gamepad or joystick, restrict the information flow from the player to the game. Firstly, the number of buttons or degrees of freedom provided by these controllers is insufficient to satisfy the infinitely large amount of information that could be transferred from the player. And secondly, the idea of representing oneself using buttons or a joystick is not an intuitive one since the player has to spend an effort to learn and memorise the mapping of their intentions to controller actions. Tremendous amount of research and development has been going on to alleviate this HCI bottleneck (Sharma et al., 1998). One example is the work on motion capturing techniques and devices (such as Kinect²), which enable one-to-one correspondence between player actions (as well as reactions) in the real world and those in the game world.

There are times when the players may need a deeper representation of themselves, rather than their overt actions. Let us consider a life-simulation game, *The Sims*³. In this game, the player controls the life of a character (or several characters) that can be customised to look like the player in terms of outfit or bodily and facial features. The character is also autonomous and its behaviour is influenced by the personality assigned to it by the player at the beginning of the game. It is inevitable that, at some time during play, this virtual look-alike of the player will not act or interact with other characters in congruence with the player's feelings or thoughts, because of either the inaccuracy of the player's personality assignment or the imperfection of the game to produce a desirable action. Consequently, this incongruence will hamper the player's sense of presence. In cases such as these, BCI can provide a translation between the psychological state of the player in the real world and the dynamics of the game world, just as Kinect provides correspondence between real-world and game-world actions. So, the additional inner state information can strengthen the feeling of presence.

*I think, therefore I
am in the game
world*

1.2.1.3 Sociality

Some people enjoy playing computer games with other people (Rouse, 2005). They play not necessarily for the challenge but just to be with others. They enjoy spending time with friends, seeing their reactions and expressions, and gloating or feeling proud upon winning (Sweetser and Wyeth, 2005). Any multi-player version of a BCI game can provide such an interactive environment. Players may cooperate or compete

² <http://www.xbox.com/kinect>

³ <http://www.thesims.com>

using BCI or they can share their experiences, such as difficulties or enjoyment with control, while playing the game. These interaction forms are, of course, not specific to BCI games. But, there are other ways which in BCI can provide sociality and which cannot be replicated easily or at all by other controllers.

Many social actions are related to expressing and perceiving emotions. Previous studies have shown that communication of heartbeat, which is a reflection of emotional activity, can improve the co-presence (Chanel et al., 2010) and intimacy (Janssen et al., 2010) of players. Heartbeat is certainly not the only nor the best indicator of emotion. BCI can recognise certain psychological states and let us share them. According to neurobiological emotion theories (for example the one by LeDoux (1995)), the brain is involved in the production and conscious registration of emotional states. So, theoretically, BCI can provide quick and direct information about our emotional state. Since involuntary brain activity, such as emotional response, is not easily controllable, BCI can provide objective information about our emotional state. For this reason, BCI can also be used in game situations where players would like to hide their psychological states from each other. For example, in a bluffing game, players can restrict their bodily movements and to some extent even their physiological activity but not their brain activity. So, BCI can be used for emotion-awareness or, more generally, psychological awareness in two opposite game logics (i.e. expressing versus concealing psychological state).

Going one step further than emotional awareness, the emotional contagion theory states that people tend to converge emotionally and, for this, they tend *“to automatically mimic and synchronise expressions, vocalisations, postures, and movements with those of another person”* (Hatfield et al., 1994). Research has confirmed that synchronisation contributes to coherence (Wiltermuth and Heath, 2009) and can be used as a measure of the intensity of the interaction between people (Hatfield et al., 1994). It has therefore been used in some game experience research (Ekman et al., 2012). This suggests that synchronisation games can strengthen the interaction between players. In such a game, BCI can enable synchronisation of psychological states, in the form of emotional synchrony (Kühn et al., 2011) or mental synchrony (Sobell and Trivich, 1989), and can provide a deeper and personal interaction between two players compared to physical synchrony.

*Psychological
synchrony via BCI*

1.2.2 An Overview of BCI Games

We find it useful to provide an overview of BCI games in order to analyse the interaction designs used in their development. We will not provide an extensive survey of BCI games in each category as this has

been done before (Plass-Oude Bos et al., 2010). If we categorise BCI games based on the neuromechanism they are based on, we end up with three categories which are mental state, movement imagery and evoked response games. A BCI game can rely on a single neuromechanism as well as on multiple neuromechanisms. The latter category of games are called hybrid BCI games (e.g. Mühl et al., 2010).

1.2.2.1 *Mental State Games*

Mental state games are usually played via two activities: relaxing or concentrating. These activities stem from clinical practice, such as relaxing to reduce anxiety or concentrating to reduce attention deficiency, but they are used in BCI games for very different purposes. Most of the mental state games allow players to move physical (Hjelm, 2003) or virtual (Oum et al., 2010) objects but there are other mechanisms such as changing the game avatar (Plass-Oude Bos et al., 2010).

Relaxing is a preferable activity in a game as it leads to a positive affective state that players would like to reach while playing games (Lazzaro, 2004). Therefore, even if the game environment is not an affective one, people may play such games for the end effect of being relaxed. Moreover, they might easily refer their acquaintances and even children to play such games. Concentration is also a preferable game activity due to its absorbing effect. According to the flow (Csíkszentmihályi, 1990; Sweetser and Wyeth, 2005) and immersion theories (Brown and Cairns, 2004), concentration is the key to successful games. Therefore, games requiring concentration or paying attention, which is one of the activities leading to concentration, ought to provide a positive play experience.

Why it is a good idea to use relaxedness and concentration in games

The speed with which we can change our state of relaxedness or concentration is much slower than the speed with which we can press buttons or use any other modality. Therefore, BCI games relying on mental state are either slow paced or in these games BCI is used as an auxiliary controller along with a primary controller which is faster than BCI. Mental state games usually allow only binary control. For example, in a relaxation game, players can either be relaxed or not relaxed so they can communicate a maximum of two discrete commands. It is possible to fit a continuous scale between these two states but validating such a scale is not trivial.

1.2.2.2 *Movement Imagery Games*

Movement imagery games originated from clinical studies for restoring the mobility and communication capabilities of disabled individuals (Wolpaw et al., 2002). They require no physical movement but imagery of limb movements, mostly the hands, fingers or feet. Players imagine

movements to navigate, as in driving a virtual car (Krepki et al., 2007), or to make selections, as in playing pinball (Tangemann et al., 2009). The latter example implies that it is possible to recognise movement imagery quite quickly, without needing to average the signal. Therefore, movement imagery games are suitable for fast interaction. On the other hand, the number of commands in these games is limited to the number of distinguishable imaginary actions players can perform. Using other modalities in combination with BCI can increase the number of commands. However, the movements made to control other modalities, such as pushing a button or speaking, might contaminate the movement imagery signal because in the absence of signal averaging, the SNR is very low (see §1.1.2).

1.2.2.3 *Evoked Response Games*

This class of games is dominated by SSVEP games, accompanied by rare examples of P300 games (e.g. Finke et al., 2009). The reason for this dominance might be due to the extra conveniences SSVEP provides compared to other evoked neuromechanisms (see §1.1.2). As Midden-dorf et al. (2000) also suggest, SSVEP games have been developed with two approaches.

The first approach is to map the strength of SSVEP which is evoked by single stimulus to game actions. For example, a weak SSVEP can steer a virtual plane to the left while a strong one to the right (Middendorf et al., 2000). Players can manipulate SSVEP strength in different ways. One way is to close and open the eyes to produce weak and strong SSVEP. But this would probably be too trivial to produce challenge in a game. Another way is to regulate SSVEP strength by the amount of attention paid. Research has shown that sustained attention can enhance SSVEP (di Russo and Spinelli, 2002). That is to say, we can infer whether a person is simply being exposed to a stimulus or they are actually paying attention to it. Sustained attention is an activity that can lead to a state of concentration (Mateer and Sohlberg, 2001). This makes SSVEP suitable for concentration games, the advantages of which we discussed in §1.2.2.1.

The second approach, which is the more popular one, is to use multiple stimuli each of which is associated with a command. In almost all games built with this approach, BCI is used to select a direction. Players can select a direction to aim their gun in a first-person shooter game (Moore Jackson et al., 2009) or to steer their car in a racing game (Martinez et al., 2007). The advantage of this approach is that, compared to the other class of games, they allow higher numbers of commands. Mental state games are limited by the levels of relaxedness or concentration states that BCI can detect while movement imagery games are limited by the number of body limbs that we can imagine

*SSVEP games allow
high number of
commands*

and BCI can differentiate. Such restrictions do not apply for evoked response games. Simply adding more stimuli can increase the number of commands. On the other hand, a computer screen is a limited space so the number of stimuli that can be placed on the screen is also limited. Moreover, as their number increases, the stimuli come closer to each other. This makes paying attention difficult for the user as multiple stimuli could interfere with each other. Furthermore, a screen cluttered with attention demanding stimuli would prevent the player from enjoying primary game elements, such as its visuals or the story line.

Evoked response games are less suitable for fast games due to the signal averaging process, which requires signals to accumulate for some time. But they are suitable for multimodal interaction thanks to their high SNR.

HIGHLIGHTS FROM THE CHAPTER

- BCI is a PC system that offers a beyond-human-like HCI by inferring our intentions and psychological states.
- It is a viable approach to employ BCI in MMI to compensate for the weaknesses of BCI.
- EEG is a preferable brain signal acquisition method for BCI applications in HCI. It is portable, inexpensive, and fast-responding.
- BCI can provide challenge, fantasy and sociality in computer games in ways that cannot be achieved by other modalities. These are the motivations for people to play computer games in general and correspond to the basic human psychological needs.
- Compared to other neuromechanisms, externally evoked neuromechanisms are easy to detect and they provide high SNR. Among them, SSEP is extra advantageous as it is simpler to analyse and also allows a potentially high number of commands to be issued. Moreover, the concentration required to play SSEP based games can contribute to flow.

...Then he chopped the tomatoes he'd peeled into the bowl. He poured some vinegar and some olive oil. He mixed all the vegetables in the bowl and tasted the salad. 'Salt!' he murmured raising his eyebrows and added a pinch of salt. Tasting it again, he nodded with satisfaction. Then he shaped the tomato skins like a rose and placed them on the top of the salad. He smiled thinking his girlfriend would like it...

This is not to practise our storytelling skills but to illustrate a situation that is familiar or at least realistic to most of us. In this chapter, we will try to answer questions such as "Why did he chop the tomatoes?", "Why did he taste the salad?", "Why did he shape the tomato skins?", "Did she like the salad in the end?". We will see how the answers to these questions help us understand what user experience (UX) is and why it is important for games and in general.

2.1 USER EXPERIENCE

What difference does it make when the tomatoes are chopped or a pinch of salt is added to the salad? Do they make the salad edible? We can still eat the salad when the tomatoes are not chopped or when there is no salt in it. We can still satisfy our hunger. But edibility is something else. It is about fitness, palatability to be eaten. When the tomato is not chopped, we cannot fit it into our mouth. If we bite, it will squirt. If we try to cut, it will roll in the olive oil. It is difficult, time consuming and frustrating to eat. When there is no salt in the salad, we find it tasteless. It is unsatisfactory to eat. Thus, these things do make the salad edible.

Edibility of the things we eat is analogous to usability of the things we use. Usability is not about whether something functions correctly. When we talk about usability, we refer to a product's learnability, efficiency, memorability, reliability, and resulting satisfaction from use; the attributes of usability as outlined by Nielsen (1993). In ISO 9241-11:1998, usability is defined as "*the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use*". Effectiveness is about the completeness and accuracy of the service provided by the product. Efficiency is the relation between effectiveness and the resources the user spends to use the product. Satisfaction is the user's comfort with and positive attitudes towards the use of the system (Frøkjær et al., 2000).

Ideally, any product should be engineered, thus evaluated before its release, for usability. Nielsen (1994a) proposed four usability evaluation techniques: automatic (usability measures computed by running a user interface specification through some program), formal (using exact models and formulas to calculate usability measures), empirical (usability assessed by testing the interface with real users) and heuristic (based on rules of thumb and the general skill and experience of the evaluators) evaluation. He claimed that automatic methods did not work and formal methods were very difficult to apply and did not scale up well to handle larger user interfaces. In agreement with his claim, empirical and heuristic evaluations have been the prevalent usability evaluation methods.

*The literature
favours practical
usability evaluation
by experts or users*

Regarding empirical usability evaluation, various questionnaires have been developed and used in research and industry. Prominent examples used for evaluating HCI systems are the Questionnaire for User Interface Satisfaction (QUIS, Chin et al., 1988), Software Usability Measurement Inventory (SUMI, Kirakowski and Corbett, 1993), Computer System Usability Questionnaire (CSUQ, Lewis, 1995), System Usability Scale (SUS, Brooke, 1996) and IsoMetrics (Gediga et al., 1999). Nielsen (1994a) reasoned that real users could be difficult or expensive to recruit in sufficient numbers. Heuristic evaluation, on the other hand, could be performed quickly by a small number of evaluators who are experts or knowledgeable in detecting usability problems. This is similar to tasting the salad while preparing it and noticing that it needs some salt. The main character of our story deemed himself knowledgeable in preparing an edible salad. Otherwise, he would have asked someone else who was more knowledgeable than him to taste the salad. Nielsen (1994b) proposed ten usability heuristics which have not only been applied extensively in academia and industry but have also been adapted to different contexts such as web sites, user interfaces and games. The ten usability heuristics are visibility of system status, match between system and the real world, user control and freedom, consistency and standards, error prevention, recognition rather than recall, flexibility and efficiency of use, aesthetic and minimalist design, helping users recognise, diagnose and recover from errors, and help and documentation.

We would like to note that sometimes usability is not evaluated as a whole but rather some of its components are assessed with respect to their importance for a product. For example workload is one of the usability components (Bevan, 1995) which can be measured within a usability questionnaire but also through dedicated questionnaires (e.g. the Task Load Index (TLX, Hart and Staveland, 1988) or other metrics (such as measuring physiological responses (Veltman and Gaillard, 1998))).

In developing HCI applications, especially computer software, usability has been widely accepted as a non-functional requirement (Grady,

1992). Bevan (1995) equates usability to quality of use, which he defines as “the extent to which a product satisfies stated and implied needs when used under stated conditions” and suggests it as the major design objective for an interactive product. ISO/IEC 9126-1:2001 includes usability as one of the software quality attributes along with functionality, efficiency, reliability, maintainability and portability. Note that, from time to time, there are differences in the context of usability. For example Nielsen (1993) treats efficiency as a usability aspect while ISO/IEC 9126-1:2001 treats them separately. What is common among the usability definitions, however, is that they are all concerned with the users’ cognitive skills, while they are pursuing a pragmatic goal. This very definition caused usability to be criticised by the *enjoyment movement*.

Usability is about user's cognitive skills while pursuing pragmatic goals

In one of the seminal enjoyment movement articles, Hassenzahl et al. (2001) quote the words of Robert Glass, a pioneer in software engineering, as follows: “If you’re still talking about ease of use then you’re behind. It is all about the joy of use. Ease of use has become a given – it’s assumed that your product will work.” Indeed, when we order a salad at a restaurant, how many of us are worried that its tomatoes will not be chopped or it will not have any salt (i.e. that it will not be edible)? They point out a shift from ease of use (referring to usability), to joy of use. They claim that enjoyment might improve the quality of work, especially of those involving emotions such as call center agency. They also state that from time to time ease and joy of use compensate for each other from the user’s perspective. For example one might perceive a very unusable software as a very usable one through the task-unrelated graphics, color, and music. Similarly, Overbeeke et al. (2002) propose an HCI that is more fun and beautiful. As opposed to product-centred usability, they place the human (the user) as a whole in the centre of design. They distinguish three levels of human skills which are cognitive, perceptual-motor and emotional skills corresponding respectively to knowing, doing and feeling. They note that usability deals only with the cognitive skills while “pure logic alone, without emotional value, leaves a person, or a machine for that matter, indecisive”. They suggest that we respect all sorts of user skills, including emotional, and develop products that are “surprising, seductive, smart, rewarding, tempting, even moody, and thereby exhilarating to use”. In a later article, Overbeeke et al. (2003) warn that enjoyment should not be reduced solely to fun. The distinction between fun and pleasure, two forms of enjoyment, made by Blythe and Hassenzahl (2003) explains this rightful warning. While fun is distraction from an activity, pleasure is focussing on an activity and a deep feeling of absorption. So, for example, regarding activities requiring attention, fun may not be the optimal enjoyment form but pleasure may be more desirable.

From product-centred to human-centred design

Pleasure and fun, two enjoyment types

Hopefully now it has become clear to the reader why the main character of our story placed the rose-shaped tomato skins on the top of

the salad. The peeled skins added nothing to edibility and they were not even meant to be eaten. Our character thought, however, that they would be *liked*, referring to the enjoyment discussion in the previous paragraph. Apparently, the guest was not coming (only) to satisfy her hunger. She was expecting to enjoy the meal with our character. As Overbeeke et al. (2003) point out, people are not interested in products, they are searching for *experiences*. In our story, the salad was just a mediator; it conveyed our character's thoughts and feelings for his girlfriend, probably contributing to her enjoyment of the meal. The success of the meal, where the goal was to spend enjoyable time together, does not necessarily depend on the edibility of the salad or other food. It depends on the overall experience of the dinner. So, rather than the edibility of the salad, experiences of the individuals should be evaluated. Analogically, instead of evaluating the usability of the products, their capability to provide positive UX should be evaluated.

People are not
interested in
products, they are
interested in
experiences

What is UX?

The ISO 9241-210:2010 definition of UX is "*A person's perceptions and responses that result from the use or anticipated use of a product, system or service.*" Hassenzahl and Tractinsky (2006) provide a more elaborate definition as follows:

UX is a consequence of a user's internal state (predispositions, expectations, needs, motivation, mood, etc.), the characteristics of the designed system (e.g. complexity, purpose, usability, functionality, etc.) and the context (or the environment) within which the interaction occurs (e.g. organisational/social setting, meaningfulness of the activity, voluntariness of use, etc.).

We can design 'for'
UX

Wright et al. (2003) claim that, although we cannot design an experience, with a sensitive and skilled way of understanding our users, we can design *for* experience. We call this sensitive and skilled way of understanding the users UX evaluation. Law and van Schaik (2010) propose that despite depending heavily on the user's internal state, UX "*is not overly subjective that prediction of and design for it is futile. It is something new.*" Therefore, UX evaluation is feasible and meaningful to perform. We take it one step further and claim that UX evaluation is actually necessary in product development since performance and usability, despite playing a role in its construction, cannot sufficiently represent UX. Here, we will try to explain our claim with an example from the automobile industry. BMW M5 is a high performance vehicle that has been in production since 1985. It is powered by an engine called S63 which delivers 560 hp and places the car's performance in a high position among the others in its class¹. However, just providing a high performance is not enough for a high performance car. The user (i.e. the driver) should be aware of the car's high performance; they should

Why evaluate UX?

¹ http://www.fastestlaps.com/cars/bmw_m5_f10.html

PRE-GAME UX	IN-GAME UX	POST-GAME UX
Motivations, expectations, needs, player background, ...	Usability, workload, flow, immersion, presence, engagement, ...	Emotion, affect, habituation, ...

Table 4: UX related concepts categorised temporally with respect to interacting with a game.

experience it. For this reason, BMW developed the 2012 M5 using the Active Sound Design which brings the sound of the engine into the cabin (BMW Group, 2011). This is actually against the classical active noise control research trying to deaden the noise caused by the engine, tires, wind and so on (Elliott and Nelson, 1993). We do not know (yet, as of writing these lines) whether this new design will improve the UX of the BMW M5 but this example illustrates that neither performance nor developing usable products with respect to logic, conventions or common sense can guarantee UX. Hence, UX evaluation is a requirement rather than a luxury in product development.

2.2 CONCEPTS FOR UX EVALUATION OF GAMES

In the context of computer games, we call UX player experience (PX) or gameplay experience. Just as UX, PX is a concept that is non-trivial to define and evaluate. Regarding the former, some researchers proposed structural models which aim at identifying the components of PX and their interaction (Leino et al., 2008; Nacke, 2009; Kultima and Stenros, 2010). Regarding the latter, some researchers have proposed measurement models which try to identify measurable phenomena correlated with PX (Sweetser and Wyeth, 2005; Ermi and Mäyrä, 2007; Bernhaupt, 2010). Describing or discussing PX models is out of the scope of our work. As per our purpose of evaluating UX of BCI games, we are rather interested in measurable PX concepts. But to identify these concepts, we surely benefit from UX and PX models. Next, we will discuss these concepts categorised temporally with respect to interacting with a game: pre-game, in-game and post-game UX (see Table 4). This categorisation also corresponds to the one in the PX model proposed by Fernandez (2008). Following the three categories, we will introduce an umbrella PX concept called playability and describe its heuristic evaluation.

PX: UX of games

2.2.1 Pre-Game User Experience

UX of a product is influenced by factors which exist before a user interacts with the product. At the product side, these factors are related to a product's functionality. In the context of games, Nacke et al. (2010a) list methods to evaluate a game system such as compatibility and regression testing. At the user side, as Hassenzahl and Tractinsky (2006) describe, the users' internal state before interacting with a product is highly influential on UX. These include their predispositions, expectations, needs, motivations and mood. To give an example, let us go back to our story. The user of the product, thus the girlfriend to eat the salad, was coming to dinner with some motivation and expectation. Let us assume two cases. In the first, her motivation was to satisfy her hunger and she was expecting to eat a sandwich that is similar to the one she ate the last time she was at our character's house. In the second, her motivation was to spend some time together with our character and she was not necessarily expecting any food. Let us also assume that in either case our character would prepare the same salad. In the first case, she could be disappointed by the salad, no matter how edible it was, while in the latter she could be happy, even by the outlook of the salad. So, the same salad can evoke very different feelings depending on her motivation and expectation. Similarly, a product can generate totally different UX even for the same user depending on the internal user state.

Other than the user's internal state, expertise is another factor affecting UX. Users' previous exposure to the product (Law and van Schaik, 2010) or to another product that is using the same technology may influence their experience of a product. Regarding the latter, let us consider the following example. Assume that you are satisfied with interacting with the automated teller machine (ATM) of the bank that you have been working with for some time, although you find the ATM somewhat slow. When you interact, for some reason, with the ATM of another bank you notice that you completed your operation quicker than you would with your usual ATM. This situation not only satisfies you for the moment but changes your opinion that ATM machines are slow. The next time you use your usual ATM, you would not be as satisfied as you used to be before using the ATM of the other bank. Perhaps the ATM of the other bank would become your usual ATM. A product is successful if it makes use of the technology in such a way that it can change its users' experiences positively compared to its competitors.

The matter is not 'Which technology to use?'; it is 'How to use technology?'

Bargas-Avila and Hornbæk (2011) point out that in UX practice pre-game evaluations (or, as they call it, before evaluations) are rare. An example evaluation method is SUXES (Turunen et al., 2009). This subjective evaluation method collects user expectations before using

a product and UX upon using it. Thereby, it reveals how adequate a product is in satisfying its users' expectations. In the context of games, pre-game evaluations have been conducted to identify player motivations (Lazzaro, 2004) or to show the differences between different player groups, such as males and females (Ogletree and Drake, 2007). No systematic pre-game evaluation methodology has been proposed specifically for games.

2.2.2 In-Game User Experience

In-game experience is the most widely evaluated category of UX. Some approaches evaluate in-game experience during the game (via methods such as think-aloud or psychophysiological assessment) while some evaluate it after playing the game (via methods such as questionnaire or interviews). We will describe these evaluation methods later in this chapter but first we will discuss which concepts can be considered for in-game UX evaluation.

2.2.2.1 Usability

Although UX originated from the dissatisfaction with the usability, the two are not mutually exclusive. Firstly, as Law and van Schaik (2010) also suggest, we can consider UX as an elaborated form of satisfaction, which is one of the usability dimensions. Satisfaction is about whether a product can fulfill its users' needs. User needs are sometimes pragmatic (such as opening a bottle) and sometimes hedonic (such as socialising with people) (Hassenzahl et al., 2010). In either case, perhaps more salient in the latter, satisfaction is not only about reaching an end-goal but also the experience of the user while working toward the goal. Secondly, usability can be regarded as a part of UX. Blythe et al. (2006) claim that *"it is not possible to have an engaging experience with a machine that doesn't work"*. Agreeing with this argument, we further claim that it might still not be possible to have an engaging experience with a machine that works. A product that is fully functional but, for instance, very difficult or painful to use still cannot yield a positive UX. This is similar to an extremely salty (thus non-edible) salad disturbing our story character's guest during the entire dinner and keeping her from enjoying the time she spent with him. Therefore, usability evaluation has sometimes been suggested as a UX evaluation method (e.g. Finstad, 2010, Note 3 of ISO 9241-210:2010). Hassenzahl (2004) mentions pragmatic quality (which we equate to usability here) as a construct of his UX model but he also claims that pragmatic quality alone cannot be the source of a positive UX but can rather enable *"the fulfillment of needs through removing barriers and, thus, dampening negative affect"* (Hassenzahl et al., 2010). A survey by Bargas-Avila and Hornbæk (2011)

*Usability and UX:
evil twins?*

revealed that 45% of the UX evaluation studies conducted between the years 2005 and 2009 assessed usability metrics alongside UX.

The imperfection of BCI as a technology makes the evaluation of usability a delicate matter. In §1.2.1.1 we proposed that the imperfection can be turned into a challenge for the player and can eventually provide a sense of virtuosity. But the game should still be predictable and controllable so that the player does not experience negative affect, such as frustration. It is difficult to assess the UX of the game itself, as a whole with its mechanics, narrative and the interface, if the player is stuck with the game controller.

2.2.2.2 *Flow, Immersion and Presence*

As we have mentioned in §1.2.1.1, Csíkszentmihályi (1990) introduced flow as a state of optimal experience resulting from performing autotelic activities. This means that the activity is *an end in itself*, it is intrinsically rewarding. We can consider game playing as such an autotelic activity. As we discussed in §1.2.1, people do not have pragmatic goals while playing games. The rewards, such as the sense of achievement or fantasy, are intrinsic to the player. Sweetser and Wyeth (2005) propose a model, identifying eight elements of flow in games. These are concentration, challenge, player skills, control, clear goals, feedback, immersion and social interaction. Regarding the connection between flow and UX, Jennett et al. (2008) state that flow is an extreme experience. Referring to the flow definition of Csíkszentmihályi (1990), which reads as “*the state in which people are so involved in an activity that nothing else seems to matter*”, they suggest that it is possible to have a positive UX while still being aware of things like *needing to leave the game soon in order to catch a bus*. Moreover, it may not be possible to have a game which can satisfy all of the flow elements. For this reason, many people opt for evaluating flow not as a whole but in terms of some of its elements.

Among the frequently evaluated flow elements, immersion is a prominent phenomenon. Witmer and Singer (1998) define immersion as “*a psychological state characterized by perceiving oneself to be enveloped by, included in, and interacting with an environment that provides a continuous stream of stimuli and experiences.*” A study by Cheng and Cairns (2005) revealed that immersion can even overcome the deleterious usability elements, showing the strong influence of immersion on UX. Brown and Cairns (2004) identify three involvement levels which, in chains, lead to immersion. These are engagement, engrossment and total immersion. To reach the engagement level, the player gets involved in the game in order to learn how to play it and get used to the controls. When we say involvement, we mean “*focusing one’s energy and attention on a coherent set of stimuli or meaningfully related activities and events.*” (Witmer and Singer, 1998). When the player is involved in the game such

that game controls become invisible to them and their emotions are directly affected by the game features, then they reach the engrossment level. And finally, when the player is so cut off from the reality that it is only the game that matters, they reach total immersion. We see that the description of the total immersion state is very similar to that of the flow state. Both concepts involve losing connection with reality, altered sense of time and spending attention. Brown and Cairns (2004) distinguish between the two explaining that total immersion is a *fleeting* experience.

Within their playability model, González Sánchez et al. (2009) propose a set of guidelines to evaluate game immersion. These include conscious awareness, absorption, realism, dexterity and socio-cultural proximity. Ermi and Mäyrä (2007) suggest using immersion to explain gameplay experience and propose a heuristic model. They describe three immersion categories. These are sensory immersion, related to the audiovisual appeal of the game; challenge-based immersion, in which player skills and game challenges are balanced; and imaginative immersion, in which the player enjoys the fantasy of the game. Apart from the heuristics, questionnaires (Jennett, 2009) and neurophysiological measures (Nacke and Lindley, 2008) have been proposed to evaluate game immersion. Some evaluation methods focus on specific immersion levels. For example, Brockmyer et al. (2009) propose a questionnaire to evaluate game engagement.

Immersion categories: sensory, challenge-based, imaginative

When we consider the proposed immersion elements in the context of BCI games, we see a good correspondence between them. Challenge-based immersion can potentially be provided by virtuosity-providing BCI games while imaginative immersion can be realised by fantasy-providing BCI games (recalling from §1.2.1). Thus, immersion is a highly relevant concept in the evaluation of BCI games.

Witmer and Singer (1998) stated immersion and involvement as the necessary conditions for another UX related concept, presence, defined by the International Society for Presence Research (2000) as:

Presence (a shortened version of the term “telepresence”) is a psychological state or subjective perception in which even though part or all of an individual’s current experience is generated by and/or filtered through human-made technology, part or all of the individual’s perception fails to accurately acknowledge the role of the technology in the experience.

Many authors indicated different views about the relation between immersion and presence. Brown and Cairns (2004) equate presence to total immersion whereas Witmer and Singer (1998) name immersion as a necessary condition for presence. On the other hand, Jennett et al. (2008) claim that immersion and presence can occur independently. They give the game Tetris as an example game which might not involve

Immersion and presence do not always co-occur

presence but can induce immersion. *“It is unlikely you will feel like you are in a world of falling blocks”* but the game might be *“leading to time loss, not noticing things around you, etc.”* Conversely, a player can experience presence but might not be immersed as in the case of *“carrying out a boring task in a virtual simulation”*. Regardless of their interrelation, both immersion and presence are considered as sub-optimal flow experiences (Calvillo-Gómez et al., 2010) since neither of them alone guarantees flow. Regarding presence evaluation, various methods have been suggested. van Baren and IJsselstein (2004) provide a comprehensive survey of these methods.

In our discussion within §1.2.1.2 about fantasy-providing BCI games, we claimed that the additional inner state information transferred by the BCI from the player to the game can strengthen the feeling of presence. This makes presence evaluation essential especially in evaluating UX of fantasy-providing BCI games.

2.2.2.3 Social Interaction

Context (or the environment) plays an important role in UX. When we say context, we do not refer only to the physical medium that encapsulates the user but also to the entities contained by the medium. Particularly, the interaction between the user and the other people (co-users or non-users) can provide a UX that is more than the sum of individual experiences (Battarbee, 2003). Battarbee and Koskinen (2005) call this sort of UX as co-experience and describe it as the social interaction between several people with the aim of lifting up their individual UX. Similarly, in his player experience model, Nacke (2009) introduces the framed context experience as encompassing both the individual UX and the interactions in the social context. Lindley and Monk (2008) illustrate in their model that the individual behaviors of co-presents in a group lead to a group behaviour, such as a conversation, which in turn affects the individual behaviours within the group. They call this loop of social behaviours as social enjoyment.

We understand social interaction as a series of actions and reactions of an individual, directed towards another individual. Battarbee (2003) suggests that the action for co-experience is creative and collaborative in nature, such as creating and sharing pictures with others. In gaming, this motivates the development of collaborative multi-player games. In these games, players are encouraged or enforced to collaborate and, thus, to interact. Manninen (2003) claims that in multiplayer games, the feeling of presence and the level of psychological immersion are increased due to the communication, coordination and collaboration aspects of interaction. Therefore, social actions can serve as indicators of UX. For example, Lindley et al. (2008) compared the vocal and bodily actions of co-players while they were playing a collaborative game using

*Co-experience and
social enjoyment*

*Social interaction as
indicator for UX*

different controllers. They claimed that social interaction is an indicator for engagement.

2.2.3 Post-Game User Experience

Post-game UX is about the things that happen beginning from the instant the player stops playing. Poels et al. (2010) identify two groups of post-game UX which are short-term and long-term post-game UX. We will now discuss these two groups in detail.

2.2.3.1 Short-term User Experience

Short-term UX occurs right after playing a game and is related exclusively to that specific playing session. To describe short-term UX, Poels et al. (2010) give the examples of *“the relief after passing through a difficult level, the warm feeling of having spent time with friends through online gaming, and guilty feelings after having gamed too long and as such neglected other people or responsibilities”*. If we inspect the words they mention to describe short-term UX (i.e. relief, warm feeling and guilty feeling), we see that they all refer to emotions. Emotion has been a prominent concept in UX. Hassenzahl and Tractinsky (2006) present UX as emotional usability. Forlizzi and Battarbee (2004) consider emotion as a resource for understanding and communicating UX. They suggest that pleasure is an emotional outcome of interaction with a product. Similarly, Wright et al. (2003) mention emotion within the four threads of UX which are compositional, sensual, emotional and spatio-temporal threads. Tan and Jansz (2008) explain within their game experience model that the essence of experience is emotional.

Many theories try to define emotion and determine its context. We will not provide a thorough overview of those but briefly present the one by Russell and Barrett (1999), which we adopt in our work. They see emotion as an umbrella term for at least two distinct phenomena. One of them is core affect which is *“the most elementary consciously accessible affective feelings (and their neurophysiological counterparts) that need not be directed at anything.”* The other is prototypical emotional episode that is a complex state that is concerned with a specific object and that includes *“core affect (...); overt behavior of the right sort (...) in relation to the object; attention toward, appraisal of, and attributions to that object; the experience of oneself as having a specific emotion; and, of course, all the neural, chemical, and other bodily events underlying these psychological happenings”*. Thus, core affect is a condition for prototypical emotional episode. Sometimes, affect was exclusively considered in UX research. Johnson and Wiles (2003) proposed that to experience positive affect is the utmost motivation to play games. Hassenzahl et al. (2010) showed

that positive affect and goodness are correlated and emphasised the role of experienced affect in the evaluation of a product.

Although we made the distinction between emotion and affect, some researchers use the two interchangeably. Moreover, from time to time, they use the very same methods to evaluate either of the concepts. In concurrence with the emotion theory of Russell and Barrett (1999) that we adopted, for the sake of this work, we will treat affect evaluation as a case of emotion evaluation and gather the methods to assess either of them under emotion evaluation methods. Axelrod and Hone (2006) provide an overview of methods used in identifying emotional responses. They mention bio-physiological measures (e.g. heart rate), questionnaires (e.g. the Self-Assessment Manikin (SAM, Bradley and Lang, 1994), performance measures (e.g. task completion time) and observation and coding (e.g. facial expression analysis).

Even though we included emotion as a candidate concept of post-game UX, it can also be treated as an in-game concept. It is difficult to assess players' emotions using subjective measures while they are in-play because the activities the players need to perform to manifest their emotions, such as filling-in questionnaires or thinking aloud, would interfere with the actual game play. Evaluating in-game emotions after the game ends is also not reliable as players' answers might be related to their post-game emotions. Thus, it is a more promising approach to use objective measures to assess in-game emotions. For example, Zeng et al. (2009) provide a comprehensive survey of collecting and analysing audio and visual human responses to assess emotions. Mandryk et al. (2006b) propose using physiological measures which correlate with emotion and can be acquired in-play. These include the following measures with their emotional correlates in parentheses: skin conductivity (arousal), cardiovascular measures (positive and negative emotions, stress and mental effort), respiratory measures (arousal, negative emotions) and electrical correlates of muscle activity (positive and negative emotions). Moreover, as we mentioned in §1.1.2, BCIs may provide emotional state information based on neurological emotion correlates.

2.2.3.2 Long-term User Experience

Poels et al. (2010) consider long-term UX as a consequence of repeated exposure to a game environment. In their work, they conduct interviews with habitual World of Warcraft players (those who play the game for more than 3 hours per day). Their interview analysis revealed long-term experiences such as associating game elements with real-life stimuli, daydreaming and fantasising about the game world and elements and difficulties in stepping back to the real world. In the study of Poels et al. (2010), the participants who played the game regularly were interviewed only once. This might lead to controversies regarding the homogeneity

of the participants. It is difficult to ensure that all the participants play a game under similar conditions. Some people play with others while some alone. Some play at home while some elsewhere. Thus, it is difficult to obtain a controlled experiment and make inferences about long-term UX with single interviews. Another approach that can potentially overcome this problem is conducting multiple controlled experiments and UX evaluations with each player over a long period. For example, Plass-Oude Bos et al. (2011) conducted UX experiments with participants who came into a lab and played the World of Warcraft game once a week over a period of five weeks. They claimed that this way, they could track the potential changes on UX over time.

Long-term UX is especially important while evaluating UX of novel products, such as BCI games. Novelty can induce a halo effect (Thorndike, 1920) on users, preventing them from spotting the problems which would significantly affect their UX of another, ordinary product. These problems would become more obvious as the users become more familiar with the product. Thus, it is more meaningful to evaluate the UX after the novelty of the product wears off.

Preventing the halo effect from novelty

2.2.4 Playability

Playability is not considered as a one-to-one correspondent of usability for games. Usability is related to a product's ability to satisfy pragmatic needs. However, as we surveyed in §1.2.1, the needs of people to play games are hedonic, such as to experience fantasy or to socialise. So, traditional usability heuristics are not very meaningful for games. González Sánchez et al. (2009) define playability as "*a set of properties that describe the Player Experience using a specific game system whose main objective is to provide enjoyment and entertainment, by being credible and satisfying, when the player plays alone or in company*". Some authors such as Federoff (2002) and Desurvire et al. (2004) proposed playability heuristics upon the traditional usability heuristics proposed by Nielsen (1994b). Conversely, Fabricatore et al. (2002) tried to construct a model of playability, independent of usability, and suggested a set of guidelines for game designers. Febretti and Garzotto (2009) showed that playability was positively correlated to the long-term engagement with computer games.

Playability $\neq$ Game usability

Many playability heuristics have been proposed by different researchers. In each, the context of playability has been differently considered. Some targeted the actual game the people play, independent of the software that makes it up. In this vein, González Sánchez et al. (2009) identified seven attributes of playability which are satisfaction, learnability, effectiveness, immersion, motivation, emotion and socialisation. Some researchers, on the other hand, do consider the game

as a software. They even talk about pragmatic quality and include usability as a playability factor (Desurvire et al., 2004). Furthermore, some researchers include even the input and output devices as part of a game. Koeffel et al. (2010) proposed evaluating the comfort of the physical setup in their set of heuristics. Besides the generic playability heuristics, there are those focussing on specific games, such as mobile games (Korhonen and Koivisto, 2006).

2.3 DATA COLLECTION METHODS FOR UX EVALUATION OF GAMES

Now that we have reviewed the concepts which can be used to evaluate UX, the remaining question is how these concepts can be evaluated. Actually, we have already mentioned the evaluation methods when we introduced the concepts. This section is more about the general guidelines and concerns for using different data collection methods to evaluate UX of games. By doing this, we will already start discussing the fitness of various methods to assess UX of BCIs thus obtaining a refined list of candidate data collection methods to evaluate UX of BCI games.

In their survey, Bargas-Avila and Hornbæk (2011) list the data collection methods used in UX studies. The list is headed by questionnaires and includes methods such as interviews, user observation, video recordings, focus groups, diaries and body movements. There is no one-to-one correspondence between the methods and the concepts. For example, as we discussed in the previous section, emotion evaluation can be conducted by a variety of methods such as questionnaire, observation or psychophysiological measures. Likewise, an interview can be conducted to evaluate both pre-game and post-game UX. Moreover, multiple methods can be employed together to evaluate UX to obtain a more complete picture. However, since all methods have their strengths and weaknesses, in some cases it might be favourable to use certain methods.

We can categorise the UX evaluation methods using two independent determinants. The first determinant specifies whether the method yields quantitative or qualitative data. Quantitative data can be represented in numbers, such as someone's age. It can be statistically described and compared. Qualitative data cannot be represented in numbers, such as someone's words. It is analysed using methods such as discourse analysis and grounded theory (Wertz et al., 2011). It is worth noting that qualitative data can be subjected to quantitative post-analysis, such as determining the most common (i.e. the mode) of the words pronounced within a population. But this is not how we classify the evaluation methods within this work. The second determinant specifies whether the method collects data subjectively or objectively. In subjective data collec-

*UX concepts and
UX methods:
many-to-many
relation*

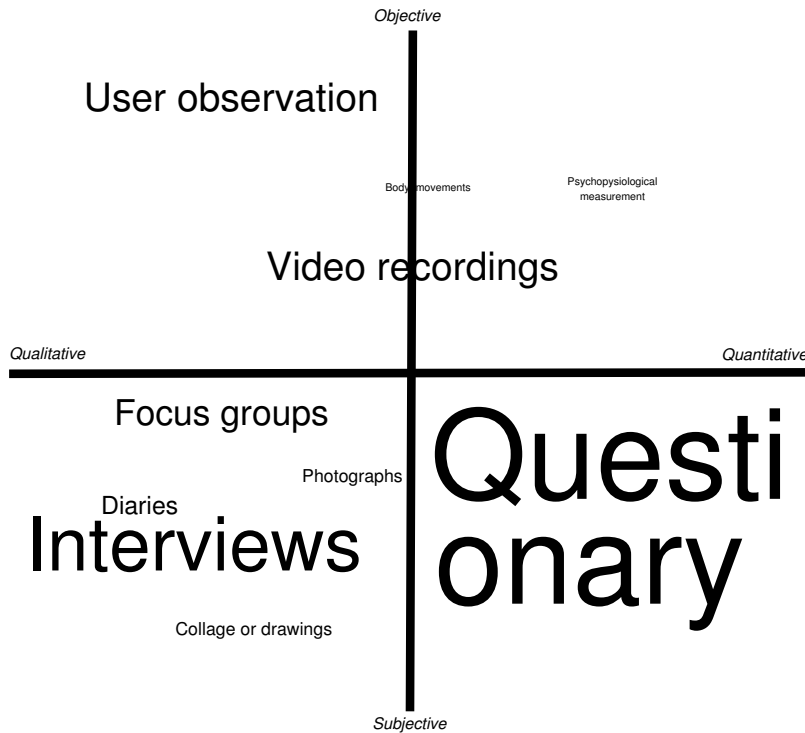


Figure 3: Data collection methods for UX evaluation, organised in four quadrants as in the work by Mandryk et al. (2006b). Horizontal axis ranges from qualitative to quantitative methods. Vertical axis ranges from subjective to objective methods. Size of the text is proportional to the number of times the method was reported to be used in UX studies (Bargas-Avila and Hornbæk, 2011).

tion, the participant under study is consciously involved in providing information about themselves. Conversely, in objective data acquisition the participant might not even be aware that data about themselves is being acquired. Mandryk et al. (2006a) constructed a Cartesian plane formed by these two determinants and placed on this plane the UX evaluation methods used for entertainment technologies. Based on the same structure, in Figure 3 we show an extended categorisation, representing the popularity of each method in terms of their frequency of being used.

We will discuss the UX evaluation methods under the quantitative and qualitative categories. Within each category, we will distinguish between subjective and objective methods.

2.3.1 Quantitative Methods

*Easy, generalisable
UX analysis*

Quantitative methods provide us with measurable data about the participant under study. Quantitative data is easy to summarise and generalise through statistics. In UX evaluation, questionnaires are the dominant quantitative tools (Bargas-Avila and Hornbæk, 2011). Ideally, every questionnaire is supported by an underlying measurement model. A valid measurement model ensures the meaningfulness and validity of the evaluations (Law and van Schaik, 2010). A measurement model contains the dimensions of a multidimensional concept (such as UX) which are called constructs or latent variables (such as immersion). Constructs are not directly measurable but are measured in terms of manifest variables. Manifest variables are the items someone sees when filling in a questionnaire. There are different ways to present the items in a questionnaire. For example they may be statements to which people indicate their levels of agreement or disagreement (known as Likert scales) or bipolar adjectives to which people indicate their tendency (known as semantic differential scale). The precision of the scale (number of points) that people use to make indications usually varies between 5 and 7. The scale may be anchored at all or some scale points with numbers, text labels or images (e.g. (Bradley and Lang, 1994)). Law and van Schaik (2010) state that the more specific the items are, the higher the correlation between the measured and the actual UX is. Therefore, instead of a small number of generic questions, a greater number of more specific questions collected under constructs are more reliable in UX evaluation. Despite the advantages we mentioned for validated questionnaires, Bargas-Avila and Hornbæk (2011) report that many researchers use self-made items or unvalidated questionnaires. They suggest that this should change for the sake of consistent and sound UX research. Finally, Law and van Schaik (2010) draw attention to the choice of the questionnaire medium. With the conveniences digital questionnaires provide, such as instant formatting of data and checking for completeness of responses, some people opt for digitising questionnaires which were originally validated in paper form. However, the digital and paper form of a questionnaire might have very different measurement properties. So, ideally, a digitised version of a questionnaire should be re-validated before use (Buchanan et al., 2005).

Questionnaires can only be filled in after one finishes interacting with a product. Thus, they are not powerful in capturing in-game UX. On the other hand, audio-visual (AV) and psychophysiological data can be collected during the game, without demanding additional effort from the participant. Regarding the former, Zeng et al. (2009) present an extensive survey of the methods that can be used to automatically recognise UX related emotional states based on behavioural cues. Regarding the latter, research has shown the correlation between UX related psychological phenomena and neurophysiological measures (Mandryk et al.,

2006b; Kivikangas et al., 2010; Dirican and Göktürk, 2011). Questionary responses are subjective and can become biased due to various factors, even simply because the respondents are aware that they participate in an experiment (known as the Hawthorne effect (Mayo, 1933)). On the contrary, AV and psychophysiological data are objective because they are natural behavioural and psychophysiological reflections of an individual. Regarding this, the latter have extra advantage because they are much harder (mostly impossible) to manipulate while certain behavioral reactions can be artificially generated and, voluntarily or not, suppressed. Despite the many potential advantages, the use of methods relying on AV and psychophysiological data have been impeded by their practical drawbacks. For example, AV recordings are subject to acoustic ambient noise and also demand that the participant remains in the angle of view. Similarly, some physiological sensors are sensitive to, even the small, body movements, which may restrict the freedom of the participant. Especially in a game this is not a favourable situation because increased bodily movements can increase the level of engagement (Bianchi-Berthouze et al., 2007). The psychophysiological sensors attached to the person under study might induce discomfort, especially when they stay on for a long time. Moreover, the measured data needs to be systematically calibrated as there is no interpersonal baseline and the intrapersonal baseline might shift during the experiment. All these practical drawbacks themselves can manipulate UX and thus make the reliability of the measurements questionable.

Games welcome movements, AV and physiological methods don't!

2.3.2 Qualitative Methods

Quantitative methods are powerful in measuring and comparing things but often we are interested in the reasons behind the resulting measurements or differences in measurements. For example, in a UX study, we might find out that someone's respiration rate increased while playing a game. We can interpret this as emotional arousal (Mandryk et al., 2006b) but why did it increase? We can answer this question by observing player-game interaction to see what was happening when the player's respiration rate increased. Alternatively, we can ask the user themselves why did their respiration rate increase. The answers we will get are unlikely to contain numbers but rather qualitative data relating to events, people, emotions and so on. So, qualitative methods can provide rich and detailed data which can explain the results provided by the quantitative methods.

Answering the question 'Why?'

Like the quantitative methods, qualitative methods can be subjective or objective. Subjective qualitative methods are those in which the participant, in written or oral ways, expresses their thoughts or emotions freely, in their own words. They can do this in response to specific

*Keeping a diary is
an endogenous
method,
interviewing is
exogenous*

questions or instructions, for example in an interview, or they can do it self-driven, for example by keeping a diary or taking photographs (Bargas-Avila and Hornbæk, 2011). Light (2006) proposes interviewing as the natural method of choice if the task is about learning about people's motivation, their emotional responses, the things they consider important in a given situation. However, she draws attention to the possibility that despite the interviewees' best intentions, their responses might become "*affected by many social and cognitive effects such as post hoc rationalisation and embarrassment, above and beyond a failure to reconstruct precise detail, or go further than rehearsed repertoires of 'stories'.*" This warning suggests that post-hoc interviews should be conducted carefully, especially while investigating in-game UX. Specialised interviewing techniques have been proposed for capturing in-game UX. The think-aloud method, in which the user provides a commentary while interacting with a product, has been criticised since the secondary task of thinking aloud can prevent or hinder the user's enjoyment of the product thus can itself manipulate the UX (Light, 2006; Mandryk et al., 2006b). Retrospective think-aloud, in which the user comments while watching the video of themselves interacting with a product, has also been criticised that watching the video would "*generate a new set of thoughts, which may interfere with recalling the original performance of the task*" (Light, 2006). Thus, the user's comments could reflect the UX relating not only to the interaction but also to watching the video (Mandryk et al., 2006b).

Objective qualitative methods rely largely on AV data collection. The collected data can be observed in real time or analysed after the interaction. Real-time analysis is useful when the experimenter would like to ask specific questions, for instance in a post-hoc interview, or adjust the experiment in relation to the participant's behaviour during the interaction. Post-hoc analysis generally involves manual coding of participant's facial, bodily or vocal behaviour. As we explained in §2.3.1, behavioural data is invaluable as it represents the participant's natural, spontaneous reactions. There are a couple of drawbacks of manually encoding behavioural cues. One of them is the vast amount of time and effort required to view, interpret and annotate such a rich data. Fisher and Sanderson (1996) report the ratio of the analysis time to the sequence time to be ranging between 5:1 and 100:1 and draws attention to the relation between analysis time and analysis quality. Another drawback is the reliability of the data. As is the case with interviews, people might suppress their behavioural cues, willingly or not, due to some factors such as embarrassment. A solution to this is to hide from the user that they are being recorded. But this is ethically problematic. People should know what kind of data has been recorded from them and how these are going to be used. A better approach might be to inform the user about the recording but conceal

*Analysis quality $\propto$
analysis time*

the recording device so that the user is less conscious of it. Another point of concern regarding manual coding of behavioural cues is the reliability of the annotations (Reidsma, 2008). Although the collected data is assumed to be objective, the coding procedure is subjective because it is carried out by human annotators. Thus, even if the coding procedure is carried out according to a well-defined protocol, annotators' interpretations might differ, consciously or not, due to personal biases or differences in their perceptions. Inter-coder (also known as inter-rater, inter-annotator, inter-observer) or synchronic agreement is a measure that describes the similarity between multiple coders' independent annotations and is used as an indicator of annotation reliability (Watkins and Pacheco, 2000). Annotation reliability is not an issue specific to the coding of behavioural cues but general to all qualitative methods. So, for example, we can talk about annotation agreement regarding annotations of interview responses as well.

HIGHLIGHTS FROM THE CHAPTER

- Humans are emotional beings so HCI should respect human emotional skills. Besides ease of use, ease of joy should be a goal.
- People are interested in experiences rather than products. We should design for UX. UX should be evaluated, rather than the product.
- UX is a complex phenomenon. It cannot be measured directly but rather in terms of other measurable concepts.
- People play computer games solely to fulfill their hedonic needs. Hedonic fulfillment is correlated to positive UX. UX is essential in game evaluation.
- Three types of UX can be observed in gaming context: pre-game, in-game, post-game.
- Pre-game UX related concepts include motivations, expectations, needs and individual characteristics. In-game UX related concepts include usability, flow, immersion, presence and social interaction. Post-game UX related concepts include emotions and habituation effects.
- UX related concepts can be evaluated quantitatively and qualitatively as well as subjectively and objectively. A multi-concept, multi-method approach to UX evaluation is desirable.
- Prominent UX evaluation methods suitable for games are questionnaire, interviewing, user observation and video recording.

EVALUATING USER EXPERIENCE OF BCI GAMES

In the previous chapters, we have introduced BCI games and UX evaluation in games. In this chapter, we will combine the two. We will start with a survey of BCI game evaluation studies to extract which aspects have been subject to evaluation. We will show that UX evaluation of BCI games is a rare practice despite all its merits we have already discussed. Then, we will explain the implications of this situation and pose our research question. Finally, we will describe our methodology and tools.

3.1 STATE OF THE ART

Now that we have discussed the merits of UX evaluation in HCI, we would expect BCIs – being HCI systems – to be evaluated with respect to UX. To verify our expectation, we took a survey approach and investigated the aspects which have recently been evaluated in the context of BCI games.

3.1.1 Data Collection

We first identified the candidate digital libraries to use in our survey. The list included the ACM DL¹, the DBLP Computer Science Bibliography², the IEEE Xplore³, Scopus⁴ and SpringerLink⁵. Since BCI is a multidisciplinary research field, the digital library we needed had to be multidisciplinary as well. Therefore, we excluded libraries such as the ACM DL and DBLP as they were indexing articles about specific domains. Since BCI articles have appeared in collections by different publishers (Hamadicharef, 2011), we excluded libraries which primarily index articles by specific publishers. For example, SpringerLink is focussed on publications by Springer whereas IEEE Xplore is focussed on those by the IEEE. From the remaining list, we opted for Scopus as it was “the world’s largest abstract and citation database of peer-reviewed literature and quality web sources”⁶. Despite being a large multidisciplinary library, there could have been imbalance between the topics covered in

¹ <http://dl.acm.org>

² <http://www.dblp.org/search>

³ <http://ieeexplore.ieee.org>

⁴ <http://www.scopus.com>

⁵ <http://www.springerlink.com>

⁶ About Scopus | SciVerse, <http://www.info.sciverse.com/scopus/about>, Accessed on 12 Feb 2012

Scopus. For this, we checked whether the list of journals publishing BCI research (Hamadicharef, 2011) were indexed by Scopus. We determined that Scopus covered 97.33% of these journals⁷ thus can represent BCI research articles well.

To capture the latest trends in BCI evaluation, we considered articles written in 2009, 2010 and 2011. We only considered journal and conference articles thus excluded review, survey and other articles potentially not reporting about a specific study with participants. We excluded in-press articles as well as those not written in English. We searched in the title, abstract and keyword fields of articles. We were interested in finding articles about BCI games so included "bci game" as well as "brain computer interface game" in our query. We also included "eeg game" and "nirs game" in the query as EEG and NIRS are the two prominent measurement methods used in BCI research in the context of HCI. Consequently, on 12 February 2012, we issued on Scopus the following query:

TITLE-ABS-KEY(("brain computer interface" OR bci OR eeg OR nirs) AND game) AND PUBYEAR > 2008 AND PUBYEAR < 2012 AND LANGUAGE(english) AND DOCTYPE(ar OR cp)

In response to our query above, Scopus returned 120 articles. The digital object identifiers (DOIs) of the returned articles are listed in Table 18. From this list, we removed the articles which we ourselves co-authored within the context of this work (#2, #3, #6 and #15). Then, we downloaded the remaining 116 articles for analysis.

3.1.2 Data Processing and Analysis

From the list of 116 articles, we removed duplicate articles and those we did not have access to. We further removed those which did not describe a BCI game. These were articles describing studies with games and/or BCIs (or EEG or NIRS) separately. A hypothetical example for such a study would be using NIRS to investigate the brain activity during playing games. At the end of all the exclusions, we were left with 31 distinct articles about BCI games (marked rows in Table 18).

For each article, we identified the concepts which were evaluated. We firstly relied on authors' own words indicating whether their study included an evaluation and, if so, what was evaluated. This information was usually contained in the abstracts of the articles but we searched through the whole articles. Secondly, we extracted the concepts which authors did not explicitly mention evaluating but did evaluate. This

⁷ The non-covered journals were Journal of Neurosciences, Proceedings of the National Academy of Sciences of the USA and Event-Related Dynamics of Brain Oscillations

EVALUATED ASPECT	RELATED CONCEPTS	N	%
Performance	Accuracy, bitrate, ITR, feeling of control, effectiveness, ...	21	68
Usability	Fatigue, workload, satisfaction, comfort, ...	6	19
UX	Fun, mood, user behaviour, ...	6	19
No evaluation	N/A	8	26

Table 5: Evaluated aspects and related concepts reported in BCI game articles along with the number of reporting articles and their percentages.

information was usually found in the analysis and/or results sections of the articles.

We organised the concepts we identified into three categories, namely performance, usability and UX. Here, we would like to note that one can place a concept into multiple categories. For example satisfaction is a concept that is related to both usability and UX. In cases such as this, we firstly considered how the authors categorise or define a concept. Only if they did not provide a categorisation, did we place the concept into the most meaningful category in accordance with our discussion in the previous chapter.

3.1.3 Results

All of the BCI games described in the articles ($N = 31$) were for single player. In relation to our categorisation of BCI games (see §1.2.2) 14 games were movement imagery games, 10 were mental state games, 4 were evoked potential games and 1 was a hybrid game. The remaining two used error potentials and actual movements.

Table 5 shows the evaluated aspects, related concepts, number of articles and their percentages for the studies described in the articles. Performance was evaluated in 21 studies while usability and UX were each evaluated in 6 studies. In 8 studies no evaluation was performed. We note that these numbers do not sum to the total number of inspected articles because in some articles more than one category of concepts were evaluated. Performance related concepts included accuracy, bitrate, information transfer rate (ITR), average number of commands, interaction duration, feeling of control and effectiveness (e.g. in the case of a therapeutic BCI game). Usability related concepts included fatigue, workload, satisfaction, comfort and usability as a whole. Finally, UX related concepts included fun, mood, user behaviour and UX as a whole.

The literature has been favouring single player games

BCI game evaluations are biased toward performance

While performance was evaluated using standard measures (Schlögl et al., 2007), such as the bitrate or the ITR, no standard method was used in usability or UX evaluations. These were usually evaluated using non-validated questionnaires or independent questions.

3.2 PROBLEM STATEMENT

McNamara and Kirakowski (2006) suggest three aspects of concern with technology which are functionality, usability and UX. These aspects correspond to those resulting from our BCI game survey, which were performance, usability and UX. What is different, though, is the amount of concern devoted to each of these aspects. McNamara and Kirakowski (2006) suggest an equal-share evaluation while our survey indicated an imbalance. The survey analysis results showed that 68% of the reported studies evaluated performance while usability and UX evaluations were each as low as 19%. We can talk about two main factors each of which play a role in the strong bias toward performance evaluations.

*Why are UX
evaluations
overlooked in BCI
game studies?*

The first is the deep-rooted trend of evaluating the performance of BCI applications. Since the time Vidal (1973) built the first BCI application, it has not been possible to achieve a BCI that can reliably replace another controller. Achieving high reliability has become the ultimate goal for researchers as disabled individuals have started using BCI systems for communication and control. Especially in life-critical BCI systems, such as a BCI wheelchair, errors are not tolerable at all. What would happen if the wheelchair turned to the wrong side at the edge of a cliff? Thus, though it is not a Millennium Prize Problem (Devlin, 2002), achieving a reliable BCI is an unsolved problem being challenged by many people. This continuous challenge has been promoting performance evaluations with BCIs. Today, BCI performance evaluations are standardised and are expected in the articles describing any sort of BCI application. While this is not, at all, a negative trend, it keeps BCI researchers from directing their efforts to anything other than performance. Apart from our survey results showing that a number of BCI researchers do care about UX, we actually believe that many others also see the value in UX evaluation. However, they are trapped between the walls of performance. They think that UX evaluation is meaningful only once the walls of performance are down. In a way, they are freezing UX evaluations until they achieve the reliable BCI. At this point, we are facing a fundamental problem. This strategy is actually delaying the creation of the reliable BCI, if it is not eradicating it completely. While developing BCI applications, developers make many choices that would maximise the performance of the application. These chains of choices, even if they agree with logic or common sense, cannot be simply assumed to provide positive UX (recall the BMW M5 story in §2.1). So,

*Performance and
UX evaluations
should be done hand
in hand*

while trying to achieve a reliable BCI, developers might end up with one that is not functioning at all because there is nobody willing to use it. Let us return to our mini story at the beginning of Chapter 2. Imagine that our character was trying different types of vinegars to prepare a better salad every time his girlfriend visited him but that actually she did not like vinegar at all. Ideally, he would be aware of this situation through her behaviour but it would only be possible if he cared about it and showed effort to understand it. Then, he would not have wasted his time with the vinegar but could have tried other ingredients to prepare a better salad. The situation is analogous to the BCI development. UX evaluation, thus the resulting opinions and feelings of the BCI users, can – and should – affect the choices that BCI developers take while building BCIs. With the ongoing BCI evaluation trend, while we are questioning the reliability of BCI today, at the time of reliable BCIs we will be questioning whether what we finally achieved has any value for the users who would have been ignored through all those years. To prevent such a disaster, it is necessary to evaluate UX *while* trying to achieve a reliable BCI.

The second factor is the development purpose of the BCI games. Including the very first BCI game, which is actually the first BCI application built by Vidal (1973), BCI games have generally been developed to demonstrate the performance of new signal analysis or machine learning methods. Hence, enjoyment has not been the primary purpose. When an application has such a pragmatic purpose, UX and usability evaluations become less relevant, even if it is a game. At this point, we are facing another fundamental problem. Due to their pragmatic purposes, so far, BCI games have been built without being paid much attention to the added-value of BCI. They have been small games that can be, or have already been, played much more accurately and useably by other controllers, for example by a mouse or gamepad. So, clearly, BCI cannot be an advantageous game controller in terms of performance or usability, unless we are talking about disabled users. Still, it might be possible that it is capable of providing positive UX. As we have explained in §2.1, performance and usability alone are not sufficient representers of UX. We even suggested in §1.2.1.1 that imperfect performance of BCI can become a challenge mechanism and contribute to positive UX. By evaluating UX, we can identify the good practices toward building better BCI games and save them from being simple demonstration tools.

Both factors we have described underline the importance of UX evaluation in developing BCI games. Our work focusses more on the second problem. In §1.2.1 we proposed that BCI games can satisfy players' motivations to play computer games in general. So, BCI games can become *games to be played* rather than *games to demonstrate*. To confirm our proposition, in this work we would like to see how, if at all, BCI

*For the disabled,
BCI can still be a
better performing
game controller*

*From 'games for
demonstration' to
'games for
enjoyment'*

control can enable a game to provide positive UX. Hence, we pursue the following research question in this work:

How does controlling a game with BCI add to UX?

3.3 METHODOLOGY OF THE THESIS

The research question we have just asked brings us back to UX evaluation. Our approach is to conduct UX evaluations on a game that can be played using BCI as well as using other controllers. This way, we can compare the UX resulting from the use of different controllers and understand the added-value relating exclusively to BCI control. Furthermore, we can discriminate the UX added by the BCI from that which the game itself already provides.

*Comparing
comparable
phenomena*

Conducting comparative research has its challenges. Perhaps the most important point of concern is to verify whether the phenomena under comparison are comparable at all. In Chapter 1 we explained that BCI is an unreliable controller compared to the others. Thus, any comparative evaluation might easily be biased against BCI simply due to its low reliability. Reliability does not necessarily vanish the true qualities of BCI but can prevent users from enjoying them and cause a floor effect in UX. As it is not easy to improve the BCI performance up to the levels of other controllers, an equivalent approach is to decrease the performance of other controllers down to the level of BCI. But simply deteriorating the recognition performance might cause an unfair bias, this time against the other controllers. In a study by Klimmt et al. (2007) participants played a computer game which, deliberately and with certain probability, processed player actions erroneously. Compared to the standard (non-erroneous) version of the game, participants reported significantly decreased level of enjoyment. In the same study, participants also played a version in which the task was harder than the standard version but no erroneous processing was done. This version did not cause a decrease in enjoyment level compared to the standard version. The results of this study suggest that, in reliability leveling, instead of artificially introducing errors to the controller, the tasks should be made challenging. The users should not conclude that the errors they face are solely due to the incapability of the controller they are using but that their actions also play a role in errors. In this work we try both approaches. In the first study, we compare the BCI version of the game to the conventional – thus too reliable to be comparable – mouse controlled version. Moreover, in all studies, we compare it against leveled – thus comparable in reliability – controller versions. The leveling procedure is explained in detail in §3.3.1.

*BCI game
evaluation beyond
performance*

Another confounding factor that can unjustly bias the evaluations is the halo effect caused by the novelty of BCI technology to the players

(recall from §2.2.3.2). We are not interested in one-time UX resulting from a single experiment with a particular BCI game but rather in players' opinions and feelings about BCI as a technology. The players might evaluate BCI favourably simply because they find the technology to be novel and the idea of controlling things solely through mind to be 'cool'. We do not object to the sincerity of such evaluations but we doubt their power in reflecting the reality. The attributes such as novelty and coolness are easily worn out by time. So, ideally, for any product, they should not be the main attributes providing a positive UX. In this work, we are interested in such BCI attributes which do not get worn out by time. One approach would be to conduct experiments over a long time period to identify the vanishing and remaining attributes but this is rather a time consuming approach. We take an alternative approach and compare BCI to both traditional and novel controllers. In our second study, we compare BCI to mouse control, which is a traditional game controller, while in the third study we compare it to an automatic speech recogniser (ASR), which is a novel game controller.

*BCI game
evaluation beyond
novelty*

While UX evaluation is already a non-trivial process, as we discussed in Chapter 2, it is even more difficult in the context of BCI games since the related research is very immature, as shown by our survey. While there are standard measures and methods to evaluate performance, neither of these have been established with respect to UX. Therefore, in our approach, we try different combinations of existing UX concepts and evaluation methods to evaluate UX of BCI games. As it would be difficult, even infeasible, to consider all the combinations of the concepts and methods in a single study, we consider different combinations of them across several studies. Besides UX related concepts and measures we also consider performance and usability evaluations to investigate the interplay between the three.

3.3.1 Equalised Comparative Evaluation

Equalised comparative evaluation (ECE) is a method we propose to compare two or more recognition technologies independent of their performances. In contrast to deliberately introducing errors to equalise the recognition performances of technologies, our approach equalises the *user-contributed* recognition performances of the technologies through parameter-adjusted task difficulty.

*User-contributed
recognition
performance*

The steps of our approach are as follows:

1. For each technology, identify the set of task-related parameters affecting performance.
2. For each technology, determine the parameter values that yield similar performance.

- a) If possible, make use of the literature to determine the parameter values.
- b) Else, conduct an experiment to identify the parameter values.
 - i. For each technology and parameter, construct a set of candidate parameter values according to literature, expertise or common sense.
 - ii. In a cued study, where ground-truth about user actions can be known, let people use each technology with each parameter value (or each combination of parameter values, if there are more than one parameters) and compute the performance (e.g. in terms of accuracy, recall and so on).
 - iii. Select the (combination of) parameters for each technology which yield the most similar (e.g. have high correlation, low difference and so on) and highest performance values.
3. Let people use the technologies with the selected parameter values and evaluate the attributes beyond performance (e.g. UX, usability and so on).

The steps of our approach might be clearer to the reader with an example application of them. Two such examples can be found in §4.2.2 and §6.2.1.

3.3.2 *Experimental Game: Mind the Sheep!*

Mind the Sheep! (MTS!) is a computer game that we developed for our experimental purposes. The game world contains a number of sheep that move autonomously, dogs that are moved by players, fences and other elements representing a meadow (see Figure 4 for the default game world with 10 sheep and 3 dogs). The aim of the players is to fence the sheep in as quickly as possible by herding them with their dogs. There are two main actions the players need to take to move a dog: selecting a dog and providing a target location. Selecting a dog can be done with the mouse, BCI or ASR. The variety of choices to make selections enables us to conduct comparative research. Providing a target location is always done with the mouse, by left-clicking on the desired location. Depending on the game version, these actions can be performed in any order as well as simultaneously. While a dog is on move, another or the same dog can be re-selected and provided with a new target location. The game can be played by a single player as well as by multiple players. Next, we will describe the different versions of

*A game capable of
accommodating
many controllers*

the game with respect to the controller used to select the dogs, followed by the multi-player version description of the game.



Figure 4: A screenshot from the game Mind the Sheep!. The game world contains 3 black dogs, 10 white sheep, a fence and other elements representing a meadow. The number of sheep to be fenced in are shown in the top-right corner.

3.3.2.1 *Point-and-Click Version*

In this version of the game (MTS!-P&C), players first click on the dog they wish to select and then on the location they wish to move the dog to. Then the selected dog starts moving to the provided location. As both selecting a dog and providing a location require mouse clicking, these actions can only be performed sequentially, the former always preceding the latter.

3.3.2.2 *BCI Version*

This version (MTS!-BCI) of the game is based on SSVEP evocation so it is an evoked response game. We chose to develop an SSVEP game due to its extra advantages compared to the other neuromechanisms (recall from §1.2.2.3). To select a dog, players keep the left mouse button pressed. As soon as the button is pressed, the dog images are replaced by circles that flicker (alternate between black and white colour) at

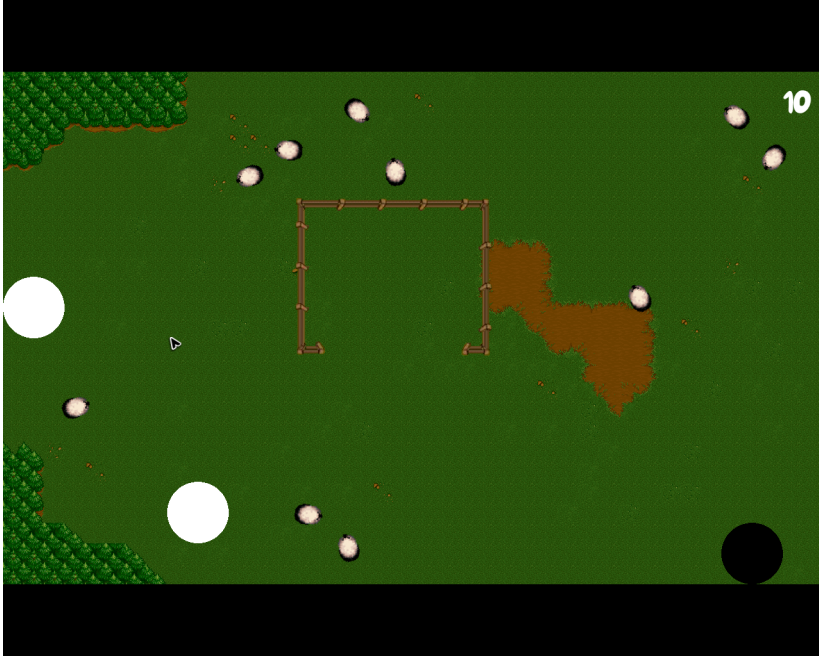


Figure 5: A screenshot from the game Mind the Sheep! while the SSVEP stimulation is on.

distinct frequencies (see Figure 5). Also, the collection of EEG data from the players is started. To select a dog, the players focus their vision on the flickering circle that replaces the dog they wish to select. The flickering keeps on until the mouse button is released. When the mouse button is released, the EEG data that has been collected until then is analysed and based on the analysis one of the dogs – hopefully the one the player wished to select – is selected. The selected dog moves to the location where the cursor is located at the time the mouse button is released. The longer the player keeps the mouse button pressed and focuses on flickering (i.e. the more is the data that BCI collects), the higher is the probability of selecting the correct dog. But also, the longer the player waits, the more the game state will change and the more delayed the player actions will be. So, in principle, there is a trade-off between making a correct selection and making a quick selection. This makes MTS!-BCI a challenge-based BCI game (recall from §1.2.1.1). Unlike MTS!-P&C, providing a location and selecting a dog can be done simultaneously in this version.

*A challenge-based,
evoked response
game*

In all the studies reported within this work, EEG analysis is done as follows. EEG signals are acquired and recorded by a number of BioSemi Active-electrodes placed in contact with the scalp. The number and locations of the electrodes vary per study. The continuous signals are

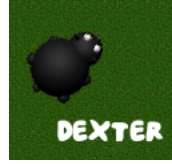


Figure 6: A dog image from MTS!-ASR accompanied by a name.

digitised at a sampling rate of 512 Hz using the BioSemi ActiveTwo system. No further processing is carried out in the hardware. In the game software, the digitised EEG data are processed using canonical correlation analysis (CCA) (Lin et al., 2006) including the three harmonics of each flicker frequency (i.e. fundamental frequency, second and third harmonics). The dog with the frequency that yields the maximum correlation in CCA is selected by the game.

3.3.2.3 *Timed Selection Version*

In this version (MTS!-TS), players keep the left mouse button pressed as in MTS!-BCI but the circles replacing dog images do not flicker simultaneously. Instead, as long as the button is pressed, the circles are displayed sequentially, in random order, for a period of time. The display period increases at a standard rate. To select a dog, the players need to release the mouse button at the time the dog they wish to select is displayed. The selected dog moves to the location where the cursor is located at the time the mouse button is released. Due to the increasing display period, the longer the players keep the mouse button pressed, the higher is their chance of selecting the right dog. So, in terms of the challenge mechanism, MTS!-TS is analogous to MTS!-BCI. As in MTS!-BCI, providing a location and selecting a dog can be done simultaneously in this version.

Trade-off between accuracy and speed, as in MTS!-BCI

3.3.2.4 *ASR Version*

We can see ASR as a novel game controller, just as BCI. In the ASR version of the game (MTS!-ASR) the dog images are accompanied by dog names (see Figure 6). The player keeps the left mouse button pressed and pronounces the name of the dog they wish to select. Meanwhile the acoustic data is collected. When the mouse button is released, the collected data is analysed and based on the analysis one of the dogs is selected. The selected dog moves to the location where the cursor is located at the time the mouse button is released. As in MTS!-BCI, providing a location and selecting a dog can be done simultaneously in this version.

A version comparable to MTS!-BCI in terms of novelty

In MTS!-ASR acoustic data analysis is done as follows. Acquired acoustic data is recorded, processed and analysed using the CMU Sphinx speech recognition toolkit (Walker et al., 2004). The Wall Street Journal model supplied within the Sphinx toolkit is used for the dictionary and the acoustic model. Word level unigrams representing the dog names are used to form the language model. Then, Sphinx constructs a search graph using the acoustic model, the dictionary and the language model. The signal processing pipeline consists of a pre-emphasiser, raised cosine windower, discrete Fourier transform, mel frequency filter bank, discrete cosine transform, cepstral mean normalization and feature extraction. Decoding is performed by a frame synchronous Viterbi search on the constructed search graph using the extracted features. The dog with the name matching the result of the decoding is selected by the game. If there is no result at the end of decoding (e.g. in the case of silence), no action is taken.

3.3.2.5 *Multimodal Version*

In multimodal version of MTS! (MTS!-MM), players can switch between the aforementioned versions of the game by pressing the Ctrl key on the keyboard. As soon as the Ctrl key is pressed, the game world is redrawn to match the corresponding version. For example, when a player presses the Ctrl key to switch from MTS!-ASR to MTS!-BCI, the dog names are removed from the game world. MTS!-MM can be configured to contain all the MTS! versions or a subset of them.

The term multimodal should not confuse the reader into thinking that the aforementioned versions of the game were unimodal games. All MTS! versions are multiplayer games in which players use the mouse to give directions to the dogs and another modality (e.g. BCI in MTS!-BCI) to select dogs. We use multimodal here in reference to the selection task. In the aforementioned versions, selections are made with a single, dedicated control mechanism (modality) whereas in MTS!-MM players can choose between the modalities to make selections.

3.3.2.6 *Multi-Player Version*

The MTS! versions that we have described so far were for a single player. But any of them can be played by two people as well. In multi-player MTS!, players have their own sets of dogs (see Figure 7) but they play in the same game world and cooperate to fence the sheep in. Each dog has a name tag under it (not shown in the figure) for easy referring. The goal of the game is the same; fencing the sheep in as quickly as possible. The players should ideally be co-located to be able to cooperate as the game itself does not provide any means of communication. But they do not necessarily need to share the same screen. Especially in MTS!-TS



Figure 7: A screenshot from the multi-player version of the game Mind the Sheep!. One of the dogs of one of the players is marked with number 1 and that of the other player is marked with number 2.

and MTS!-BCI the amount of visual load doubles and might become excessive. So, players can play on separate screens. This sort of setup in which each of multiple users has *“separate control over an identical version of the task ... within their own private screen space, that is visible to both participants”* is known as the Separate Control of Shared Space architecture and has been found to encourage children to contribute to a computer-based task (Kerawalla et al., 2008).

Yuill and Rogers (2012) identify two main approaches to collaboration design. On the one hand there are the approaches that enforce collaboration, for example through turn-taking, so that everybody is given an equal chance to contribute. On the other hand are those that give away control in such a way that individuals act independently and collaborate only if they wish. In between these two extremities, they identify a third design approach which encourages collaboration, but does not enforce it, by *“providing an added benefit, or incentive, for users if they work together”*. In multi-player MTS! we did not enforce collaboration either. If one player stops playing the game, the other one can still keep on playing and has some chance to finish the game. However, if players play together, they are supposed to finish the game quicker. Therefore, our collaboration design closely follows the third – encouraging – approach.

*Encouraging
collaboration rather
than enforcing it or
setting it free*

STUDY	APPROACH	GAMES	CONCEPTS	METHODS
1	non-ECE, ECE	MTS!-BCI, MTS!-P&C, MTS!-TS	Social inter- action	Observational analysis, questionary, interview
2	ECE	MTS!-BCI, MTS!-TS	Immersion, affect, per- formance	Questionary, log analysis
3	ECE	MTS!-BCI, MTS!-ASR, MTS!-MM	Workload, engagement, product quality, per- formance, expectation, modality switching	Questionary, log analysis, interview

Table 6: Summary of the studies conducted within this work. The columns show whether ECE was done, which game versions were used, which UX related concepts were evaluated and which UX evaluation methods were used.

3.4 OVERVIEW OF THE THESIS

The rest of this work is organised as follows. Part II contains 3 chapters each of which describe a UX evaluation study conducted for answering our research question. For each study, we describe the purpose, related work, tools used, experimental setup, analysis, findings and implications. In Table 6 we provide a summary of these studies. We specify whether we did an ECE (see §3.3.1), which versions of the game MTS! were used (see §3.3.2), which UX related concepts were evaluated (recall from §2.2) and which UX evaluation methods were employed (recall from §2.3). In Part III, we discuss the collective implications of the studies in relation to our research question, point out the limitations of the work and bring up possible future directions.

3.5 CONTRIBUTIONS OF THE THESIS

The contributions of our work are manifold. In §3.2, we claimed that BCI researchers seem to be freezing UX evaluations until they achieve the reliable BCI but doing the former is actually hindering the occurrence of the latter. As achieving the reliable BCI is not so easy, an

alternative might be to simulate it and so already start considering other factors, such as UX. However, the simulation of the case “*What if the BCI performs reliably?*” is not easy. Since there is no ground truth about a user’s input to a BCI (i.e. the user’s intent or state), a classical Wizard-of-Oz simulation (Salber and Coutaz, 1993) of a reliable BCI is not feasible. Instead, with our ECE approach, we propose a simulation of the case “*What if the BCI performs as reliably as other modalities?*”. This case actually seems more useful as the current dominant trend in BCI research is to replicate other modalities’ functionalities using BCIs.

Simulating the future BCIs

As there are no guidelines established for evaluating the UX of BCI games, our work can be regarded as an exploratory one that identifies the preferable (and non-preferable) measures and methods of doing it. This can encourage BCI game developers to evaluate, and perhaps even advertise, the UX of their products. If UX becomes a standard evaluation matter, just as performance, then the developers can start enjoying its benefits which we demonstrated throughout Chapter 2 in the context of games.

Exploring the ways to evaluate UX of BCI games

The UX evaluation studies we conduct on our experimental BCI game reveal the strong and weak features related not only to our game but also to SSVEP and/or challenge-based BCI games in general. These can shape the way people develop BCI games. Furthermore, as showed by our mini survey reported in this chapter, no multi-player BCI games have been reported in the literature within the past 3 years. Our experimental game sets an example of a multi-player BCI game and its evaluation shows the points of concern in developing such games.

Challenge-based, evoked response games: any good?

The two frameworks we propose in Chapter 1, the BCI framework (Figure 1) and the updated PC framework (Figure 2), let BCI developers situate their applications among the others and make them aware of the decisions they need to make and the options they can choose from while developing BCI applications. Similarly, the categorisation of BCI games that we propose according to player motivations (§1.2.1) and neuromechanisms (§1.2.2), let BCI game developers describe their games easily and show them the potential capabilities and drawbacks of their games.

HIGHLIGHTS FROM THE CHAPTER

- BCI game literature shows a strong bias toward performance evaluation and single player games.
- BCI cannot provide an advantage over the other game controllers in terms of performance. But it might provide a better UX.
- The ECE approach we propose enables evaluating BCI independent of its reliability and simulating an ideal scenario.

- The game that we developed, MTS!, is an experimental platform suitable to conduct ECE.

Part II

STUDIES

In this part, we will describe the three studies we conducted in order to answer the research question we have formulated in Part I. We will discuss each study in a separate chapter and for each study we will describe its purpose, the related work, the methodology we followed, the experiment we conducted and the implications of our findings.

EVALUATING SOCIAL INTERACTION AND CO-EXPERIENCE

The purpose of the study that we will report in this chapter was twofold. Firstly, we wanted to investigate how BCI control influenced the social interaction between collaborating players. As we explained in §2.2.2.3 social interaction catalyses co-experience, which, in turn, influences individual UX. So, analysing social interaction can provide us with clues about UX. Secondly, we aimed at showing the benefits of the ECE approach (see §3.3.1). We compared social interaction during BCI control to two other control mechanisms. In one of our comparisons we did an ECE while in the other we did not. This way, we could also assess BCI control independent of its performance.

4.1 RELATED WORK

Next, we will describe the work related to our study. We will start with previous HCI studies on social interaction evaluation and then continue with UX studies on multi-player BCI games.

4.1.1 *Social Interaction Evaluation*

Social interaction is made up of social behaviours of multiple individuals. Social behaviour can be explicitly directed to other people, as in the case of asking a question or nodding in agreement by looking at a person's face. But it can also be implicitly directed. For example, you may ask yourself a rhetorical question while reading some text, without looking at a co-present or calling their name. Although the question is not directed, it can still be responded to by the co-present, as you had actually intended (Heath et al., 2002).

*Explicitly and
implicitly directed
social behaviour*

As the reader might have already noticed, the social behaviours we have just mentioned contain either vocalisations (e.g. asking a question) or gestures (e.g. nodding). Lindley et al. (2008) suggest categorising and coding player vocalisations and gestures of interest in collaborative multi-player games according to the definitions in the Autism Diagnostic Observation Schedule (Lord et al., 2000). Player vocalisations can emerge in forms of speech or utterances. In a collaborative game, speech potentially contains collaborative communication, such as developing and executing a joint strategy. Obviously, it can contain non-collaborative communication as well, such as humour. Utterances,

*Co-experience =
Collaborative
Interaction +
Emotional
Interaction*

such as a laughter or a pause, potentially manifest inner state information, such as enjoyment or puzzlement. Gestures can be classified as instrumental or empathic gestures. Instrumental gestures are actually those that aim to change the immediate behaviour of someone, such as holding an index finger on one's lips to mean 'be quiet' (Attwood et al., 1988). But, Lindley et al. (2008) keep this category broader and include deictic gestures (pointing) and response gestures (nodding, shrugging and so on). So, instrumental gestures can be generalised to those that facilitate collaborative interaction, just as speech does. Empathic or expressive gestures are those which express inner feeling states or respond to feeling states of others (Attwood et al., 1988). Empathic gestures include covering one's face with one's hands in the case of embarrassment or putting an arm around someone for consolation. So, speech and instrumental gestures mainly mediate collaboration while utterances and empathic gestures carry emotional information. Collectively, they construct the co-experience.

*Non-directed
behaviour and social
interaction*

There are times when we produce an utterance or an empathic gesture without directing it to someone, or even unintentionally. For example, a pause can be used deliberately as a conjuncture but can also be produced involuntarily out of hesitation (Rochester, 1973). Similarly, you might cover your face reflexively when you lose in the game but also deliberately to show your team mate that you accept that your mistake caused your team to lose. Even in these cases, non-directed behaviours can have social implications. When you produce a pause, your conversational partner might help you out of your hesitation or when you cover your face your team mate might try to console you. Thus, in evaluating social interaction and co-experience, non-directed behaviours are definitely worth consideration. In our study, we followed the categorisation of Lindley et al. (2008) to code social behaviours and we treated directed and non-directed behaviours equally.

There are other approaches to assess social interaction and co-experience. Lindley and Monk (2008) analysed the number of turns, turn overlaps, conversational equality and conversational freedom as indicators of social enjoyment while people were interacting with an electronic photograph display. Their work aimed at assessing the formality of conversations with respect to different seating arrangements and distributions of control. Battarbee (2003) analysed the social interaction within a group of friends through the multimedia messages they exchanged. She observed and categorised the message content into categories such as communication, greeting and humour.

4.1.2 *UX Evaluation with Multi-Player BCI Games*

Not as much research has been conducted for multi-player BCI games as for single-player BCI games. In a study conducted by O'Hara et al. (2011), groups of participants were asked to videotape themselves while one of the group members was playing a BCI (mental state regulation) game in a social setting. Then, their recordings were analysed to investigate how people use bodily actions to facilitate control of brain activity and how they make their actions and intentions visible to, and interpretable by, others playing and watching the game. The analysis results showed that players tried to provide explicit cues and explain their actions to compensate the lack of visible embodiment or other indicators of their thoughts. Moreover, they used bodily action (e.g. postures, gestures) as strategies to influence their brain activity. All of these results point out the positive influence of bodily movements while playing BCI games in social contexts.

In the study of O'Hara et al. (2011), only one person played a game using BCI within a group of people. In our study, both people belonging to a pair played a game using BCI. Therefore, our study is the first one to evaluate the UX of multiple people who play at the same time using BCI.

4.2 METHODOLOGY

We let people play different versions of MTS!. We evaluated social interaction during each game as described in §4.2.1. We compared social interaction in MTS!-BCI to that in MTS!-P&C and MTS!-TS. For the latter, we did an ECE, the details of which we explain in §4.2.2. We conducted a pilot study to optimize the SSVEP recognition performance in MTS!-BCI (see §4.2.3).

4.2.1 *Evaluating Social Interaction*

We captured player vocalisations and gestures as audio-visual (AV) data. We logged the completion time of each game as an indicator of player performance. At the end of the experiment, players filled in a questionnaire. The questionnaire items were statements with a 7-point Likert scale anchored at the extremities with strong disagreement and strong agreement. The items were "I felt inclined to work together with my partner during this experiment" and three times "I found it difficult to select a dog with *controller*" where each time *controller* referred to a different control mechanism (i.e. P&C, TS, BCI). The former item tested whether the game was able to induce social interaction while the latter items tested the levels of perceived control during different games.

A last item asked players to rate the games with respect to the level of collaboration they had with their team mates. We also conducted an interview to learn more about participants' playing strategies and collaboration rankings of different control mechanisms.

We used Anvil (Kipp, 2001) to code player vocalisations and gestures in the AV data, according to the coding scheme we described in section 4.1.1. For the former, we marked the onset and offset timestamps while for the latter we marked only the onset. We identified the average total lengths of speech and utterances a pair produced per game and we normalised the values to seconds per minute, as game completion times could vary between pairs. We computed the average numbers of instrumental and empathic gestures and we normalised them to number of times per minute. We extracted game completion times from the logs. To summarise vocalisations, gestures, game completion times and questionnaire responses we computed median¹ and interquartile range (IQR)² values. To summarise participants' ratings of games with respect to the level of collaboration they had with their team mates, we computed each game's mean reciprocal rank (MRR)³ (Voorhees, 1999).

The game version (i.e. the control mechanism) was the independent variable while the dependent variables were the normalised lengths of speech and utterances, normalised counts of instrumental and empathic gestures, game completion times and questionnaire responses. For questionnaire responses, the analysis unit was the participant while for the other dependent variables the pair was considered as the analysis unit. We compared the analysis results of MTS!-BCI to those of MTS!-P&C and MTS!-TS in order to assess the effect of perceived control on dependent variables. The significances of differences were assessed by Wilcoxon signed-rank tests ($p < 0.05$).

BCI control as the
baseline

4.2.2 ECE Details

We followed the ECE steps described in §3.3.1 as follows. In Step 1, we determined the task-related parameters affecting performance. For MTS!-BCI this is the stimulation duration while for MTS!-TS these are the initial stimulus display period and the increase rate of the display period. Since stimulation duration in MTS!-BCI is determined by the player, it could not be fixed. Therefore, we tried to adjust the MTS!

-
- ¹ Median, a measure of centre, is the midpoint of a distribution such that half of the observations are smaller and the other half are larger than it.
 - ² IQR, a measure of variation, is the difference between the first and third quartiles. The first (third) quartile is the median of the observations whose position in the ordered list is to the left (right) of the location of the overall median.
 - ³ Given a ranking of items, the reciprocal rank of an item is the multiplicative inverse of its rank. Given rankings of items by multiple rankers, the MRR of an item is the mean of the item's reciprocal ranks by all rankers. The greater is the MRR of an item, the better is its rank.

TS parameters so that the maximum correct selection chances were equalised in MTS!-BCI and MTS!-TS.

In Step 2, we resorted to the literature to adjust the parameters. The CCA algorithm we used in SSVEP detection was found to converge its maximum performance (in terms of average accuracy) after a stimulation period of 2.25 seconds (Lin et al., 2006). So, in MTS!-BCI, after 2.25 seconds of attending to the stimulation, the players should have the highest chance of making a correct selection. This means that, also in the MTS!-TS, the players should have the highest chance of making a correct selection after 2.25 seconds. In MTS!-TS, making a correct selection depends on the timing of the players to release the mouse button as soon as the circle for the dog they wish to select appears on the screen. If the circle remains on the screen long enough for players to react – in other words, if it remains as long as the human visuomotor reaction time – then the players can make a correct selection. There are different reports about human reaction time as it is dependent on the individual, stimulus and the context. For our study, due to its similarity, we relied on the experiment by Lansing et al. (1959) in which participants reacted to light stimulus by button presses. The experiment results showed that for alerted individuals the mean reaction time was 225 msec. So, in MTS!-TS, with a display duration of 225 msec the players should have the highest chance of making a correct selection. But this display duration should be reached only after 2.25 seconds so that the correct selection chances are equalised for MTS!-BCI and MTS!-TS. Therefore, the display duration starts with 100 msec and increases 5% at each subsequent highlight so that it reaches 225 msec after 2.55 seconds – thus, not equal but close to 2.25 seconds. The increasing stops when the display period reaches 500 msec. While our approach potentially levels the best selection probabilities during MTS!-BCI and MTS!-TS, we acknowledge that the probability distributions of the correct selections may not be similar for the two games. However, equalising the probability distributions is a non-trivial task and requires a separate, controlled study.

In Step 3, we evaluated social interaction in both games as we described in §4.2.1.

4.2.3 *Pilot Study: Optimising BCI Performance*

In §4.2.2 we explained that the SSVEP recognition algorithm we used reaches its maximum performance after 2.25 seconds. However, the value of the maximum performance depends on a number of other parameters. We conducted a pilot study to identify the parameter values that yield the maximum performance. As parameters, we considered stimulus (i.e. flickering circle) diameter, stimulus frequencies, electrodes

to use for analysis and re-referencing method. For stimulus diameter, we tested for 2 cm and 3 cm. For stimulus frequencies, we tested for 6 Hz, 6.67 Hz, 7.5 Hz, 8.57 Hz, 10 Hz, 12 Hz and 15 Hz. For analysis electrodes we tested for the sets $S_1 = \{PO_3, O_1, O_z, O_2, PO_4\}$ and $S_2 = S_1 \cup \{P_3, P_z, P_4\}$ (for electrode locations see Jasper, 1958). For the re-referencing method we tested for common average referencing (CAR) and linked ears (McFarland et al., 1997). Seven people participated in the study. The setup and procedure is described elsewhere in detail (Hakvoort et al., 2011).

For each combination of parameters, we computed the recall⁴ of the recognition performed by the BCI as described in §3.3.2.2. The greater diameter size (3 cm) consistently yielded a significantly higher performance than the smaller one (2 cm). Regarding analysis electrodes, the set with the smaller number of electrodes (set S_1) performed as well as the other one (set S_2). Similarly, the re-referencing method requiring the smaller number of electrodes (linked-ears) performed as well as the other one. Therefore, we fixed diameter size to 3 cm, analysis electrodes to set S_1 and re-referencing method to linked-ears, and we reduced our analysis to stimulus frequencies. Since there are three stimuli in the game, we computed the average recall for all the 3-combinations of stimulus frequencies (see Table 19). We decided to use 7.5 Hz, 10 Hz and 12 Hz which yielded the maximum average recall of 83.62%.

4.3 EXPERIMENT

Next, we will describe the setup of the experiment we conducted, the participants who took part in the experiment and the results of our data analysis.

4.3.1 Setup

The setup consisted of five computers: two for the participants to play on, two for the EEG data processing and one for the recording and storing of AV data. The participants were seated next to each other, as seen in Figure 8, so bodily interaction, such as pointing, was possible while playing the game. They both looked at their own LCD screens which were placed approximately 50 cm apart from each other. This gave the participants the opportunity to see each other's screen. We informed the participants that their bodily movements could hinder the BCI performance but did not instruct them to refrain from movements. So players needed to take care of the trade-off between BCI performance

⁴ Recall is the number of true positives divided by the total number of items that actually belong to the positive class. A recall of 100% means that every item from a class was labeled as belonging to that class.



Figure 8: One player pointing at his team mate's screen while the team mate is following him.

and collaborative interaction. The EEG caps and electrodes were placed on participants at the start of the experiment and removed at the end of the experiment. A camera and a microphone were placed in front of each participant but behind their screens.

Each pair started with a short training to learn the game and the three different selection methods. Once the training was finished, they played three versions of the game (MTS!-P&C, MTS!-TS and MTS!-BCI) in random order. Each game lasted until all the sheep were fenced in or a time limit of 20 minutes was reached. Each game was played on a pre-made map that differed across the games to prevent familiarity of the participants with the game world. The maps differed only with respect to the layout so that the difficulty was not altered. Maps were assigned randomly to games. During the whole experiment, the experimenter stayed in the same room with the participants. At the end of the experiment, participants filled in the questionnaire and took part in an interview.

4.3.2 *Participants*

Twenty participants (2 female), divided into 10 pairs, took part in the experiment. They had an average age of 25.25 ($\sigma = 7.20$), ranging from 18 to 54 years. They had normal or corrected vision, used a computer every day and had at least some experience with computer games. One pair of participants spoke Romanian as mother tongue while the rest were native Dutch speakers. We asked all participants to bring a friend.

If no friend was available, we teamed them up with another participant. Familiarity of pair members was not a confounding factor due to the within-pair design of the experiment (i.e. all pairs played all games). The participants participated voluntarily in this study, and signed a consent form for their participation. To motivate the pairs to do their best, we promised a pair of cinema tickets to the pair that completed all the games within the shortest time. We instructed the players to talk in their native language during the game.

4.3.3 Results

On the 7-point Likert scale, participants rated the questionnaire item "I felt inclined to work together with my partner during this experiment" with a mode of 7 (9 out of 20 answered with a 7) and a median of 6. Using one-sample Wilcoxon signed-rank test, we compared the ratings to the median of the scale (that is, 4) assuming that it represented an average level of collaboration. The test indicated that the game induced a high level of collaboration ($Z = 3.98, p < 0.001$).

Table 7 shows the median and IQR values for speech and utterance lengths, instrumental and empathic gesture counts, game completion times and difficulty ratings for MTS!-BCI, MTS!-P&C and MTS!-TS. The results which are significantly different from MTS!-BCI results are marked with an asterisk.

Participants produced a significantly greater number of utterances and significantly greater number of empathic gestures during MTS!-BCI than MTS!-P&C. They also produced less speech during MTS!-BCI than MTS!-P&C, though the difference was marginally significant ($p = 0.059$).

Participants found selecting dogs significantly easier in MTS!-P&C than in MTS!-BCI but no difference was observed between MTS!-BCI and MTS!-TS. So, our effort to equalise perceived controllability of MTS!-BCI and MTS!-TS was effective. Despite the lower difficulty of MTS!-P&C, there was no significant difference between the games in overall performance, in terms of game completion times. This suggests that, with all control mechanisms, regardless of the difficulty in control, the participants were able to come up with strategies to remain in control. This is actually a desirable situation because according to the flow theory (Csíkszentmihályi, 1990), the players should experience some challenge during play but the challenge should match their skills and should be manageable.

The MRRs of the games with respect to the subjective level of collaboration were 0.46, 0.82 and 0.55 for MTS!-BCI, MTS!-P&C and MTS!-TS respectively. So, participants reported that their level of collaboration was the highest in MTS!-P&C and the lowest in MTS!-BCI.

The game MTS! did induce collaboration

BCI control was more difficult but still manageable

Participants collaborated less during BCI control

DEPENDENT VARIABLE [UNIT]	MTS!-BCI	MTS!-P&C	MTS!-TS
Speech [sec/min]	6.19 (2.95)	7.43 (3.69)	5.58 (2.34)
Utterances [sec/min]	2.00 (1.15)	1.27* (0.75)	1.34 (0.83)
Instrumental gestures [1/min]	0.19 (0.19)	0.30 (0.53)	0.30 (0.31)
Empathic gestures [1/min]	1.74 (1.02)	1.05* (0.96)	2.05 (1.25)
Game completion times [min]	6.95 (4.56)	5.22 (5.31)	9.27 (4.69)
Difficulty ratings	4.50 (1.00)	1.00* (3.00)	5.50 (1.75)

Table 7: Median and IQR (in parentheses underneath) values for dependent variables per game. * Result differs significantly from the corresponding MTS!-BCI result.

4.4 DISCUSSION

To explain the results we obtained, we analysed the AV data qualitatively. We tried to identify the events leading to the production of speech, utterances, and instrumental and empathic gestures. In the next subsections, we will discuss these in detail. Table 8 provides a summary of our discussion.

4.4.1 Speech

As we expected, players produced speech that was collaboration oriented. We identified three main categories of collaborative speech: instructive, consultative and awareness-creative. Participants used instructive speech when they wanted their team mates to do something “Try to stay out of the way and work them downwards so that in a minute I can take them back up” (Pair 4, MTS!-TS) or not to do something “You mustn’t do that yet” (Pair 10, MTS!-P&C). The former was often in question form for politeness “Can you send Rex down?” (Pair 2, MTS!-BCI). They used consultative speech to get the opinion of the team mate on individual or joint actions “Shall I let Max run on ahead?” (Pair 2, MTS!-TS). These were mostly in question form indicating that a response was expected and almost always they were answered by the

Instructive speech

Consultative speech

BEHAVIOUR	REASON	INTERACTION
Speech	Instruction, consultation, creating awareness, joking, encouragement	Collaborative, Emotional
Utterances	Success or failure in game	Emotional
Instrumental gestures	Pointing to location	Collaborative
Empathic gestures	Success or failure in game, encouragement	Emotional

Table 8: Social behaviour observed during the experiment. The columns show the category of social behaviour, its reason and the type of social interaction it contributes to.

*Awareness-creative
speech*

team mate “*Yes, go ahead*” (Pair 2, MTS!-TS). Awareness-creative speech aimed at drawing the team mate’s attention to game events, upcoming self-actions or difficulties encountered. To create awareness for game events, participants often ran a commentary “*There are already 3 in there*” (Pair 1, MTS!-BCI); “*Now they are just standing there staring*” (Pair 7, MTS!-P&C). They made their team mates aware of their upcoming self-actions as well: A: “*I am trying to get the sheep in the top left corner*” B: “*And I am trying to get the 3 on the right in the middle to go with me*” (Pair 2, MTS!-BCI). To create awareness for their difficulties, they did not direct their speech to their team mates. Instead, they addressed the game elements or they self-spoke. In MTS!-P&C, they addressed the sheep “*Sheep! Stick to the rules*”; “*Gotcha*” (Pair 2) while in the other games they also addressed the dogs “*No, not you*” (Pair 2, MTS!-BCI). This implies that during MTS!-P&C herding the sheep was the challenge while during the other games selecting the dogs was also a challenge. When the participants expressed their difficulties in this fashion, their team mates often responded – through speech or utterances – despite the non-directed nature of speech A: “*Hey, that’s not fair*” B: (laughter) (Pair 3, MTS!-TS).

*Non-collaborative
speech*

Participants spoke also for non-collaborative purposes. For example, they encouraged or appreciated each other’s successful actions “*Awesome*” (Pair 2, MTS!-BCI). They also made jokes “*Boy, or have you been hired in as the mole? [referring to the TV show ‘The Mole’]*” (Pair 3, MTS!-TS) and became sarcastic A: “*You need to look at the dog*” B: “*Yes, but listen, if I click randomly then at least one of the other dogs can get a turn [implying that the BCI was always selecting the same dog]*” A: “*Well, good*

luck with that [sarcastic]" (Pair 10, MTS!-BCI). Collaborative speech was more densely observed than non-collaborative speech. The aim of the game and the condition to win the prize was to finish the games as quickly as possible so collaborative speech had a more important role in this.

As we reported in section 4.3.3, participants produced a greater amount of speech during MTS!-P&C than MTS!-BCI. Our qualitative analysis showed that speech was mainly to collaborate. This suggests that participants collaborated more during less difficult control. This was not because the participants knew that their physical actions could degrade BCI performance because there was no difference between MTS!-BCI and MTS!-TS with respect to the speech produced. Our interview suggests that this was because during more difficult control, participants paid more attention to controlling the game – in particular, to selecting dogs – than collaborating A: *"The selection was also very difficult in the second game [MTS!-TS]. So we were more focussed on that than on collaborating"* B: *"When the selection did not work we tried to discuss it but ..."* A: *"... then we were more busy controlling the dogs"* (Pair 3); *"You can select them so easily [referring to MTS!-P&C] that you don't have to focus on selection and you don't lose much concentration on selecting the dogs ... so you have more time to discuss things with each other"* (Pair 9). This is supported by the subjective ratings of level of collaboration. Participants ranked MTS!-P&C as the game in which they collaborated the best. So, the challenge of selecting dogs in MTS!-BCI and MTS!-TS did not drive participants to collaborate with each other. Although collaboration would not help making correct selections, it could have compensated for the time lost in incorrect selections.

Attention split between collaboration and selection

4.4.2 Instrumental Gestures

As we showed in section 4.3.3, we found no difference between the games regarding the number of instrumental gestures produced. This seems to be a floor effect caused by the extremely low number of instrumental gestures. Although we tried to facilitate instrumental gestures, such as pointing by keeping the monitors within distance of reach, we noticed that during some games no instrumental gestures were produced at all. Regarding pointing, we noticed that participants actually compensated their gestures by collaborative speech. For example, when they wanted to specify a location, they used the game world and its elements as reference points A: *"Where is Jack?"* B: *"Top left"* (Pair 10, MTS!-BCI); *"Can you place Atlas next to Rex?"* (Pair 10, MTS!-P&C). Sometimes, when both players were busy in the same location, there was no location specification *"[Can you] Move Max closer?"* (Pair 2, MTS!-TS). Participants used pointing gestures when they wanted to

Gestures were compensated by speech

instruct their team mates “*You must place one here [pointing to team mate’s screen]*” (Pair 6, MTS!-P&C). They sometimes used their own screen to point and sometimes their team mate’s. In the former case, they often looked at their team mate while pointing to make sure that they were receiving attention.

Participants almost never produced response gestures, such as nodding and shrugging. This is not surprising because team mates were seated next to each other so they were not in direct view of each other. So, it was highly probable that their gestures would go unnoticed by their team mates. But we think that even if they were seated differently, for example oppositely, this would not change the situation because they would still be devoting more of their attention to looking at the screen than to following the gestures of their team mates. This situation is not specific to our game but can be observed in many computer games, except for ‘head-up games’ (Soute et al., 2010) which are developed with the purpose of facilitating face-to-face communication.

4.4.3 Utterances and Empathic Gestures

Participants produced utterances to indicate their emotional states. These included whistling when things were on the way (Pair 1, MTS!-P&C; Pair 2, MTS!-BCI) as well as utterances such as “Pff”, “Oops” or sighing when things were wrong (Pair 2, MTS!-P&C). Things were wrong when, for example, participants selected a wrong dog or when a sheep missed the fence. It is difficult to tell whether these participants directed utterances toward their team mates but sometimes their team mates responded to utterances with a smile or laughter. So, intentionally or not, utterances provided social interaction.

Participants produced empathic gestures along with utterances but also exclusively. As with utterances, empathic gestures conveyed emotional state information. These were sometimes directed and sometimes not. As an example of the former, when one of the participants fenced a sheep in, he started to dance. He kept on dancing until his team mate noticed him and groaned slightly disapprovingly (Pair 9, MTS!-TS). Examples for the non-directed empathic gestures were placing a hand on the forehead (Pair 2, MTS!-BCI) or head shaking (Pair 5, MTS!-TS) as if in desperation. Empathic gestures manifesting positive emotions arose when some sheep were fenced in or the game ended. These included thumbs up, clapping and smiling to team member. Some participants sighed with the relief of completing a difficult task.

Our results in section 4.3.3 revealed that participants produced a significantly smaller number of utterances and significantly fewer empathic gestures during MTS!-P&C than MTS!-BCI. No difference was observed between MTS!-BCI and MTS!-TS for the two variables. As

supported by our qualitative analysis, participants became more emotionally expressive when they used a difficult control mechanism. This does not necessarily mean that they experienced negative emotions during difficult control because, they produced utterances and empathic gestures not only when they encountered an error but also when they succeeded. In a way, difficulty of control served as an emphasiser for utterances and empathic gestures.

HIGHLIGHTS FROM THE CHAPTER

- Comparison of MTS!-BCI and MTS!-P&C showed that social interaction was influenced mainly by the challenge of control. ECE of MTS!-BCI and MTS!-TS showed that factors beyond performance, such as the novelty of using a BCI or the SSVEP stimulation, did not influence social interaction.
- Challenge of control reduced co-experience by dampening collaborative social interaction but also strengthened it by emphasising emotional social interaction.
- Collaborative social interaction was dampened during MTS!-BCI not because the participants suppressed their behaviour to prevent BCI performance from degrading. Rather, they were more concerned with remaining in control so they could not devote enough attention to collaborating, despite indicating that the game, in general, induced a high level of collaboration. So, there seems to be a trade-off between collaborating and being able to control the game.

Our first study, which we reported in the previous chapter, showed that unreliability of BCI control can influence UX and therefore mask the mechanisms playing a role in UX, such as social interaction. This supports our approach to evaluate UX of BCI games, independent of the BCI performance. In the previous study, we investigated UX of a multi-player BCI game and compared it to another controller using ECE. We did not find any significant difference between the two controllers. The purpose of the second study which we will report in this chapter was to investigate UX of a single-player BCI game in terms of immersion and affect. For this, we used MTS!-BCI and MTS!-TS as experimental games and, as in the previous study, we did an ECE to exclude the influence of performance on UX.

5.1 RELATED WORK

In §2.2 we explained the relevance of immersion and affect (or emotion in general) to UX and BCI games. Immersion has been studied in the context of some BCI applications. Friedman et al. (2007) investigated the effect of immersion on BCI control. They showed that an immersive environment can improve the sense of presence while carrying out navigational tasks through imaginary movements. Donnerer and Steed (2010) explored the influence of using various P300 stimuli on control in immersive VEs. They suggested that P300 can be used successfully in 3D environments. Some researchers evaluated presence, a concept related to immersion, in BCI controlled VEs. The experiment of Groenegrass et al. (2010) revealed that P300 based navigation lowered the sense of presence compared to gaze-based navigation. The studies we have mentioned so far considered immersion while navigating in VEs using BCI. However, immersion upon playing a game using BCI has not been studied before. This is what we report in this chapter. Evoked response BCI games, such as the MTS! game we evaluate, in which players pay attention to in-game stimuli can especially be immersive, since focussing attention is an essential ingredient of immersion.

Affect has rarely been studied before in the context of BCI applications. For example, Prasad et al. (2010) evaluated through visual analogue scales the mood of stroke patients who used a motor imagery based assistive BCI. They showed that the patients' mood improved over time, as they kept using the assistive BCI. Regarding BCI games,

Mühl et al. (2010) investigated players' states of relaxedness through physiological sensors. But they aimed at verifying whether their relaxedness induction protocol was successful. So, the game was used just as an experimental platform rather than a tool for enjoyment. In our study, we investigated the affect relating exclusively to playing a BCI game.

5.2 METHODOLOGY

In this study, we asked people to play MTSI-BCI and MTSI-TS. We did an ECE exactly as we described in §4.2.2 except that in Step 3, we evaluated immersion and affect instead of social interaction.

To evaluate the level of immersion the players experienced, we used the Immersive Experience Questionnaire (IEQ) developed by Jennett (2009). The IEQ measures game immersion in 5 dimensions, namely, cognitive involvement, real world dissociation, emotional involvement, challenge and control. It contains 31 questions with 7-point scales. Dimension scores and the total immersion score are reached by summing the corresponding individual question ratings. Higher scores indicate higher levels of immersion during the game while the median of the scale (i.e. 4) a neutral experience.

We used the Self-Assessment Manikin (SAM) (Bradley and Lang, 1994) to evaluate the affect players experienced. The SAM is a non-verbal pictorial assessment technique that directly measures the pleasure, arousal, and dominance associated with a person's affective reaction to a wide variety of stimuli. We used 9-point scales in our experiment. Higher ratings indicate more pleasure/arousal during stimulation and the median of the scale (i.e. 5) indicates a neutral experience.

We also analysed game completion time and number of selections for performance assessment. For these data, a smaller value indicates a better performance.

5.3 EXPERIMENT

In this section we will describe the setup of the experiment we conducted, the participants who took part in the experiment, the data analysis steps and the results of our data analysis.

5.3.1 Setup

A large screen projector is a semi-immersive display (Bowman et al., 2001)

Participants sat on a chair behind a table. There was only a mouse on the table. A projector that was mounted on the ceiling projected the game on a screen that was approximately 3 metres away from the participant. Stimulus properties were set in accordance with the

results of the pilot study we did for the first study (see §4.2.3). The sizes of the stimuli (3 cm on a monitor) were scaled in proportion to the increased distance to the screen. For stimulus frequencies, we used 10 Hz, 12 Hz and 15 Hz which performed comparably to the set of frequencies we used in the previous study. We used electrode locations PO_3 , O_1 , O_z , O_2 , PO_4 , P_3 , P_z and P_4 to analyse the EEG signals and used CAR for re-referencing.

The experiment consisted of two sessions. In one session participants played MTSI-BCI while in the other they played MTSI-TS. The order of the sessions were counterbalanced across the participants. In each session, participants played a familiarity trial, an easy trial and a difficult trial. The difficulty levels were simply to train the participants so not to investigate their effect on affect or immersion. After each trial, the participants filled in the SAM and were given a short break. After each session, they filled in the IEQ. At the end of the experiment, they were asked which game they would like to play again if they were given the opportunity.

In the familiarity trial participants could get used to the selection method by selecting and moving the dogs. During this trial, participants had to collect the 10 static objects placed across the playground. In the easy trial, participants had to pen a small flock of 5 sheep using the dogs. During this trial two pens were placed on the playground, one on the left and one on the right of the screen to make the task easier for the participants. In the difficult trial, participants had to gather the 10 sheep that were scattered across the playground, into one pen that was placed in the center of the playground. The layout of the playgrounds across the trials were kept the same to ensure no playground was more difficult for one of the selection methods. However to ensure that participants did not create a strategy for a specific trial, the positions of the dogs, collectible objects and sheep were altered for the different selection methods. A timeout was set for each trial for the participants to finish the level by collecting all objects or gathering all sheep into a pen. Participants had 3 minutes, 5 minutes and 10 minutes for the familiarity, easy and difficult trials respectively. Since immersion in games is often accompanied by losing track of time, the time left was not visible for the participants. Otherwise it could have influenced their perception of the elapsed time.

5.3.2 *Participants*

Seventeen people (7 female) participated in the experiment. They had an average age of 22 ($\sigma = 4.74$), ranging between 17 and 37 years. All participants except for one had normal or corrected-to-normal vision and described themselves as daily computer users. Three of them had

previous experience of BCIs. Before the experiment, all participants signed an informed consent form and they were paid according to our institution's regulations.

5.3.3 Analysis

The game version (i.e. the control mechanism) was the independent variable while the dependent variables were the IEQ and SAM ratings, game completion time and number of selections. We considered, for applicable variables, only the easy and difficult trials. So we excluded the familiarity trials. To summarise the data we computed median and IQR values. We compared MTS!-BCI and MTS!-TS in order to assess the effect of perceived control on dependent variables. We assessed the significances of differences by Wilcoxon signed-rank tests ($p < 0.05$). We also computed Spearman's ρ values to assess the correlation between the IEQ dimension scores and between the SAM dimension scores.

5.3.4 Results

Our IEQ analysis results showed that the median total immersion score for MTS!-BCI (162) was significantly higher than that of MTS!-TS (148). With respect to IEQ dimensions, the median scores for MTS!-BCI were significantly higher than those for MTS!-TS, except for the challenge (see Table 9).

IEQ DIMENSION	MTS!-TS	MTS!-BCI
Cognitive Involvement* [40]	55.0 (8.5)	61.0 (7.5)
Real World Dissociation* [24]	27.0 (9.5)	28.0 (7.0)
Emotional Involvement* [48]	53.0 (12.5)	59.0 (10.5)
Challenge [20]	22.0 (4.5)	20.0 (3.5)
Control* [32]	31.0 (6.0)	38.0 (9.5)

Table 9: Median and IQR (in parentheses) values for the IEQ dimensions per game. In brackets are the scores indicating the neutral experience for each dimension. * Significant difference between the games.

The SAM analysis results revealed that there was no difference between the two games in valence, arousal or dominance (see Table 10 for the median and IQR values).

Log analysis results indicated no difference between the two games in terms of completion time. However, both easy and difficult versions of the games were significantly different from each other in terms of the number of selections. Moreover, the difficult games lasted significantly

SAM DIMENSION	MTS!-TS		MTS!-BCI	
	EASY	DIFFICULT	EASY	DIFFICULT
Valence	7.0 (1.5)	7.0 (2.5)	7.0 (1.0)	7.0 (1.5)
Arousal	5.0 (3.0)	6.0 (4.0)	5.0 (4.0)	6.0 (4.5)
Dominance	5.0 (4.0)	6.0 (4.5)	6.0 (2.0)	6.0 (3.0)

Table 10: Median and IQR (in parentheses) values for the SAM dimension scores per game.

VARIABLE	MTS!-TS		MTS!-BCI	
	EASY	DIFFICULT	EASY	DIFFICULT
Completion time	53.28 (43.60)	113.38 (72.46)	44.47 (24.20)	105.18 (47.55)
Number of selections*	39.00 (30.50)	80.00 (54.50)	29.00 (13.00)	72.00 (42.00)

Table 11: Median and IQR (in parentheses) values for the game completion times (in seconds) and numbers of selections per game. * Significant difference between the games.

longer and took significantly more selections to complete than the easy games confirming that our difficulty manipulation was successful (see Table 11).

When participants were asked which game they would have liked to play if they were given the opportunity, 12 of them chose MTS!-BCI, 3 of them chose MTS!-TS and 2 of them chose neither.

Our analysis of correlation between the IEQ dimensions revealed significant correlations only between cognitive involvement, emotional involvement and control. The correlation coefficients and significance values can be seen in Table 12. For the SAM dimensions, only valence and dominance dimensions were consistently correlated in both MTS!-TS and MTS!-BCI (only the difficult game) (see Table 13).

5.4 DISCUSSION

Although there was no difference between MTS!-TS and MTS!-BCI in terms of affect, the immersion ratings and performance measures demonstrate that participants had a better experience of the game using BCI than using the mouse. This is probably why the majority of them indicated MTS!-BCI as the game they would like to play again.

IEQ DIMENSION	MTS!-TS	MTS!-BCI
CI-C	0.58*	0.68**
EI-C	0.76**	0.75**
CI-EI	0.79**	0.79**

Table 12: Coefficients for the significant correlations between the IEQ dimensions. CI=Cognitive Involvement, C=Control, EI=Emotional Involvement. * significant at the 0.05 level, ** significant at the 0.01 level.

DIFFICULTY	MTS!-TS	MTS!-BCI
Easy	0.72*	0.02
Difficult	0.88*	0.64*

Table 13: Correlation coefficients for the SAM valence and dominance scores. * Significant correlation ($p < 0.01$).

Regarding performance, the IEQ challenge dimension scores were not significantly different suggesting that, as we desired to do, we were able to offer equally challenging controllers. However, participants indicated in their IEQ control dimension responses that they felt more in control using BCI and they made fewer selections during BCI control. Thus, our efforts to equalise the perceived controllability of the two games was not successful. Therefore, the UX evaluations might be biased due to the differences in controllability. Indeed, for example, we identified a significant correlation between the valence and dominance dimensions of the SAM (see Table 13).

*Level of control
influencing
participant actions*

Even though the participants indicated a difference between the games in terms of level of control, they finished the games within similar times. This confirmed our finding in §4.3.3 that players compensate for the low level of control through other means, such as following different strategies. For example, in our experiment, the participants compensated for lower level of control during MTS!-TS by making a greater number of selections.

While we claimed that the level of control influenced the number of selections, an opposite relationship was also the case. While MTS!-TS provided obvious feedback about the dog that was likely to be selected¹ (i.e. the dog that is replaced by the circle appearing on the screen), MTS!-BCI provided no feedback about the dog that is likely to be selected or the likelihood of selection. For this reason, participants waited longer²

¹ Of course, the likelihood of selection was dependent on player's visuomotor skill.

² Non-significant, 0.5 seconds longer on average

(i.e. kept the mouse pressed for longer time) during MTS!-BCI than during MTS!-TS to ensure that they made a correct selection. Conversely, while playing MTS!-TS, participants released the mouse quickly with the help of the feedback and their self-confidence. However, the likelihood of selecting the correct dog in MTS!-TS improved by selection duration. As the participants waited shorter, they made more mistakes.

*Participant actions
influencing level of
control*

Personal IEQ dimensions (cognitive involvement, real world dissociation and emotional involvement) are more difficult to interpret than the game-related IEQ dimensions, which we explained in the previous paragraph (challenge and control). MTS!-BCI scored significantly higher in all personal IEQ dimensions but scores alone cannot explain why this was the case. If we had collected qualitative data, for example through interviews, then we could have had some insight about it. Nevertheless, we can speculate based on our observations and talks with the participants.

Regarding cognitive involvement, it is possible that controlling the game by thoughts – or, at least, the illusion of doing so – might have increased the scores for MTS!-BCI. Moreover, the sustained attention required to elicit SSVEP and select dogs in MTS!-BCI might have improved cognitive involvement. MTS!-TS also demands attention as the players need to release the mouse as soon as the dog they wish to select is replaced by a circle. However, the visuomotor task of releasing the mouse in time is more challenging than the attentional task of perceiving the circle because human is quicker in the latter task than in the former task. Therefore, the attentional task might become masked by the visuomotor task.

*Sustained attention
and cognitive
involvement*

Regarding emotional involvement, we tried to explain the difference between the games by the SAM scores. However, the SAM scores for the two games were not significantly different. Thus, the SAM and the IEQ emotional involvement scores did not support each other. This is not totally unexpected because the IEQ emotional involvement score only indicates whether the participants were affected or not while the SAM scores indicate the quality of the affect (e.g. valence, arousal). Therefore, we resorted to our correlation analysis results. We identified that emotional involvement was highly significantly correlated with the cognitive involvement and control scores (see Table 12). The latter correlation is also supported by the correlation between the SAM valence and dominance scores³ (see Table 13). The cause-effect relationship between the correlated IEQ dimensions is difficult to explain with the information at hand so we can only conclude that high emotional involvement is accompanied by high cognitive involvement and feeling of control.

³ A similar correlation has been reported before by Koelstra et al. (2012).

HIGHLIGHTS FROM THE CHAPTER

- MTS!-BCI provided a higher level of immersion than MTS!-TS. In both games, the IEQ dimensions cognitive involvement, emotional involvement and control were significantly correlated pairwise. The SAM valence and dominance dimensions were also significantly correlated. Thus, sense of control interacted with cognitive and emotional state.
- The perceived level of control in MTS!-TS was lower than that in MTS!-BCI but both games were completed within similar durations. Participants compensated for the lower level of control in MTS!-TS by making a greater number of selections.
- In MTS!-BCI the only task in making a selection was paying attention to stimuli while in MTS!-TS one task was paying attention to the stimuli and the other was releasing the mouse on time. Although in both games the chance of making a correct selection increased by attention duration, participants sometimes overestimated their visuomotor skills and did not wait long enough before releasing the mouse. So, they made more selection mistakes.
- Sustained attention while playing MTS!-BCI facilitated cognitive involvement in the game.
- Quantitative data is helpful to compare two phenomena but is insufficient to explain the differences and similarities. In exploratory studies, it should always be backed with qualitative data.

Our previous study indicated that people were indulgent to BCI. They accepted that it took time until BCI had correctly classified their thoughts. This, in turn, improved their sense of control. The factors that might have influenced their positive approach include the novelty of using a BCI and the idea that they were commanding things with thoughts. The purpose of the study which we will report in this chapter was to address the former factor, novelty. In this study, we did an ECE to compare the UX of BCI to that of another novel game controller, ASR. Despite having a long history, ASR is not as popular as other game controllers such as the keyboard or gamepad. Therefore, it is still a novel game controller to many players.

So far, we have evaluated in-game and post-game UX in our studies. In this study, we also considered pre-game UX. As we have discussed in §2.2.1 the UX of a product starts to be shaped even before using the product. UX is influenced by users' values, abilities, prior experiences and knowledge as well as the context of use. Every experience a user has with a product affects not only their next experience with the product but their future experience with any product using the same technology as control input. So, a product can change the conceptions or conclusions about a technology. If the change is in a positive way, then we can say that the product provides a positive UX. The key to success is to find the right interaction design which enables the technology to meet or surpass users' expectations.

Until now, we have always evaluated BCI and compared it to other control mechanisms (modalities) in a mutually exclusive fashion. That is, the players played games with single dedicated control mechanisms. Another approach, which we also took in this study, is to let players play games in which they can choose between different controllers. By analysing the factors affecting their modality choices, it might be possible to derive some information that relates to UX of each modality.

6.1 RELATED WORK

In this section, we will discuss the research on using multiple modalities for interacting with computers, known as multimodal interaction. We differentiate between two cases of multimodal interaction: complementary and redundant (Coutaz et al., 1995). In complementary multimodal interaction, people use multiple modalities simultaneously

*Complementary vs.
redundant
multimodality*

or sequentially in order to complete a task. This way, people can use the different modalities in performing the tasks that the modalities are the best for. For example, in a pen-voice system, people can use the pen to point to a place and speak to provide a command (Oviatt et al., 2000). In redundant multimodal interaction, people choose a modality from multiple others and use it exclusively in order to complete a task. This way, they can choose the modality that fits to their capabilities or needs the best. For example, a user with a broken hand bone would prefer speech over the mouse, and a user who can draw well would prefer pen input over speech. Oviatt (2000) found that if multiple modalities were available for use, then people could switch to the modality they believed to be the most accurate and efficient for conveying particular content. In this way they can improve their own and also the system's overall accuracy. However, accuracy might not always be the end goal. While a user with a pragmatic need would prefer an accurate modality, one with a hedonic need (e.g. a gamer) might prefer an enjoyable modality that is not necessarily efficient. Therefore, in our study, we considered several factors, including but not limited to performance, that might affect people's modality choices.

When multiple modalities are available, people prefer the one that they think is the 'best' for them

6.2 METHODOLOGY

In this study, we evaluated UX in three ways. The first way was to assess UX through post-hoc evaluation. We asked people to play MTS!-BCI and MTS!-ASR, which were equalised in recognition performance as we will explain in §6.2.1. Then, we collected people's opinions about the games through questionnaires as we will describe in §6.2.2. The second way was to evaluate UX with respect to user expectations. As we will describe in §6.2.3, through a questionnaire we collected people's expectations about BCI and ASR as input modalities before they played MTS!-BCI and MTS!-ASR. Then, in relation to their expectations, we assessed how successful the modalities were in meeting or surpassing people's expectations. The third way was to investigate people's modality preferences and switching behaviour as indicators of their UX. As we will describe in §6.2.4, we interviewed people after they had played MTS!-MM, which provided both BCI and ASR control, about their experience of the game and the modalities. We also inspected game logs to extract information about their modality usage.

6.2.1 ECE Details

We followed the ECE steps described in §3.3.1 as follows. In Step 1, we identified the task-related parameters that influence recognition

performance. For MTS!-BCI this was the set of stimulus frequencies while for MTS!-ASR the set of dog names to be pronounced.

In Step 2, we relied on our previous experiment results to determine the candidate parameter values for MTS!-BCI. We described this experiment and our findings in §4.2.3. To determine the candidate parameter values for MTS!-ASR, we conducted a pilot experiment. For the pilot experiment, we first identified a number of candidate dog names. These were Hector, Victor, Dexter, Pluto, Shadow and Lassie. Since we aimed at equalising the BCI and ASR performances – thus achieving an imperfect recognition performance with ASR – we tried to choose the dog names in such a way that some of them sounded alike and the others different. Next, we asked seven people to pronounce each candidate dog name 10 times while we recorded their voices. We noticed that if the microphone was located in front of the participants, the recognition was too accurate to be matched by the BCI recognition performance. Therefore, we placed the microphone to the right, behind the participants. After the experiments, for each 3-combination of dog names, we computed the average recall of the recognition performed by the ASR as explained in §3.3.2.4 (see Table 20). We had already computed the average recall for 3-combinations of various stimulation frequencies for our previous experiments (see Table 19). We paired each dog name 3-combination with each stimulus frequency 3-combination and we assessed their difference in average recall through Wilcoxon rank-sum tests. Among the least different pairs, we selected the 3-combination pair with the highest average recall. In this way, we decided to use Dexter, Lassie and Shadow (yielding an average recall of 82.86%) as dog names and 7.5 Hz, 10 Hz and 12 Hz (yielding an average recall of 83.62%) as flicker frequencies. This pair yielded a p-value of 0.97.

Finally, in step 3, we evaluated the UX of both games as we will describe in §6.2.2.

6.2.2 Questionnaires

To evaluate the UX resulting from playing MTS!-BCI and MTS!-ASR, we used three questionnaires. The first, NASA-TLX (Hart and Staveland, 1988), is a brief yet powerful questionnaire used frequently in BCI research to evaluate workload. It measures the subjective workload for a task using six items which assess mental demand, physical demand, temporal demand, performance, effort and frustration. Each item is rated using a 20-step bipolar scale resulting in a score between 0 and 100. The low and high ends of each scale are anchored with a word pair indicating the two extremes for the item (e.g. word pair perfect-failure for performance). An average or a weighted average of item scores provide the overall workload score. A higher score implies a higher

subjective workload associated with a task. In our study we used an unweighted version of the NASA-TLX.

The second questionnaire, the Game Engagement Questionnaire (GEQ, Brockmyer et al., 2009), is a validated tool developed specifically for games. It measures the subjective level of game engagement in four dimensions which are absorption, flow, presence and immersion. Nineteen items, each formed as a statement, are marked on a 5-point bipolar scale with respect to the level of agreement. Columns corresponding to the low end, middle and high end points of the scale are anchored with words No, Maybe and Yes respectively. The points are averaged over the items to reach the overall engagement score. A higher score indicates a higher level of engagement in the game.

The last questionnaire, AttrakDiff2 (Hassenzahl et al., 2003), evaluates product quality in three dimensions which are pragmatic quality, hedonic quality and attractiveness. The questionnaire contains twenty-one items rated using a 7-point semantic differential scale. For each item, the scale is anchored at extremes by opposite word pairs (e.g. simple-complicated). Ratings averaged over the items imply an overall product quality score. The higher the score, the better the subjective quality of the product. AttrakDiff2 has a reduced version (Hassenzahl and Monk, 2010) which is more convenient to fill in and analyse. In our study we used this reduced version.

In addition to the questionnaires, we extracted the number of selections and game durations from the game logs as performance indicators.

6.2.3 *Evaluating UX With Respect to Expectations*

The rationale of our method to evaluate UX with respect to user expectations is based on the SUXES (Turunen et al., 2009) method, which can be used to evaluate UX of multimodal systems. It consists of two questionnaires one of which is filled in before using a product (the expectations questionnaire) and the other after using it (the perceptions questionnaire). Both questionnaires include 9 items in statement form. The statements are related to speed, pleasantness, clearness, error free use, robustness, learning curve, naturalness, usefulness and future use. An example statement is “*Speech input is quick to use*”. In the expectations questionnaire, users mark their acceptable and desirable levels of quality on two separate 7-point scales. The range of values between the acceptable and desirable levels are identified as the zone of tolerance. In the perceptions questionnaire, users mark their perceived level of quality on a single 7-point scale. Then, the perceived level of quality is compared to the expected levels.

We started applying the SUXES method with the original questionnaires but after a couple of experiments we noticed some problems with

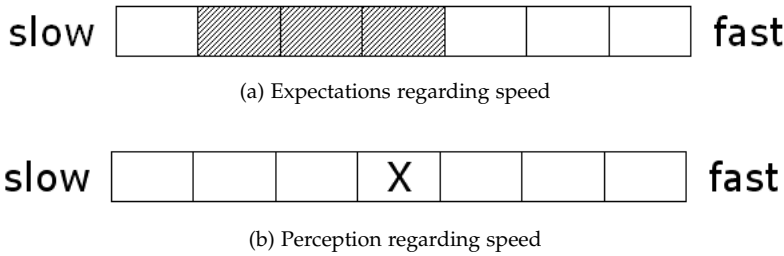


Figure 9: Interpreting expectations and perceptions: (a) implies that the user will be surprised if the interface is faster than level 4 and will be disappointed if it is slower than level 2. So the zone of expectations (ZoE) is $\langle 2,4 \rangle$ (b) indicates that the user rated the speed as level 4 which is within the ZoE thus meets the expectations.

it. Firstly, the term zone of tolerance was misleading, suggesting that any perceived level of quality falling outside the zone was intolerable. However, a perceived level higher than the desirable level is not an intolerable situation. It is actually favourable. To address this issue, we simply renamed the term zone of tolerance to zone of expectations (ZoE).

Secondly, despite our best efforts in perfecting the written instructions, the participants experienced difficulty in figuring out the 2-scale mechanism of the expectations questionnaire. Even after additional verbal explanation, some participants still filled in the questionnaire incorrectly (e.g. the desired level for an item was sometimes lower than the acceptable level). In order to address this issue, we decided to reduce the 2-scale expectations questionnaire design to a single-scale one. We used a 7-point semantic differential scale which is anchored by opposite phrase pairs at the ends (see Figure 9a). This way, participants could indicate their ZoE for each item by shading the box scale. Other than the phrase pairs, the scale contained no additional anchoring. The perceptions questionnaire was identical to the expectations questionnaire with the exception that the participants did not shade the boxes but put a cross inside the box that represented their experience (see Figure 9b). If in the perceptions questionnaire the users marked their experience for an item lower than the ZoE they would have negative UX and if they marked it higher they would have positive UX.

Thirdly, we noticed that the questionnaire items were not fully fitting for BCI systems and particularly games. For example the questionnaire items asked about the usefulness of the system while in a game usefulness is not a major concern. On the other hand the items lacked dimensions such as fatigue and fun which are relevant to BCI games. To overcome this issue, based on our experience with BCI systems and particularly with games, we chose to include the following items and corresponding

phrase pairs in parentheses (both in the given order), in our questionnaires: speed (slow–fast), pleasantness (pleasant–unpleasant), accuracy (erroneous–error-free), fatigue (tiring–effortless), learnability (easy to learn–hard to learn), naturalness (natural–unnatural) and enjoyability (boring–fun). Ordering of the phrases was consistent in expectations and perceptions questionnaires.

For easy administration, we appended the SUXES perception questionnaire items to the AttrakDiff2 questionnaire since both questionnaires contained a 7-point semantic differential scale.

The original and modified versions of the SUXES questionnaires can be found in §A.2.

6.2.4 *Modality Switching Behaviour and UX*

We prepared a semi-structured interview to be conducted at the end of the experiment to learn about users' reasons to switch or stay with a particular modality. The interview was videotaped for post-experiment analysis. The interview began with the question "*Did you switch modality during the last game?*", followed by "*Why?*" or "*Why not?*" depending on the user's answer to the first question. This way, without cueing the user, we identified the factors affecting their preference. Then we asked specifically about possible factors which might have influenced their preference: "*Was [factor] a reason for switching from or staying with one modality?*" The factors were usability, engagement, task load, performance, tiredness and curiosity. If the answer was just a Yes/No, then we asked "*Can you explain?*". Finally we repeated our first question as: "*Are there any other reasons for switching from or staying with one modality?*" in order to extract possible additional reasons that might have been triggered during the interview.

6.3 EXPERIMENT

In this section we will describe the setup of the experiment we conducted, the participants who took part in the experiment, the data analysis steps and the results of our data analysis.

6.3.1 *Setup*

Participants sat on a comfortable chair approximately 60 cm away from a 20" screen with a resolution of 1280×960 . They played MTS! three times in total. Once they played MTS!-BCI and once MTS!-ASR in counterbalanced order. Then they played MTS!-MM. They played each game until all the sheep were fenced in or the play time reached 10 minutes. Before playing MTS!-BCI and MTS!-ASR, they filled in the

SUXES expectations questionnaire. After these games, they filled in the three questionnaires, NASA-TLX, GEQ and AttrakDiff2 (containing the SUXES perceptions questionnaire), in the given order. In the NASA-TLX, the "task" was defined as "selecting a dog". For the AttrakDiff2, the "product" was replaced with "the interface for commanding the dogs" and participants were instructed to complete the questionnaire with respect to the devices they would need to use and tasks they would need to perform to select a dog. In the end of the experiment (after playing MTS!-MM), participants took part in an interview. While the participants were playing the games, the experimenter stayed outside the experiment room.

Games ran on full screen. In MTS!-ASR, BCI control was not available and brain signals were not analysed. Sound was acquired by the microphone located to the right, behind the participants. This particular location was chosen in order to match the ASR recognition performance with that of the BCI, as described §6.2.1. In MTS!-BCI game, ASR was not available and speech was not recognised. Brain signals were acquired by five EEG electrodes placed on the participant's head. While playing MTS!-ASR or MTS!-BCI participants could not switch between modalities. While playing MTS!-MM both ASR and BCI controls were available, one active at a time. The starting modality was selected randomly. The players could switch between the modalities at any time by pressing the Ctrl key. They could recognise the active modality through the game visuals as in MTS!-ASR the dogs had names while in MTS!-BCI they did not. During all games, each key press and mouse click was logged along with a timestamp. The game world layout was different in each game but comparable in difficulty.

The electrode locations and stimulus parameters were the same with those we used in our first study (see §4.2.3), also in accordance with our pilot study which we reported in §6.2.1.

6.3.2 Participants

Twenty people (3 female) participated in the experiment. They had an average age of 24.9 ($\sigma = 2.87$), ranging from 19 to 29 years, and normal or corrected vision. None of them were native English speakers. Eight of them had previous experience with BCIs and fourteen of them with ASRs. Six of them indicated that they played games more than five hours per week. Informed consent was obtained from all participants and they were paid according to the regulations of our institution.

As we explained in §6.2.3, the first 6 participants filled in the original SUXES questionnaires while the rest (14 participants) filled in the ones that we had modified. Therefore, in our SUXES analyses, we excluded these 6 participants. The remaining 14 participants had an average age

*Two sample spaces,
one contains the
other*

of 24.5 ($\sigma = 2.88$), ranging from 19 to 28 years. Two of them were female. Four of them had previous experience with BCIs and nine of them with ASRs. Four of them indicated that they played games more than five hours per week.

6.3.3 Analysis

In questionnaire and log analysis, the game version (MTS!-BCI and MTS!-ASR) was the independent variable while the dependent variables were the NASA-TLX, GEQ and AttrakDiff2 ratings, number of selections and game duration. We computed and compared the median ratings for MTS!-BCI and MTS!-ASR. We assessed the significances of differences by Wilcoxon signed-rank tests ($p < 0.05$).

For SUXES analysis, we computed two measures, measure of adequacy (MoA) and measure of superiority (MoS), for each item. MoA is the difference between the experienced level and the lower end of the ZoE while MoS is the difference between the experienced level and the higher end of the ZoE. If the experience is within the expectations, then the MoA is non-negative and MoS is non-positive. If the experience is below the acceptable level then the MoA is negative. If the experience surpasses the desirable level then the MoS is positive. In the example in Figure 9, the ZoE is $\langle 2, 4 \rangle$ and the experienced level is 4. So, MoA is 2 and MoS is 0 indicating that the experience was within the expectations. Based on median MoA and MoS values, we compared MTS!-BCI and MTS!-ASR. We assessed the significances of differences by Wilcoxon signed-rank tests ($p < 0.05$).

*Active switchers
and non-switchers*

For modality switching analysis, we grouped the participants according to their modality switching motivations based on the interview answers. We defined two groups: active switchers and non-switchers. Non-switchers were those who did not switch modality at all and those who switched only at the beginning of the game to decide on which modality to use. These people continuously used a single modality until the end of the game and fenced in the sheep using one modality. The rest of the participants were active switchers. They reported switching modality during the game after some errors or knowing that certain selections were easier using a particular modality. To study the modality switching motivations of the participants, we extracted from the interviews the factors for switching or staying with a modality. We combined these with the pre-determined factors which we explicitly asked during the interview. We computed the total number of participants reporting each factor.

To investigate whether using multiple modalities improved performance (recall our discussion in §6.1), for each of the three games played by active switchers, we calculated the number of selections (i.e. num-

ber of times the mouse button was released) and game duration as indicators of performance. We expected that the performance of active switchers in the multimodal game would be higher than that in both unimodal games, due to the switching capability and/or a learning effect. We calculated the same statistics for the multimodal game played by non-switchers as well as for the preferred and non-preferred unimodal games. For example if a non-switcher stayed at the BCI modality during the multimodal game, then their preferred unimodal game would be the one that they had played using the BCI and their non-preferred unimodal game would be the one that they had played using the ASR. We expected that in the multimodal game non-switchers would achieve a performance better than that in their non-preferred unimodal game and comparable or better (due to a potential learning effect) to that in their preferred unimodal game. We assessed the significances of differences by Wilcoxon signed-rank tests with the Bonferroni correction ($p < 0.017$).

*Preferred and
non-preferred games*

6.3.4 Results

Next, we will provide the results of the analyses we did. Since we performed a multi-method analysis, we will group our findings for each method we used.

6.3.4.1 Questionary Analysis

Table 14 shows the results of the NASA-TLX analysis. The games did not differ in terms of overall subjective workload. MTS!-BCI was more mentally demanding and marginally more effortful than MTS!-ASR.

NASA-TLX DIMENSION	MTS!-BCI	MTS!-ASR
Mental Demand*	57.50 (40.00)	32.50 (42.50)
Physical Demand	25.00 (20.00)	30.00 (31.25)
Temporal Demand	42.50 (47.05)	45.00 (33.75)
Performance	25.00 (18.75)	25.00 (32.50)
Effort†	65.00 (31.25)	57.50 (25.00)
Frustration	55.00 (52.50)	37.50 (43.75)
Overall workload	42.9 (21.45)	39.6 (25.41)

Table 14: Median and IQR (in parentheses) values for NASA-TLX dimension are overall workload scores. * Significant difference between the games. † trend toward significant difference ($p = 0.076$).

In Table 15, the results of the GEQ analysis are shown. The games did not differ in terms of the level of engagement they provided.

GEQ DIMENSION	MTS!-BCI	MTS!-ASR
Presence	2.75 (1.00)	2.50 (0.75)
Absorption	1.60 (1.15)	1.80 (0.95)
Flow	2.44 (1.25)	2.38 (0.84)
Immersion	3.50 (2.00)	4.00 (2.00)
Overall engagement	2.26 (0.86)	2.24 (0.51)

Table 15: Median and IQR (in parentheses) values for GEQ dimension are overall engagement scores.

Table 16 displays the AttrakDiff2 analysis results. The games did not differ in terms of their overall quality. However, MTS!-BCI provided a lower pragmatic quality. On the other hand, it provided marginally higher hedonic quality.

ATTRAKDIFF2 DIMENSION	MTS!-BCI	MTS!-ASR
Pragmatic Quality*	4.25 (1.30)	5.37 (1.70)
Hedonic Quality [†]	5.25 (0.90)	4.62 (0.90)
Attractiveness	4.50 (1.50)	5.00 (1.00)
Overall quality	4.65 (0.80)	5.00 (1.00)

Table 16: Median and IQR (in parentheses) values for AttrakDiff2 dimension are overall product quality scores. * Significant difference between the games. [†] trend toward significant difference ($p = 0.063$).

6.3.4.2 Log Analysis

The box plots¹ in Figure 10 display the numbers of selections and game durations corresponding to MTS!-BCI and MTS!-ASR. The numbers of selections were not different. MTS!-BCI lasted longer than MTS!-ASR.

6.3.4.3 SUXES Analysis

Table 17 shows the median MoA and MoS values. Neither metric differed significantly between the two games in any SUXES item. For

¹ A box plot is a graph of the five-number summary. A central box spans the first and third quartiles, a line in the box marks the median and lines extend from the box out to the smallest and largest observations.

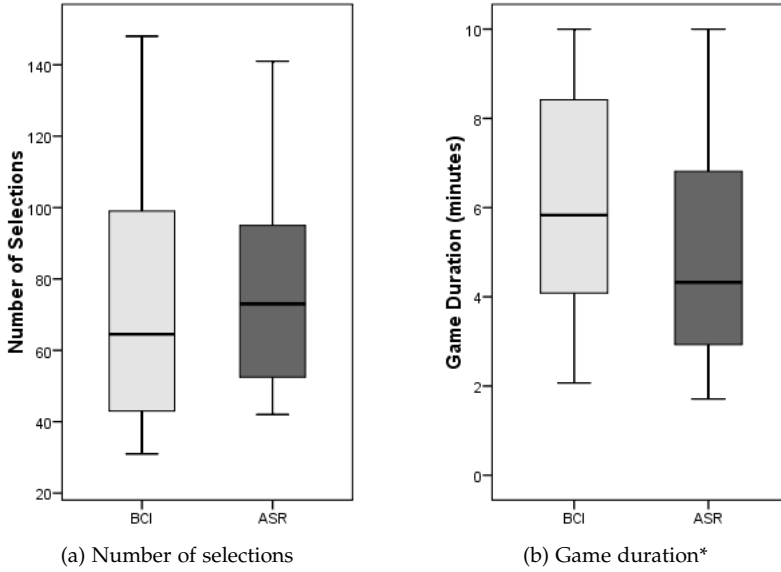


Figure 10: Numbers of selections and game durations for MTS!-BCI and MTS!-ASR. * Significant difference between the games.

all items, MoA was non-negative and MoS was non-positive. So, the perceived level for each item was within the ZoE for both games

6.3.4.4 Modality Switching Analysis

Of the 20 participants, 6 were identified as non-switchers and 14 as active switchers while playing MTS!-MM. As illustrated in Figure 11, during the interview, all of the participants indicated that performance was a factor affecting their choices. This factor was followed by usability (14 people), curiosity (12), engagement (7), task load (6) and fatigue (5). No factors different than those we had set before the experiment were identified during the interviews.

The box plots in Figure 12 display performance indicator values for the unimodal (i.e. MTS!-ASR and MTS!-BCI) and multimodal (MTS!-MM) games played by active switchers. The number of selections did not differ significantly across the games. MTS!-BCI lasted longer than MTS!-ASR and marginally longer than MTS!-MM ($p = 0.041$).

Figure 13 illustrates the performance related scores across non-switchers for the multimodal game they played as well as for their preferred and non-preferred unimodal games. While playing MTS!-MM, 4 people stayed with BCI control while 2 people preferred ASR control. The number of selections did not differ significantly across the games. MTS!-

ITEM	MTS!-BCI		MTS!-ASR	
	MOA	MOS	MOA	MOS
Speed	1.0	-1.5	1.0	-1.5
Pleasantness	1.0	-1.0	2.0	-0.5
Accuracy	0.0	-2.0	0.0	-2.0
Fatigue	1.0	-1.0	1.0	-1.0
Learnability	2.0	0.0	2.0	0.0
Naturalness	1.0	-1.0	2.0	0.0
Enjoyability	1.5	-0.5	1.5	-0.5

Table 17: Median measures of adequacy and superiority (MoA and MoS) for MTS!-BCI and MTS!-ASR.

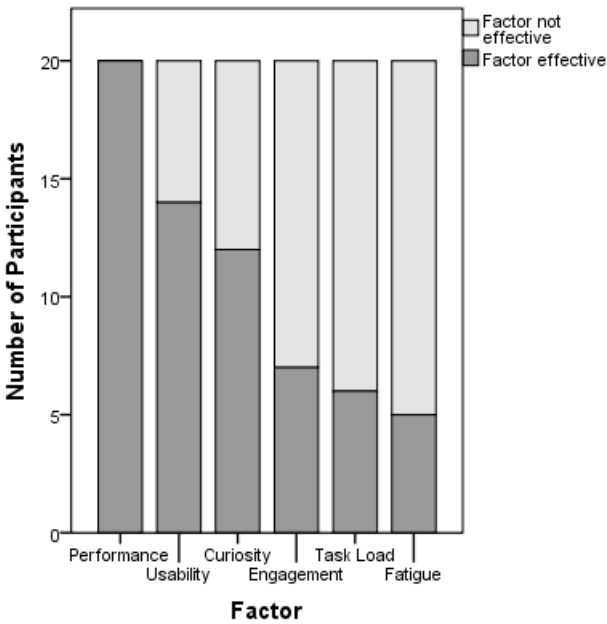


Figure 11: For each factor, number of participants reporting whether it was effective or not in modality switching or preference.

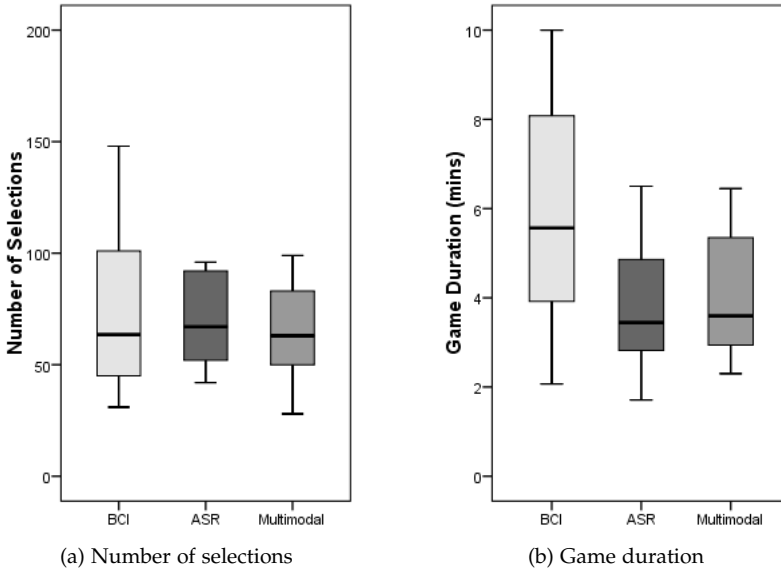


Figure 12: Performance indicators for the BCI, ASR and multimodal games played by active switchers.

MM lasted marginally shorter than non-preferred unimodal games ($p = 0.046$).

6.4 DISCUSSION

Next, we will discuss the causes and implications of our analysis results. We will divide our discussion into sections so that we can discuss each method separately.

6.4.1 Questionnaire and Log Results

We used three questionnaires NASA-TLX, GEQ and AttrakDiff2 to assess workload, engagement and product quality respectively. We analysed logs to support our data. We will now interpret the results of each questionnaire.

6.4.1.1 NASA-TLX Results

NASA-TLX analysis results showed that there was no difference between MTS!-BCI and MTS!-ASR in terms of subjective level of workload. Among the questionnaire items, the two games differed only by the mental demand. MTS!-BCI was more mentally demanding. This is not

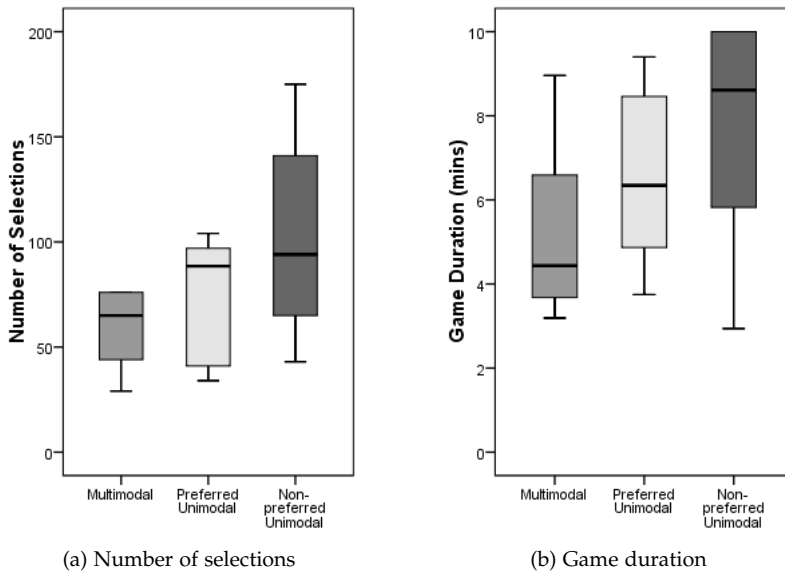


Figure 13: Performance indicators for the multimodal, preferred unimodal and non-preferred unimodal games played by non-switchers.

*The effort to satisfy
the mental demand
of BCI was not
frustrating*

an unexpected situation since MTS!-BCI requires much more mental and perceptual activity (e.g. looking, focussing at visual stimulation) compared to MTS!-ASR. Despite the high level of mental demand and marginally high level of effort MTS!-BCI induced, the games did not differ significantly in terms of the level of frustration. This implies that the participants enjoyed spending mental effort to succeed in the game.

With respect to physical and temporal demand, there were no significant differences. We understand that the physical acts of looking at the stimulation in MTS!-BCI and speaking in MTS!-ASR were equally demanding. There was no deadline for completing a task in both games so the temporal demand, which is the pace of the task, was independent of the control modality. Therefore the absence of difference in temporal demand is in line with the game logic.

The participants' perceived level of performance scores did not differ between the games confirming that our performance leveling procedure was successful. This was supported by the non-different number of selections made during both games. On the other hand, the completion times of the games were different. However, as we also noted in §5.4, the game completion time is not dependent only on the recognition performance of the modalities but also on modality specific strategies. For example, as illustrated in Figure 14, the average duration of selections was significantly longer in MTS!-BCI than that in MTS!-ASR

*Selections with BCI
took longer than
those with ASR*

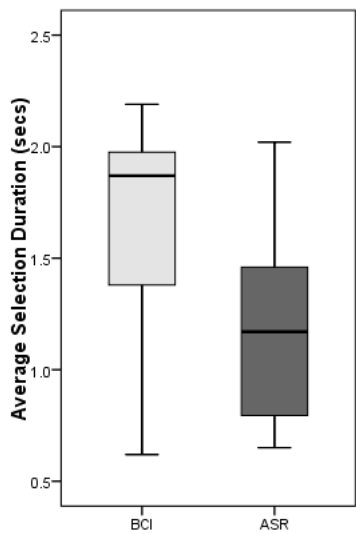


Figure 14: Average selection durations in MTS!-BCI and MTS!-ASR. The difference between the games is significant.

($p = 0.003$). In this case, while making a selection in MTS!-BCI, the game state could change more than it would in MTS!-ASR. This might necessitate recreating a strategy after some selections in the BCI game thus increase the game completion time.

6.4.1.2 GEQ Results

The GEQ scores for the two games did not differ. Since participants rated both games low on engagement, this may be because the basic components of the game – such as its visuals, rules, and so on – were not engaging and they created a floor effect. Another influential factor in low engagement scores is the tool itself we used to assess engagement. While filling in the GEQ, some participants noted several peculiarities with the questionnaire. Firstly, some of them found some items too intense. They indicated that they could not imagine themselves being ‘spaced out’ or ‘wound up’ (two example terms from the questionnaire items) while playing any game. Secondly, some participants were confused in scoring the negated items. For example, in response to the statement “I can’t tell that I’m getting tired” they could have answered with “Yes, I can’t tell that I’m getting tired” or “No, I can’t tell that I’m getting tired”, both indicating an agreement. If we had used an agreement scale instead of the Yes/Maybe/No anchors, this could have been prevented. But we prepared the questionnaire as it was validated by its developers. Thirdly, some participants stated that they did (or could) not have an opinion for some items. For example, in answering the item “If someone

*Low engagement:
game design or
methodological
issue?*

talks to me, I don't hear them" one participant said that nobody was talking during the game so this question could not be answered. Thus, it is doubtful whether the information we obtained out of the GEQ was reliable.

6.4.1.3 *AttrakDiff2 Results*

AttrakDiff2 overall quality scores of the two games were not different. Nevertheless, MTS!-ASR pragmatic quality scores were higher than those for MTS!-BCI. Participants found ASR more suitable than BCI as an input modality for completing selection tasks. This finding may seem to contradict the non-different subjective performance ratings (Table 14) and number of selections (Figure 10a) corresponding to the two games. But pragmatic quality, which we equated to usability in §2.1, is not only about the performance of an interface. It is also about consistency, flexibility, efficiency and so on. For instance, it took more effort (Table 14) and longer time (Figure 10b) for the participants to finish MTS!-BCI than to finish MTS!-ASR. This suggests that despite performing equally, BCI was not as efficient as ASR. Such factors might have influenced participants' pragmatic quality ratings.

*Low pragmatic,
high hedonic quality*

Despite being rated lower at pragmatic quality, MTS!-BCI was rated marginally higher at hedonic quality in comparison to MTS!-ASR. AttrakDiff2 hedonic quality dimension contains two sub-dimensions. The first sub-dimension, stimulation, refers to a product's ability to motivate its users to be creative with the new possibilities it offers or the challenges it poses. The other one, identification, refers to the product's ability to let its users represent themselves through the product. Both sub-dimensions are related to basic human needs; the former to personal development, the latter to social connectedness (Hassenzahl, 2004). When we analysed these two sub-dimensions, we found that MTS!-BCI was rated higher in terms of stimulation (Wilcoxon signed-rank test, $p < 0.05$). Thus, BCI control motivated the participants more than ASR control did. There was no difference between the games for identification. This was probably because the game, independent of the version, did not have a mechanism that would let the participants represent themselves to others. Thus, the controller used could not make a difference.

The games did not differ in terms of their attractiveness. Since game visuals did not differ between the games (except for the SSVEP stimulation in MTS!-BCI), the game hardware (the microphone in MTS!-ASR and the EEG system in MTS!-BCI) could have made a difference. Both the EEG electrodes (residing on participant's head) and the microphone (located behind the participant) were not in view of participants. Therefore, it might not have been possible for the participants to make a judgment about the attractiveness of the devices.

6.4.2 SUXES Results

The SUXES results showed that participants' perceptions about the games they played were within their expectations. To investigate the effect of familiarity with gaming or with the input modality on perceptions and expectations, we performed some extra analyses. We compared the participants who played games regularly to those who did not. We also compared the participants who had previous experience with BCI or ASR to those who did not. We used the Wilcoxon signed-rank test to assess the significance of any difference ($p < 0.05$). Although we hypothesised that we would find differences, especially in terms of expectations, we did not find any differences with respect to gaming habit or familiarity with the input modality.

We have identified several methodological shortcomings that might have influenced the results. Firstly, in some cases, SUXES can become overly optimistic. Let us assume the case that the users are evaluating a system for its accuracy and that the technology behind the system is intrinsically an erroneous one. Especially if the users have expertise with this technology, they would rate their expectations for the accuracy of the system low. The typical system based on this technology would function matching their expectations so users would rate their experience for the accuracy also low. At this point, our evaluation method would conclude that the system is accurate enough to satisfy users. This is not a wrong conclusion but an incomplete one. This conclusion explains where this particular system is located among its competitors using the same technology. However, it hides the information on how accurate the users found the system, independent of their expectations. This might mislead the evaluators such that they would think that there is no room, at least no need, for improving the accuracy of this system. Especially for a commercial system produced for users with a broad range of expertise levels, this might result in a serious failure on the market.

Secondly, the qualitative anchoring (i.e. the lack of quantitative labels) in the expectations questionnaire scale may direct users to form large ZoEs. The less the users are knowledgeable about the input technology, the less precise are their expectations thus the broader are their ZoEs. In this situation, the user perceptions are more likely to fall within the ZoEs. Although this is the desired case in general, in this particular situation it is misleading and uninformative.

Both shortcomings we have mentioned have to do with the characteristics of the sample space. The first issue emphasises the risk of too many experienced raters while the second the risk of too many non-experienced raters. Thus, in evaluating user expectations, the number of raters should be carefully balanced with respect to their expertise.

Importance of the sample space characteristics in evaluating expectations

6.4.3 Modality Switching Results

The participants unanimously indicated that performance was a factor in their modality preference. Non-switchers did not change modality knowing that otherwise their commands would not be understood. There were several reasons for this. For example, regarding ASR control, some suffered from their heavy foreign accent while some from their soft voice. Although non-switchers had a preference for a certain modality, their performance while using either modality did not differ significantly in terms of number of selections or game duration. This also indicated that there was no learning effect. However, the sample size for non-switchers (6 participants) may not be large enough to yield significant difference. Because, non-significantly, we did observe the performance trend we had hypothesised. The game the participants performed the best was MTS!-MM followed by their preferred unimodal game and lastly their non-preferred unimodal game. If there had been more non-switchers, the results could have been significant.

Active switchers indicated that both modalities functioned non-perfectly and that they changed modality when their commands were not understood by the game *"...when I used the speech, when the wrong one [dog] went or actually when I used the other mode, BCI, also when the wrong one went, I thought 'let's switch and try if I can make the right one go'..."* (Participant 19); *"Some of the dogs are easier to move by speech than by thought and some of those the other way so a combination of the two makes it more easier in general"* (Participant 15). Active switchers improved their game performance (in terms of game completion time) in MTS!-MM in comparison to MTS!-BCI, as we expected. However, MTS!-MM and MTS!-ASR performances were not different. We hypothesised that during MTS!-MM players used ASR more frequently than BCI. To investigate this, we extracted the number of selections per modality during MTS!-MM. As Figure 15 displays, during MTS!-MM more selections were performed using ASR than that using BCI, although the difference was not significant (Wilcoxon signed-rank test, $p > 0.05$). Nevertheless, interview results provided some clues supporting our hypothesis. Of the 14 active switchers, 7 indicated that they could not command the dog "Dexter" so switched to BCI when they wanted to control this particular dog *"One of them [the dogs] was not recognised at all, Dexter, ... with speech"* (Participant 9); *"I thought Dexter wasn't hearing me"* (Participant 3). If we assume that they used BCI to control one dog and ASR to control the remaining two, then they would be in ASR mode more frequently than in BCI mode. This can explain why the game durations were comparable in the MTS!-ASR and MTS!-MM. The non-different numbers of selections made during each game indicate that, as was the case with non-switchers, there was no learning effect. Regarding learning, the participants' comments were twofold. Some participants,

*People switched to
BCI just to control
the dog Dexter*

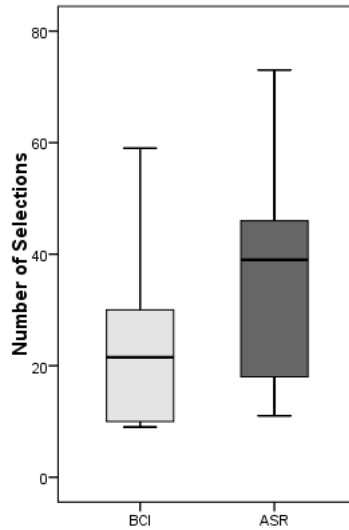


Figure 15: Number of selections made by active switchers per modality while playing MTS!-MM.

those who could not control Dexter using ASR, simply played the game with the remaining two dogs *“With speech I could not control Dexter ... at the end I didn’t use Dexter anymore”* (Participant 13). Apparently they could not learn how to control Dexter. On the other hand, some others tried until they learned how to have their speech commands understood *“Well, at first, it was because Dexter really didn’t respond to my speech commands ... but when I got Dexter moving using speech I didn’t really switch anymore”* (Participant 20); *“I also switched to operate Dexter at the start. But at the end I figured out how to operate Dexter so then I switched just for fun”* (Participant 16).

People learnt to have their speech better recognised

Besides performance, 14 people reported usability as an influencing factor. They drew attention to several differences between the modalities. Some mentioned that ASR was a faster modality *“I think the speech one is better, faster”* (Participant 1) as we have also found out through game logs (see Figure 14). Some participants explained that it was easier to see the game world while using ASR *“When you use speech it is easier to see the complete world moving and if you use thoughts you have to concentrate on the dog, you cannot see what happened with the sheep and where you have to go next”* (Participant 3); *“[With BCI] ...my eyes are more active in the game. I am selecting the target meanwhile I am selecting the dog. Instead of this I preferred speech. This way I could use both by eyes and voice in parallel”* (Participant 12). Some participants indicated that with BCI at least one dog is selected at any selection attempt while with ASR it was possible that no dog was selected (e.g. in case of silence) *“Thought input is easier*

A split-attention effect with BCI

because every time when I want a dog to move at least there is one to move. But if I use the speech input sometimes none of the dogs move" (Participant 4).

Curiosity for
(self-)performance

Twelve people indicated that curiosity was a reason affecting their modality usage. Active switchers mentioned two sorts of curiosity. The first originated when their commands were interpreted incorrectly so that they wanted to see whether the other modality would recognise the command correctly *"When I was not getting the response I wanted to see if the other one would help or not" (Participant 1)*. So this sort of curiosity was tightly coupled to performance. The second sort was curiosity about self improvement. Participants switched modality to check whether their performance improved simply by time, without any specific reason *"See if it will work better this time" (Participant 16); "Let me see if I improved. To compare my performance. Both competition with myself and competition between the two modalities" (Participant 14)*. Non-switchers also mentioned curiosity as an influential factor. They saw BCI as a novel technology and wished to play the game using it *"In daily life we have already some [speech recogniser] applications so for me, first I am trying with BCI" (Participant 18); "Flickering, that's new to me so I wanted to try that out to see how that works" (Participant 7).*

Curiosity for the
novelty

During our interviews, we encountered one unexpected reason for staying with BCI control. One of our participants indicated that he tried to use BCI more often because he knew that our ongoing research was about developing BCI games *"Also, I wanted to help you ... so that it's not always speech" (Participant 5)*. Although we did our best to hide our experimental hypotheses from our participants, our work was known to some of them. So, it is possible that other people had also behaved as this participant but no other participant explicitly stated having done so.

HIGHLIGHTS FROM THE CHAPTER

- Mental demand might not be desirable in using a pragmatic user interface but it is not necessarily the case in playing games. Participants spent effort to satisfy the mental and perceptual demands of BCI but they were not frustrated.
- Participants found SSVEP based BCI of lower pragmatic quality (less usable) than ASR for making selections in a dynamic game world. Selections took longer using BCI than ASR, yet the subjective performance ratings for the modalities did not differ. Moreover, stimulus dependent BCI control required participants to split their attention between concentrating (to make selection) and observing the game world (to keep track of game events and make plans).

- Participants found BCI more stimulating (motivating) than ASR. According to the interviews, the stimulation was mainly due to its novelty. Although the use of ASR in games is not common, the ASR technology is more familiar to people than the BCI technology is. Therefore, ASR was not a comparable modality to BCI in terms of its novelty.
- On two main occasions participants stayed with or switched to BCI control. The first was when they could not command the game using ASR. This was either a temporary (e.g. trouble with pronouncing a particular dog name) or permanent (e.g. having an accent or a voice characteristic that obstructs commanding) situation. The second was when they wanted to get more familiar with a technology that they have not interacted with before.
- Participants should not be informed about the experimental hypotheses or the research conducted in general by the investigator so that their behaviour or evaluations are not influenced. Moreover, in evaluating user expectations from a technology, it is important to ensure a balance between the users who are experienced and those who are inexperienced with the technology.

Part III

CONCLUSION

In this final part of our work, we will situate the experimental findings we showed in Part II, within the problem space we formed in Part I. We will try to answer the research question we have posed and also discuss other issues regarding the UX of BCI games. We will conclude by the possible limitations of our work and some pointers to future research directions.

IMPLICATIONS OF THE STUDIES

Before we start discussing the implications of our studies, we would like to recapitulate what we have done in each of them. We conducted three studies using different versions of our experimental game. These game versions were MTS!-BCI, MTS!-ASR, MTS!-P&C, MTS!-TS and MTS!-MM in which players made selections through BCI, ASR, pointing and clicking, visuomotor action and switching between BCI and ASR. In our first study (Chapter 4) we evaluated social interaction via observational analysis, questionnaire and interview. We compared MTS!-BCI, MTS!-P&C and MTS!-TS. In the second study (Chapter 5), we evaluated immersion, affect and performance via questionnaire and log analysis. We compared MTS!-BCI and MTS!-TS. In our last study (Chapter 6) we evaluated workload, engagement, product quality, performance, player expectations and modality usage via questionnaire, log analysis and interview. We compared MTS!-BCI, MTS!-ASR and MTS!-MM.

7.1 ANSWERING THE RESEARCH QUESTION

The collective aim of our studies was to answer the research question we posed in §3.2. To recall, we would like to understand how BCI control adds to the UX of a game. We claimed that the answer(s) to this question could reveal the good practices towards building enjoyable BCI games, some of which we mentioned in §1.2.1.

Through the UX evaluation studies we have conducted with our experimental game, MTS!, we can identify the following characteristics of BCI that contribute to a positive UX while playing a computer game: challenging control, cognitive involvement and novelty. Next we will discuss each characteristic in detail.

7.1.1 *Challenging Control*

In our first study, the participants found BCI control more difficult than mouse control. Despite this difference in difficulty, the performance of participants during the two games did not differ. This means that, while playing using BCI, the participants compensated for the difficulty of control by investing more effort into the game. This effort connected them more to the game so that their reactions upon success and failure were emphasised. They produced more utterances and empathic ges-

tures. These contributed to their emotional social interaction with their partners and, thus, to their co-experience.

Similarly, in our third study, participants indicated that they invested more effort in controlling the game using BCI than doing so with ASR. However, their performance did not differ between the two games. Moreover, they did not report different levels of frustration. Thus, the effort of controlling the game using BCI contributed positively to their UX.

In our first study, we also compared BCI to an equally-performing control mechanism (timed selection). We found out that the implications of the challenging control (on co-experience) was not specific to BCI. Timed selection also heightened the emotional social interaction between co-players. This suggests that there are more practical ways – for the game developers – than using BCI to induce challenges in a game and to enrich co-experience, such as challenging players' visuomotor skills.

In our second study, participants reported higher levels of control with BCI than timed selection, even though their actual performance did not differ between the two games. The log analyses suggested that this was because they made longer and – since the recognition accuracy improved by time – more accurate selections using BCI than using timed selection. We suggested uncertainty as the factor that could have encouraged the participants to make longer selections using BCI. With timed selection, the participants could evaluate the certainty of making a correct choice through their vision. On the other hand, with BCI, they did not have any means, such as a perception, that they could rely on to evaluate the correct choice certainty. Therefore, they figured out and, more importantly, respected – due to lack of certainty – the fact that longer selections improved the chance of making a correct choice. Although this fact also applied to timed selection, by evaluating the visual evidence, they made shorter but incorrect choices. Thus, uncertainty about mental activity drove players to be more patient while playing with BCI and thus improved their control ability.

*Improved UX
'thanks to'
uncertainty*

7.1.2 Cognitive Involvement

In our second study, we found that cognitive involvement and sense of control interacted with each other. These two, in turn, contributed to the sense of immersion. Participants indicated higher levels of cognitive involvement and control while using BCI. This is not unexpected since executing mental processes, such as attention, is the key to successful BCI control. While playing MTS! using BCI, the participants needed to pay attention to the stimulation to be able to select the right dog. This sustained mental activity not only helped them succeed in control

*Executing mental
activity improves
BCI control*

but also dissociated them from the real world. Thus, the participants experienced a higher level of immersion with BCI control.

Similarly, the results of our third study showed that BCI control demanded a higher level of mental activity than ASR control yet it was not a more frustrating experience for the participants. Thus, while for a pragmatic user interface high mental demand is not a desirable characteristic, for a game this might not necessarily be the case.

Collectively, our findings indicate that controlling a game using brain activity or at least the idea, or the feeling, of doing so can enhance the UX of the game.

7.1.3 *Novelty*

In our third study, participants rated BCI higher on hedonic quality than ASR. When we further analysed the ratings, we found that BCI stimulated (i.e. motivated) the participants more. The interviews we conducted with the participants indicated that they were curious for the technology, BCI, that was new to them. They were motivated to try it out despite its shortcomings. They indicated that as an input technology ASR was not as novel to them as BCI.

Novelty motivates

In §7.1.1 we explained that due to uncertainty participants made longer and more accurate selections. We suggest that the novelty of BCI played a role on uncertainty. Since BCI control, especially the SSVEP generation, was new to the participants, they were more patient with it. They spent time to understand how BCI worked and to learn how to interact with it. Thus, the novelty made participants indulgent to the obligatory nature of the SSVEP based BCI.

Collectively, novelty motivated participants to interact with BCI and mitigated any negative UX that could have arisen due to challenging control.

7.2 REVERSING THE RESEARCH QUESTION

The three characteristics of BCI that we have mentioned in the previous section can also hinder the UX of a game. In this section, we will discuss the disadvantages of each characteristic for UX.

7.2.1 *Challenging Control*

As we explained in §7.1.1, in our first study, BCI enriched emotional social interaction. On the other hand, it dampened collaborative social interaction. Participants had to execute multiple mental processes by attending to the visual stimulation while making collaborative decisions. So, they produced less instructive, consultative and awareness-creative

speech for the sake of remaining in control. This detracted their co-experience. Thus, we suggest that collaborative BCI games should encourage the collaboration rather than enforcing it.

The longer lasting selections with BCI that we observed in our second and third studies improved the feeling of control but also deteriorated the usability evaluations of participants. In our third study, participants rated the pragmatic quality of BCI lower than ASR. The longer lasting selections might have played a role in this. In MTS!, while the player was busy with selecting a dog, the sheep kept on moving. The longer the selections took, the more the game world changed. This made it necessary to re-construct a strategy after each selection. Thus, we suggest that SSVEP based BCIs are more usable for slow paced games.

*SSVEP based games
are for slow paced
games*

7.2.2 Cognitive Involvement

In §7.1.2 we explained that executing mental processes improved the sense of immersion. On the other hand, some mental processes, especially visual attention, can hinder UX of a game. The major feedback channel in computer games is the visual one. Thus, sharing this channel with the control mechanism can decrease the amount of game feedback the players receive and negatively influence their control capability. During the interviews we conducted within our third study, participants indicated that it was more difficult to see the game world and make plans while the SSVEP stimulation was on.

There are several ways of involving the players mentally in a game, without ceasing the feedback flow from the game. For SSVEP based games, an option is to integrate the stimuli into the game in such a way that the game world is still visually accessible for the player. This can be done by using a small number of stimuli and embodying the stimuli seamlessly into the objects in the game world. For example, Legény et al. (2011) incorporated SSVEP stimuli into animated butterfly wings – which also flicker in nature – and placed the butterflies in a forest-like VE. Another way is to consider covert attention in SSVEP detection. Although previous studies reported overt attention to be more reliable than covert attention for SSVEP based BCIs (Walter et al., 2012; Kelly et al., 2004), a combination of the two may improve the overall performance. SSVEP is, of course, not the only way of cognitively involving the players. For example, using other non-visually evoked potentials, such as the steady-state auditorily evoked potential (SSAEP) (Picton et al., 2003), would separate the stimulation and feedback sensory channels. Alternatively, neuromechanisms such as movement imagery, which do not require any stimulation, would allow players to fully appreciate the game visuals.

*Mitigating the load
on the visual
channel*

7.2.3 *Novelty*

In §2.2.3.2 we explained that the novelty of a product can make the users overly optimistic and bias their evaluations. If the novelty masks the problems with the product, then when it vanishes (i.e. after the product is used a number of times) then the problems will start to emerge and the UX will deteriorate. Ensuring that the UX of a product does not deteriorate over time is crucial for the acceptance of the product by the public and for its success on the market in the long run.

Besides its positive influence on UX (see §7.1.3), the novelty of BCI poses a potential threat to its long term UX. It may not be right to make immediate conclusions about the quality of a BCI application based on the positive UX ratings resulting from one-time use. For example, in our third study, many participants indicated that interacting with BCI was new to them, compared to using an ASR. So, their behaviour and actions were mainly motivated by their curiosity for a novel interaction paradigm. It is probable that their evaluations were also influenced by this novelty. Thus, it is questionable whether they would like to play MTS!-BCI or any other game using BCI for the second time or longer.

Novelty deceives

7.3 HOW TO EVALUATE THE UX OF BCI GAMES?

Our studies have shown us some good and bad practices in evaluating UX of BCI games or games in general. These were related to data collection methods, evaluated concepts, participants (i.e. evaluators) and comparing BCI to other modalities.

7.3.1 *Data Collection Methods*

In all of our studies we used several data collection methods together to assess UX. Using multiple methods provided at least two main advantages. Firstly, complementary data collected by different methods helped us explain the results we obtained with one method by those obtained through the other method. For example, in our third study, we analysed the game logs to profile participants' modality usage. If we had not supported our analysis with the interview responses, we would not be able to understand why people used certain modalities or when and why they switched modality. Secondly, redundant data obtained through different methods let us cross-check the analysis results for the different data. For instance, in our second study, we confirmed the IEQ control dimension responses by the number of selections.

Multimodal UX evaluation

In our first study, we did quantitative and qualitative observation analysis on the AV data. Quantitative analysis allowed us to summarise and compare participant behaviour while playing different games. However,

The importance of qualitative data for exploratory studies

numbers alone fell short to explain the reasons behind certain behaviour. Qualitative analysis allowed us to relate participant behaviour to game events thus to understand the events that triggered certain behaviour. For example, by analysing the content of participants' speech, we figured out that they compensated for instrumental gestures by speaking and therefore the amount of speech was low. The benefit of qualitative analysis for quantitative analysis is not limited to observation analysis but can be observed across different analysis methods. For example, in our second study, relying on the IEQ responses, we found that emotional involvement was higher while playing MTS!-BCI than that while playing MTS!-TS but we could not explain the reason behind this with full confidence. If we had, for instance, interviewed the participants then we could have had insight about their emotional state during the game. Therefore, quantitative analysis is suitable for comparison studies but qualitative analysis is a must for exploratory studies.

The interviews we conducted allowed us to obtain information that we would have otherwise not been able to reach through observation. For example, in our first study, through the interview responses, we could determine that the participants collaborated less while playing MTS!-BCI due to the challenging control but not, for example, for the sake of preventing movement related noise in the EEG data. One might claim that it is possible to reach such information through any subjective data collection method, such as a questionnaire. Questionnaires are static media that rely on pre-determined theories or hypotheses. However, UX while playing a game may be influenced by factors so many in number that they cannot be represented in a questionnaire. Thus, questionnaires are not as powerful as interviews in capturing the vast range of factors playing a role in UX.

Influencing UX while evaluating it

A critical point of concern in UX evaluation is to minimise the effort that the participant spends in evaluating UX. Especially if the evaluations are conducted in between multiple sessions and take too much time or effort then this might influence their upcoming UX. Moreover, the boredom or the cognitive load might negatively bias the quality of evaluations. One way of minimising the participant effort is to reduce the complexity of the evaluation methods. This means avoiding unnecessary questions during an interview or using questionnaires that are as brief – but still powerful – as possible. For instance, in our third study, we used the reduced AttrakDiff2 for pragmatic quality evaluation. Another way to reduce participant effort is to use passive data collection methods, such as psychophysiological measurement or AV recording. This way, participants do not spend any effort in providing UX related data. They are perhaps not even aware that they are doing so.

The conclusions we have made so far about the data collection methods were rather generic. There are also points of concern specific to evaluation of BCI games. One of the most prominent points is with

observation analysis. Since BCI is a private communication channel, one cannot obtain a ground truth about player intention simply by observing. It is not possible to relate game events or player behaviour to player intention. This makes it difficult to use observation analysis to evaluate the UX of BCI games. In our first study, we were able to use observation analysis but this was just because the participants ran commentaries to interact with their team mates. This way we could infer their intentions and relate these to game events. In a single-player game, this would be less likely to happen. Due to lack of ground truth for player intention, subjective methods such as questionnaires and interviews are very helpful in the evaluation of BCI games.

*Observation
analysis during
private control*

7.3.2 *UX Related Concepts*

The choice of concepts to consider while evaluating UX of BCI games is crucial. The characteristics of BCI that influence the UX of a game, such as challenging control or cognitive involvement (see §7.1), are far different from, and actually against, the desirable qualities of other control modalities. For example, mental demand may not be a desirable characteristic of a word processing software but, as we found in our third study, it did not have a negative influence on UX of our BCI game. Therefore, care should be taken to interpret the results obtained through evaluation tools that are not validated specifically on BCI hardware or software.

7.3.3 *Participant Selection and Treatment*

In our third study, regarding the UX evaluation with respect to user expectations, we explained the importance of ensuring a population of participants (i.e. UX evaluators) that is balanced with respect to the participants' familiarity with BCI. We suggested that too many inexperienced or too many experienced users might bias UX evaluations and deceive the investigators. Therefore, in the evaluation of the UX of BCI games, it is viable to collect data regarding the participants' familiarity with BCI applications or games before the experiments and either ensure a balance between the groups before the experiments or check for any differences between the groups after the experiments.

Participants' behaviour during the experiment might be influenced by their knowledge or suppositions about the experimental hypotheses (Jones et al., 2003). In our third study, we found out that one of the participants tried to use BCI more frequently partly because they knew that the investigators were conducting research into developing better BCI games. To prevent such biased behaviour, investigators should refrain from experimenting with people who know about the experiment

*Avoiding 'The Man
Who Knew Too
Much'*

or the general ongoing research. Moreover, the calls for experiment participation should not form a, positive or negative, prejudice in potential participants. Care should be taken not to make any subjective comments about the BCI (or any other technology) while making the call. For example, a call such as "Come! Play this cool game using only your brain waves! Yes, that's true, only brain waves!" is sure to form a positive pre-game UX that is destined to bias the experiment. To set a more appropriate example, we show in §A.1 the text we used to call our participants to our third experiment.

7.3.4 *Comparing BCI to Other Modalities*

UX(BCI) versus UX(game)

Evaluating UX of BCI control in comparison to other controllers enabled us to understand whether the positive or negative UX arose due to the BCI or the game. For example, in our second study, MTS!-BCI provided a high level of immersion to the players. This immersion could have been provided by the game, the controller or a combination of the two. By comparing MTS!-BCI to MTS!-TS we found that it was the BCI that contributed to the immersion.

While comparing UX of BCI to other controllers, unless it is an exploratory study such as the ones we conducted within this work, it is important to ensure that the comparison is a fair one. For example, we saw in our third study that the novelty of BCI, which is a transient property, played a role in participants' behaviour and evaluations. So, when evaluating UX of a BCI game, it is important to investigate whether the novelty led to a halo effect and masked the actual UX.

HIGHLIGHTS FROM THE CHAPTER

- Challenging BCI control can emphasize emotional interaction and enhance co-experience. Moreover, the effort it takes to control the BCI and uncertainty about mental activity can improve the UX. On the other hand, challenging control can dampen collaborative interaction and hamper co-experience. Furthermore, it can reduce usability.
- Cognitive involvement facilitated by BCI can help in providing immersion and can improve UX. Conversely, the visual stimulation in an evoked potential game can prevent feedback flow from the game to the player and deteriorate UX.
- The novelty of using a BCI can motivate people to interact with the game. However, it can adversely influence their post-game UX when the novelty vanishes.

- Using multiple data collection methods can improve the quality of UX evaluations. Quantitative methods are suitable for comparison purposes but qualitative methods are necessary for exploratory studies. Questionnaires are practical for comparing phenomena but they are not as powerful as interviews in capturing the vast range of factors that play a role in UX.
- Since BCI is a private (i.e. not observable) communication channel, subjective data collection methods, such as questionnaires or interviews, are much more suitable than objective methods, such as observation or log analysis, to relate player intention to game events.
- The effort and time the evaluators spend in performing the evaluation can adversely affect their UX and evaluations.
- The factors that can improve the UX of BCI may not necessarily improve the UX of another modality.
- To understand for sure that BCI plays a positive role in UX of a game, it should be compared to, ideally, the same game played with a different controller.

LIMITATIONS OF THE THESIS AND FUTURE WORK

Doing research is so fruitful that there is no end to it. Besides answering questions and enlightening us, humanity in general, doing research enables us to ask new questions and conduct new research that let the enlightenment go on. To keep on conducting research, we, the researchers, feed from our uncertainty and even incomprehension. We are motivated by the mistakes we make or the unexpectedness we are faced with despite our meticulous works. In this chapter, we will discuss how the limitations of our studies bear new questions and point at future research directions.

8.1 THE GAME AND BCI HARDWARE

We conducted all our studies using a single game, MTS!, which we developed exclusively for our work. To be able to conduct ECE, we gave precedence to the flexibility of the game to support various input modalities. This necessitated a rather simple game that could be commanded in the same fashion using the different modalities. The simplicity of MTS! allowed us to conduct controlled experiments but it might have also influenced the UX negatively. In §1.2.1.2 we discussed that while playing games players would like to do the things that they could not have done in real life. MTS! did not have a mechanism, such as a story or an avatar, that could mediate the transformation of the player from the real world to the game world. So, in the context of the player motivations we hypothesised in §1.2.1, although MTS! succeeded in providing a challenging and social environment, it lacked in providing a fantastic environment. Furthermore, MTS! did not have rich visuals, which was a deliberate choice so as not to distract the player while attending to the stimulation. On the other hand, visuals could have motivated people to play the game and mediated immersion. It is worthwhile to investigate the UX of BCI control while playing, for example, a commercial game that is rich in narrative and visuals. However, we would like to note that such a game with rich elements can also create a ceiling effect on the UX that the controller (e.g. BCI) does not matter anymore.

*UX of fantasy
providing BCI
games?*

In the BCI version of MTS!, MTS!-BCI, we chose to use a single neuromechanism, the SSVEP. SSVEP facilitated making selections in the game, as the majority of the SSVEP games or other applications do. The difference is that instead of selecting directions by attending static stimuli, players select dogs by attending movable stimuli. Nevertheless,

*Some findings were
specific to
neuromechanism*

MTS!-BCI can be regarded as a classical SSVEP game. The consequences of this were twofold. On one hand, we had the opportunity to evaluate the ongoing trend in using SSVEP based BCI applications and show the good and bad practices to guide developers in designing SSVEP games in the future. On the other hand, our studies might have been hindered by the disadvantages of the SSVEP. Some of our findings, such as the ones we described in §7.2.2, were related specifically to SSVEP and might not have shown up if we had used another neuromechanism. Thus, these findings cannot necessarily be generalised to all BCI games. Thus, future research should keep on evaluating UX of BCI games that make use of different neuromechanisms so that a common ground can be reached for all BCI games.

*Influence of EEG
device on UX*

The professional EEG equipment that we used in our studies required us to apply some electrolyte gel on each participant's scalp and required participants to wear an electrode cap with a number of wires hanging along their shoulders. As we discussed in §2.2.4 UX of a game can be influenced by the controller that the players play with. The preparation and cleaning up procedures for professional EEG equipment do not only take time but also some effort for the participant. So, the hardware we used might have, consciously or not, influenced participants' evaluations, such as their usability assessments in our third study. Several companies have released EEG headsets (Liao et al., 2012) that are quicker to deploy, look more attractive and do not require application and removal of electrolyte gel. However, there seems to be a trade off between high-quality signal acquisition and easy and fast deployment. For example Duvinage et al. (2012) found significant differences in P300 recognition performance between a commercial and a professional EEG device. In all of our studies, to be able to conduct ECE, we tried to achieve a BCI recognition performance that was high enough to reach to the performance of other controllers. This is why we opted for using a professional EEG device. Although we did not receive any comments against the hardware, we cannot neglect the potential influence of hardware on UX. Therefore, future research should investigate the influence of BCI hardware on the UX.

8.2 UX EVALUATION

We proposed the ECE approach to compare the UX evaluations for multiple controllers independent of their recognition performances. In doing this, ensuring the equalised performances of the controllers is crucial. In our second study, we could not equalise the recognition performances in MTS!-BCI and MTS!-TS. This was because there was actually no recogniser in MTS!-TS. The selection accuracy was completely dependent on the player's visuomotor skill. MTS-BCI also required

some player skills but part of the selection accuracy was dependent on the recogniser (i.e. BCI) performance. We partly relied on the literature to equate MTS!-TS correct selection chance to that in MTS!-BCI. Despite the theoretically equalised selection chances, in practice participants acted in a way that we could not foresee. They were less tolerant thus less patient and thus less successful while playing MTS!-TS. If we had conducted a pilot study instead of relying on the literature to estimate player visuomotor skills, we might have identified this user behaviour. Therefore, in the context of ECE, it is preferable to conduct pilot experiments to obtain an understanding of user behaviour.

In §2.3 we mentioned several data collection methods suitable for UX evaluation. As each method offered certain advantages, we made use of several of them throughout our studies to assess their usefulness. While we used qualitative and quantitative methods together in our first and third studies, in our second study we only used the latter method. In the absence of qualitative data, we could not explain, for example, why the participants experienced different levels of emotional involvement while playing different games. We also could not identify the cause-effect relationship between correlated measures. Therefore, future exploratory UX studies should make use of qualitative data.

Although we tried to include as many different methods as possible, we did have to exclude some. For example, we did not use psychophysiological measurements – which could have enhanced especially the in-game UX evaluations to a great extent – for several reasons in addition to those we have already mentioned in §2.3.1. One reason was the immaturity of the research in psychophysiological UX assessment. Although some studies suggested correlations between physiological measurements and UX related concepts (Mandryk et al., 2006b; Nacke et al., 2010b), no consensus has been formed so far. Therefore, in our evaluations we used the more conventional methods. Another reason against using psychophysiological measurements was the potential disturbance of the additional physiological sensors on the participants, which could have influenced their UX. Even EEG based UX evaluation could have been problematic because while in some of our studies we used only a few localised EEG electrodes to reduce the preparation time and disturbance to the participant, UX evaluation would ideally need a greater number of electrodes. Thus, in our evaluations, we tried to use methods that would influence the UX as little as possible. Nevertheless, as the research in psychophysiological UX evaluation keeps on, future studies should make use of physiological sensors in combination with other methods to evaluate in-game UX.

*Why not
psychophysiological
measurements?*

We mentioned in §2.2 several UX related concepts that could capture UX from different viewpoints. In our studies, we evaluated various concepts that represented pre-game, in-game and post-game UX. But, as with the UX evaluation methods, we could consider only a subset of

UX related concepts. For example, we did not evaluate presence, which is highly relevant to fantasy fulfillment (see §1.2.1.2). Actually, due to this very reason we have not included presence in our evaluations. As we discussed in §8.1, our game did not have a mechanism for fantasy fulfillment. So, presence evaluations might not have been useful. But fantasy providing BCI games should definitely be evaluated for presence. Another concept that we left out was the long-term UX. The reason for this was the amount of time it required. However, especially our findings related to the novelty of BCI (discussed in §7.2.3) showed that long-term UX is essential in conducting realistic evaluations.

*Inter- vs.
intra-coder
agreement*

During our discussion in §2.3.2 we stressed the importance of inter-coder agreement while analysing qualitative data. In our first study, we manually coded player vocalisations and actions by observing the AV recordings. We did not report inter-coder agreement because the coding was performed by a single coder. One reason was the lack of manpower to re-code the recordings. Another reason was the comparative nature of our analysis. In comparative analysis, as long as a coder is consistent in coding, the occurrence of disputable codings would be equally likely for all the conditions under comparison. So, rather than inter-coder agreement, intra-coder agreement (i.e. self consistency) is essential. Nevertheless, computing inter-coder agreement may prevent contamination in codings and improve coding reliability. We suggest that both inter- and intra-annotator agreement should be computed while analysing qualitative data.

8.3 PARTICIPANTS

In each of our studies, we did experiments with around 20 participants. When we consider experiments in behavioural sciences, such a sample size might not be large enough to yield statistical significance in analyses. For us, the main reason for keeping the sample size small was the amount of time required to conduct experiments with professional EEG devices. Unlike usual experiments with the other controllers, BCI experimentation requires effortful and time consuming preparation of each participant by the experimenter. According to our experience, the time that goes into the preparation of a participant is at least 15 minutes for a 5-electrode montage and this number increases with the number of electrodes and other sensors used. Added to this is the time required for the cleanup procedure (cleaning and drying the head cap and electrodes) after the experiment, which could take another 15 minutes according to the number of electrodes used. Although the sample sizes in our experiments were not large, we did identify significances in our analyses. Only when we categorised the sample space, as we did in §6.3.4.4 for modality switching behaviour, did we end up with

sample sizes not large enough to yield statistical significance. Future BCI studies should try to include as many participants as possible in experiments by using more practical EEG devices or by involving multiple experimenters.

For each of our experiments, we tried to spread out the call for participation as far as possible so as to obtain a sample space that was as diverse as possible. However, since we started spreading out the calls from the university (due to its proximity), we could not avoid having sample spaces consisting of mainly young students. We did have participants from outside the university, indicating that our calls did reach outside the university. However, even in those cases, the participants turned out to be young people, who were more enthusiastic about technology or games. Therefore, the UX evaluations might have been influenced by the extra motivation the participants had. We suggest that future studies should try to obtain a more diverse sample space to achieve trustworthy UX evaluations.

BCI games require investment of time (to learn to control) and money (to purchase the hardware) to play. This is not uncommon as similar requirements also apply to the games played by other novel controllers such as the Wii or the Kinect. Even non-core players are interested in buying such games. However, the acceptance level of BCI games has not yet reached to the level of other novel controllers so we foresee that the potential *first* users of BCI games would be the hardcore gamers, rather than the casual gamers. In this case, it might be more meaningful to conduct BCI game experiments with hardcore gamers. In each of our three studies, we did have participants who indicated playing games casually (e.g. more than 5 hours per week) but they were not hardcore gamers who, for example, would play for longer times to train themselves and compete against other players. Future studies should try to involve more hardcore gamers to obtain a realistic assessment of UX.

*Importance of
hardcore gamers to
BCI game
evaluation*

HIGHLIGHTS FROM THE CHAPTER

- The results of the studies we conducted on one challenge based, evoked response BCI game played through a professional EEG device may not necessarily be generalisable to all genres of BCI games. Future studies should evaluate the UX of BCI games which satisfy different player motivations and/or depend on other neuromechanisms and/or are played using other EEG equipment.
- The data collection methods and UX related concepts we used do not necessarily constitute an optimal list for UX evaluation. Future studies should back quantitative data with qualitative data and consider using physiological measurements to objectively assess

UX. Moreover, they should consider long-term UX to obtain a realistic UX evaluation.

- The sample size (i.e. the number of participants) we had for data analysis might not have been large enough to yield statistical significance. Future studies should try to improve this number by involving multiple experimenters or using more practical BCI hardware. Furthermore, hardcore gamers should take part in the experiments as the first potential users of BCI games.

Part IV

APPENDIX

In the appendix, we will provide experimental material that has not been shown before within or outside this thesis (such as the modified questionnaires) followed by long tables and author publications.

EXPERIMENTAL MATERIAL

A.1 CALL FOR PARTICIPATION IN THE THIRD STUDY

We are searching for participants who are willing to play a thought and speech controlled game and share their experiences by filling in some questionnaires. The entire experiment will last for 1-1,5 hrs depending on your performance to complete the games. During the experiment your brain activity will be measured by the 7 EEG sensors placed on your scalp.

The experiments will take place at the University of Twente, SmartXp Lab, A124 (MMC).

Non-UT-workers will be compensated for their participation at the rate of # Eur/hr. UT-students enrolled to the course 201000076: User Studies in Human Media Interaction (2010) will choose between # Eur/hr and course credits.

The experiments will start on 1 March and continue until 18 March. If you are willing to participate, please e-mail your intent to H. Gürkök (xxx@cs.utwente.nl) so that you can be contacted back for the schedule and further instructions.

Please feel free to distribute this call to anybody who might be interested to participate in this experiment.

A.2 SUXES QUESTIONNAIRES USED IN THE THIRD STUDY

As we explained in §6.2.3, in our third study we modified the SUXES questionnaires. Next, we will provide the original and modified versions of the questionnaires. We will only provide the questionnaires we used for evaluating MTS!-BCI since they are identical to the ones we used for evaluating MTS!-ASR except that the former asked for the 'thought input' while the latter for the 'speech input'.

A.2.1 *Original Expectations Questionary*

In the next game, you will play using thought input (i.e. using a brain-computer interface). Based on your experiences or knowledge about brain-computer interfaces, please circle below the acceptable and desirable levels for yourself about the statements related to using thought input. 1 means the lowest level while 7 means the highest level.

Statement	Acceptable Level (lower than this level would not be acceptable)							Desirable Level (I don't expect higher level than this)						
	low..						..high	low..						..high
Using thought input is fast	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Using thought input is pleasant	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Using thought input is clear	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Using thought input is error-free	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Thought input functions free of errors	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Using thought input is easy to learn	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Using thought input is natural	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Thought input is useful	1	2	3	4	5	6	7	1	2	3	4	5	6	7
I would use thought input in the future	1	2	3	4	5	6	7	1	2	3	4	5	6	7

A.2.2 *Original Perceptions Questionary*

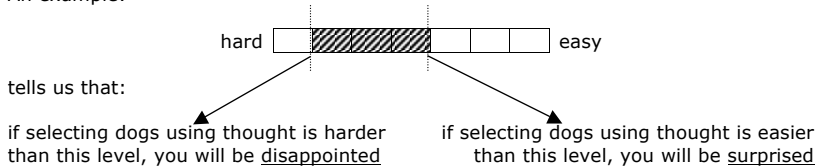
You have just played a game using thought input. Based on your experience, please circle the level you believe the thought input has the feature described by the statement. Once again, 1 means the lowest level while 7 means the highest level.

Statement	Experienced Level						
	low..						..high
Using thought input is fast	1	2	3	4	5	6	7
Using thought input is pleasant	1	2	3	4	5	6	7
Using thought input is clear	1	2	3	4	5	6	7
Using thought input is error-free	1	2	3	4	5	6	7
Thought input functions free of errors	1	2	3	4	5	6	7
Using thought input is easy to learn	1	2	3	4	5	6	7
Using thought input is natural	1	2	3	4	5	6	7
Thought input is useful	1	2	3	4	5	6	7
I would use thought input in the future	1	2	3	4	5	6	7

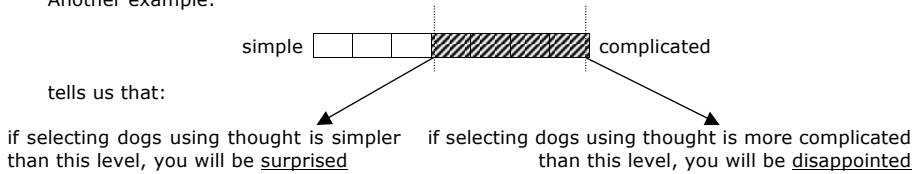
A.2.3 Modified Expectations Questionnaire

As already explained to you, in the next game, you will use thought input to select the dogs. Based on your current experience or knowledge about brain-computer interfaces, please shade the boxes below to indicate your zone of tolerance for "selecting dogs using thought" using the word pairs.

An example:

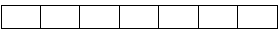
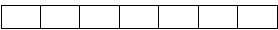
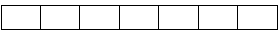
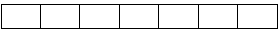
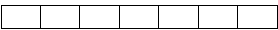


Another example:



You can shade any number of boxes (between 1 and 7) as you wish. Think about the devices you need to use and tasks you need to do to select a dog. Keep in mind that there is no right or wrong answer. Your personal opinion is what counts!

Please shade your zone of expectations for the following aspects of selecting dogs using thought:

slow		fast
pleasant		unpleasant
erroneous		error-free
tiring		effortless
easy to learn		hard to learn
natural		unnatural
boring		fun

TABLES

RECORD #	ARTICLE DOI
1*	10.1109/IEMBS.2011.6091554
2	10.1145/2070481.2070491
3	10.1145/2071423.2071490
4	10.1145/2073370.2073385
5	10.1134/S0362119711070048
6	10.1007/978-3-642-24500-8_9
7	10.1109/TSMCA.2011.2116000
8	10.1016/j.neuropsychologia.2011.09.043
9	10.1016/j.neulet.2011.08.023
10*	10.1016/j.neuroimage.2011.06.060
11	10.1177/1046878110378140
12	10.1111/j.1528-1167.2011.03252.x
13*	10.1109/CGAMES.2011.6000344
14*	10.1007/978-3-642-24091-1_60
15	10.1007/978-3-642-23774-4_12
16*	10.1109/ISCE.2011.5973847
17	(not available)
18*	10.1109/ICIEA.2011.5975562
19*	10.1109/VS-GAMES.2011.27
20	10.1111/j.1469-8986.2011.01189.x
21*	10.1109/CCMB.2011.5952128
22*	10.4156/aiss.vol3.issue7.34
23	10.1109/NFSI.2011.5936807
24	10.1016/j.brainres.2011.05.052
25	10.1007/978-3-642-21663-3_64
26*	(not available)
27	10.1016/j.dcn.2011.02.003
28	10.1007/978-3-642-21869-9_65
29	10.1007/978-3-642-21869-9_8

30*	10.1145/1978942.1978947
31*	10.1145/1978942.1978994
32	10.1145/1979742.1979828
33*	10.1007/978-3-642-21501-8_45
34*	10.1007/978-3-642-21501-8_48
35*	10.1109/TNSRE.2011.2121919
36	10.1108/17549451111149278
37	10.1007/s00221-011-2658-3
38	10.1109/CSPA.2011.5759914
39	10.1109/ISSCC.2011.5746204
40	10.1007/s00371-011-0551-5
41	10.1080/17470919.2010.506139
42	10.1016/j.dcn.2011.01.001
43	10.3233/978-1-60750-706-2-606
44	10.1016/j.jesp.2010.11.014
45*	10.1007/978-3-642-17829-0_12
46*	10.1080/10447318.2011.535755
47	10.1109/CISP.2010.5648081
48*	10.1145/1864431.1864450
49*	10.1186/1743-0003-7-60
50	10.1371/journal.pone.0014187
51	(not available)
52	10.1097/WNR.ob013e32834065f5
53	10.1109/CW.2010.56
54*	10.1109/ICEGIC.2010.5716901
55	10.1109/CW.2010.22
56	10.3791/2320
57	10.1109/IEMBS.2010.5626708
58	10.1109/ICSSENS.2010.5690277
59	(not available)
60*	10.1109/IEMBS.2010.5627376
61	(not available)
62	10.1109/IEMBS.2010.5626555
63	(not available)
64	10.1145/1920778.1920801
65	(not available)

66	(not available)
67	10.3389/fnhum.2010.00235
68*	10.1109/ICMA.2010.5589104
69	10.1109/ICAL.2010.5585311
70	10.3844/jcssp.2010.706.715
71	10.1109/ICALT.2010.81
72	10.1109/ICALT.2010.143
73	10.1109/ESIAT.2010.5568293
74*	10.4018/jgcms.2010100103
75	10.1016/j.ecolecon.2010.07.032
76*	10.1109/ICINDMA.2010.5538051
77	10.1007/s10548-010-0147-9
78	10.1016/j.intell.2010.06.001
79	10.1007/978-3-642-13437-1_50
80*	10.1016/j.clinph.2010.02.157
81	10.1016/j.neulet.2010.06.051
82	10.1016/j.yebbeh.2010.04.011
83	10.1016/j.apergo.2008.08.005
84	10.1016/j.neuropsychologia.2010.03.012
85	(not available)
86	10.1177/0956797609360750
87	(not available)
88	10.1007/s11517-009-0572-7
89*	10.1007/s12193-010-0046-0
90*	(not available)
91	10.1162/pres.19.1.1
92	10.1080/17470910903202666
93	10.1111/j.1469-8986.2009.00925.x
94	10.1109/LEOS.2009.5343270
95	10.3233/978-1-60750-017-9-189
96	10.3233/978-1-60750-017-9-37
97*	10.1109/APCIP.2009.111
98	10.1145/1690388.1690452
99	10.1109/IEMBS.2009.5332722
100	(not available)
101	10.1109/BCI.2009.21

102	(not available)
103	(not available)
104	(not available)
105	10.1109/IEMBS.2009.5334557
106	(not available)
107	10.1109/IEMBS.2009.5333456
108	(not available)
109	10.1109/MHS.2009.5352071
110	10.1109/ACIL.2009.5349483
111	10.1088/1741-2560/6/4/046005
112	10.1007/978-3-642-03655-2_50
113*	10.1016/j.neunet.2009.07.003
114	10.1016/j.neuroimage.2009.06.076
115*	10.1007/978-3-642-02574-7_17
116	10.1109/ISCAS.2009.5118043
117	10.1007/978-3-642-01811-4_34
118*	10.1186/1743-0003-6-14
119	10.3171/2008.10.JNSo8466
120	10.1016/j.entcom.2009.09.007

Table 18: Record numbers and corresponding article DOIs returned by Scopus for the query given in §3.1.1. The articles that remained after data processing ($N = 31$) are marked with an asterisk.

FREQUENCY COMBINATION	MEAN RECALL	STDEV RECALL
6, 6.67, 7.5	73.52	17.36
6, 6.67, 8.57	71.24	19.44
6, 6.67, 10	69.90	13.46
6, 6.67, 12	69.33	19.01
6, 6.67, 15	71.62	19.01
6, 7.5, 8.57	78.29	17.21
6, 7.5, 10	77.33	10.21
6, 7.5, 12	76.38	16.75
6, 7.5, 15	74.29	16.70
6, 8.57, 10	76.76	11.85
6, 8.57, 12	74.48	17.85
6, 8.57, 15	77.52	15.31
6, 10, 12	73.90	10.55
6, 10, 15	74.86	14.04
6, 12, 15	75.43	15.75
6.67, 7.5, 8.57	75.81	19.99
6.67, 7.5, 10	75.43	14.70
6.67, 7.5, 12	77.71	19.16
6.67, 7.5, 15	72.00	21.72
6.67, 8.57, 10	74.67	14.77
6.67, 8.57, 12	77.33	18.89
6.67, 8.57, 15	78.10	19.37
6.67, 10, 12	76.19	12.82
6.67, 10, 15	73.52	15.69
6.67, 12, 15	78.48	17.75
7.5, 8.57, 10	81.14	11.86
7.5, 8.57, 12	82.48	17.49
7.5, 8.57, 15	78.48	18.66
7.5, 10, 12	83.62	7.49
7.5, 10, 15	76.19	13.87
7.5, 12, 15	78.67	14.89
8.57, 10, 12	82.1	11.04
8.57, 10, 15	80.00	13.15
8.57, 12, 15	82.67	15.03
10, 12, 15	81.33	10.69

Table 19: Recall (in percent) for all the 3-combinations of candidate stimulus frequencies to be used in MTSI-BCI.

NAME COMBINATION	MEAN RECALL	STDEV RECALL
Hector, Victor, Dexter	45.71	10.49
Hector, Victor, Lassie	69.52	18.80
Hector, Victor, Pluto	71.43	13.72
Hector, Victor, Shadow	74.29	11.17
Hector, Dexter, Lassie	63.81	6.51
Hector, Dexter, Pluto	64.76	7.42
Hector, Dexter, Shadow	56.19	14.84
Hector, Lassie, Pluto	89.52	14.33
Hector, Lassie, Shadow	85.71	11.17
Hector, Pluto, Shadow	80.00	12.77
Victor, Dexter, Lassie	60.95	13.01
Victor, Dexter, Pluto	60.95	12.43
Victor, Dexter, Shadow	65.71	6.00
Victor, Lassie, Pluto	87.62	14.10
Victor, Lassie, Shadow	91.43	13.17
Victor, Pluto, Shadow	87.62	19.41
Dexter, Lassie, Pluto	82.86	20.68
Dexter, Lassie, Shadow	82.86	11.45
Dexter, Pluto, Shadow	77.14	16.27
Lassie, Pluto, Shadow	87.62	13.01

Table 20: Recall (in percent) for all the 3-combinations of candidate dog names to be used in MTS!-ASR.

AUTHOR PUBLICATIONS CONTRIBUTING TO THE THESIS

The following author publications contribute to the corresponding chapters:

- Chapter 1: Gürkök and Nijholt (2012); Gürkök et al. (2012b)
- Chapter 2: Gürkök et al. (2011); van de Laar et al. (2011, 2012)
- Chapter 4: Gürkök et al. (2012); Obbink et al. (2012)
- Chapter 5: Hakvoort et al. (2011)
- Chapter 6: Gürkök et al. (2011); Gürkök et al. (2011a,b,c, 2012a)

- H. Gürkök and A. Nijholt. Brain-computer interfaces for multimodal interaction: A survey and principles. *International Journal of Human-Computer Interaction*, 28(5):292–307, 2012.
- H. Gürkök, G. Hakvoort, and M. Poel. Evaluating user experience with respect to user expectations in brain-computer interface games. In *Proceedings of the 5th International BCI Conference*, pages 348–351. Verlag der TU Graz, Graz, Austria, 2011.
- H. Gürkök, G. Hakvoort, and M. Poel. Evaluating user experience in a selection based brain-computer interface game: A comparative study. In *Entertainment Computing - ICEC 2011*, pages 77–88. Springer-Verlag, Berlin/Heidelberg, Germany, 2011a.
- H. Gürkök, G. Hakvoort, and M. Poel. Modality switching and performance in a thought and speech controlled computer game. In *Proceedings of the 13th International Conference on Multimodal Interfaces*, pages 41–48. ACM, New York, NY, USA, 2011b.
- H. Gürkök, G. Hakvoort, M. Poel, and A. Nijholt. User expectations and experiences of a speech and thought controlled computer game. In *Proceedings of the 8th International Conference on Advances in Computer Entertainment Technology*, pages 53:1–53:6. ACM, New York, NY, USA, 2011c.
- H. Gürkök, D. Plass-Oude Bos, B. van de Laar, F. Nijboer, and A. Nijholt. User experience evaluation in BCI: Filling the gap. *International Journal of Bioelectromagnetism*, 13(1):54–55, 2011.
- H. Gürkök, G. Hakvoort, M. Poel, and A. Nijholt. Meeting the expectations from brain-computer interfaces. *Computers in Entertainment*, 2012a. To appear.

- H. Gürkök, A. Nijholt, and M. Poel. Brain-computer interface games: Towards a framework. In *Entertainment Computing - ICEC 2012*, pages 373–380. Springer-Verlag, Berlin/Heidelberg, Germany, 2012b.
- H. Gürkök, A. Nijholt, M. Poel, and M. Obbink. Evaluating a multi-player brain-computer interface game: Challenge versus co-experience. *Entertainment Computing*, 2012. Under review.
- G. Hakvoort, H. Gürkök, D. Plass-Oude Bos, M. Obbink, and M. Poel. Measuring immersion and affect in a brain-computer interface game. In *Human-Computer Interaction - INTERACT 2011*, pages 115–128. Springer-Verlag, Berlin/Heidelberg, Germany, 2011.
- M. Obbink, H. Gürkök, D. Plass-Oude Bos, G. Hakvoort, M. Poel, and A. Nijholt. Social interaction in a cooperative brain-computer interface game. In *Proceedings of the 4th International ICST Conference on Intelligent Technologies for Interactive Entertainment*, pages 183–192. Springer-Verlag, Berlin/Heidelberg, Germany, 2012.
- B. van de Laar, H. Gürkök, D. Plass-Oude Bos, F. Nijboer, and A. Nijholt. Perspectives on user experience evaluation of brain-computer interfaces. In *Universal Access in Human-Computer Interaction. Users Diversity*, pages 600–609. Springer-Verlag, Berlin/Heidelberg, Germany, 2011.
- B. van de Laar, H. Gürkök, D. Plass-Oude Bos, F. Nijboer, and A. Nijholt. Brain-computer interfaces and user experience evaluation. In *Towards Practical Brain-Computer Interfaces: Bridging the Gap from Research to Real-World Applications*, pages 237–252. Springer-Verlag, Berlin/Heidelberg, Germany, 2012.

AUTHOR PUBLICATIONS RELEVANT TO THE THESIS

The following author publications are relevant to the theme of the thesis but do not contribute to the content.

- C. Mühl, H. Gürkök, D. Plass-Oude Bos, M. Thurlings, L. Scherffig, M. Duvinage, A. Elbakyan, S. Kang, M. Poel, and D. Heylen. Bacteria hunt: Evaluating multi-paradigm BCI interaction. *Journal on Multimodal User Interfaces*, 4(1):11–25, 2010.
- D. Plass-Oude Bos, B. Reuderink, B. van de Laar, H. Gürkök, C. Mühl, M. Poel, D. Heylen, and A. Nijholt. Human-computer interaction for bci games: Usability and user experience. In *2010 International Conference on Cyberworlds (CW)*, pages 277–281. IEEE Computer Society, Los Alamitos, CA, USA, 2010a.
- D. Plass-Oude Bos, B. Reuderink, B. van de Laar, H. Gürkök, C. Mühl, M. Poel, A. Nijholt, and D. Heylen. Brain-computer interfacing and games. In *Brain-Computer Interfaces: Applying our Minds to Human-Computer Interaction*, pages 149–178. Springer-Verlag London Ltd., London, UK, 2010b.
- D. Plass-Oude Bos, H. Gürkök, B. van de Laar, F. Nijboer, and A. Nijholt. User experience evaluation in BCI: Mind the gap! *International Journal of Bioelectromagnetism*, 13(1):48–49, 2011.
- B. van de Laar, F. Nijboer, H. Gürkök, D. Plass-Oude Bos, and A. Nijholt. User experience evaluation in BCI: Bridge the gap. *International journal of bioelectromagnetism*, 13(3):157–158, 2011.

BIBLIOGRAPHY

- B. Z. Allison and J. Polich. Workload assessment of computer gaming using a single-stimulus event-related potential paradigm. *Biological Psychology*, 77(3):277–283, 2008.
- A. Attwood, U. Frith, and B. Harmelin. The understanding and use of interpersonal gestures by autistic and Down’s syndrome children. *Journal of Autism and Developmental Disorders*, 18(2):241–257, 1988.
- L. Axelrod and K. S. Hone. Affectemes and all affects: A novel approach to coding user emotional expression during interactive experiences. *Behaviour & Information Technology*, 25(2):159–173, 2006.
- J. A. Bargas-Avila and K. Hornbæk. Old wine in new bottles or novel challenges: a critical analysis of empirical studies of user experience. In *Proceedings of the 2011 annual conference on Human factors in computing systems*, pages 2689–2698. ACM, New York, NY, USA, 2011.
- K. Battarbee. Defining co-experience. In *Proceedings of the 2003 international conference on Designing pleasurable products and interfaces*, pages 109–113. ACM, New York, NY, USA, 2003.
- K. Battarbee and I. Koskinen. Co-experience: user experience as interaction. *CoDesign*, 1(1):5–18, 2005.
- H. Berger. Über das elektrenkephalogramm des menschen. *Archiv für Psychiatrie und Nervenkrankheiten*, 87(1):527–570, 1929.
- R. Bernhaupt, editor. *Evaluating User Experience in Games: Concepts and Methods*. Springer-Verlag, London, UK, 2010.
- N. Bevan. Measuring usability as quality of use. *Software Quality Journal*, 4(2):115–130, 1995.
- N. Bianchi-Berthouze, W. W. Kim, and D. Patel. Does body movement engage you more in digital game play? and Why? In *Proceedings of the 2nd international conference on Affective Computing and Intelligent Interaction*, pages 102–113. Springer-Verlag, Berlin/Heidelberg, Germany, 2007.
- J. Bieger and G. Garcia Molina. Light stimulation properties to influence brain activity: A brain-computer interface application. Technical Report TN-2010-00315, Philips Research, Eindhoven, The Netherlands, 2010.

- M. Blythe and M. Hassenzahl. The semantics of fun: Differentiating enjoyable experiences. In *Funology: From Usability to Enjoyment*, pages 91–100. Kluwer Academic Publishers, Dordrecht, The Netherlands, 2003.
- M. Blythe, J. Reid, P. Wright, and E. Geelhoed. Interdisciplinary criticism: Analysing the experience of riot! a location-sensitive digital narrative. *Behaviour & Information Technology*, 25(2):127–139, 2006.
- BMW Group. The new BMW M5, September 2011. URL http://www.bimmerpost.com/goodiesforyou/new_m5.pdf.
- D. A. Bowman, E. Kruijff, J. J. LaViola, and I. Poupyrev. An introduction to 3-D user interface design. *Presence*, 10(1):96–108, 2001.
- M. M. Bradley and P. J. Lang. Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, 25(1):49–59, 1994.
- J. H. Brockmyer, C. M. Fox, K. A. Curtiss, E. McBroom, K. M. Burkhart, and J. N. Pidruzny. The development of the game engagement questionnaire: A measure of engagement in video game-playing. *Journal of Experimental Social Psychology*, 45(4):624–634, 2009.
- J. Brooke. SUS: a ‘quick and dirty’ usability scale. In *Usability Evaluation in Industry*, pages 189–194. Taylor & Francis Ltd., London, UK, 1996.
- E. Brown and P. Cairns. A grounded investigation of game immersion. In *CHI '04 extended abstracts on Human factors in computing systems*, pages 1297–1300. ACM, New York, NY, USA, 2004.
- T. Buchanan, T. Ali, T. Heffernan, J. Ling, A. Parrott, J. Rodgers, and A. Scholey. Nonequivalence of on-line and paper-and-pencil psychological tests: The case of the prospective memory questionnaire. *Behavior Research Methods*, 37(1):148–154, 2005.
- G. C. Burdea. Haptics issues in virtual environments. In *Proceedings of the Computer Graphics International, 2000*, pages 295–302. IEEE, Piscataway, NJ, USA, 2000.
- E. H. Calvillo-Gómez, P. Cairns, and A. L. Cox. Assessing the core elements of the gaming experience. In *Evaluating User Experience in Games*, pages 47–71. Springer-Verlag, London, UK, 2010.
- A. T. Campbell, T. Choudhury, S. Hu, H. Lu, M. K. Mukerjee, M. Rabbi, and R. D. Raizada. NeuroPhone: brain-mobile phone interface using a wireless EEG headset. In *MobiHeld '10: Proceedings of the second ACM SIGCOMM workshop on networking, systems, and applications on mobile handhelds*, pages 3–8. ACM, New York, NY, USA, 2010.

- J. M. Carroll and J. C. Thomas. Fun. *ACM SIGCHI Bulletin*, 19(3):21–24, 1988.
- J. Cassell, J. Sullivan, S. Prevost, and E. Churchill, editors. *Embodied conversational agents*. The MIT Press, Cambridge, MA, USA, 2000.
- G. Chanel, J. J. Kierkels, M. Soleymani, and T. Pun. Short-term emotion assessment in a recall paradigm. *International Journal of Human-Computer Studies*, 67(8):607–627, 2009.
- G. Chanel, S. Pelli, N. Ravaja, and K. Kuikkaniemi. Social interaction using mobile devices and biofeedback: effects on presence, attraction and emotions, 2010. Presented at BioS-Play 2010.
- K. Cheng and P. A. Cairns. Behaviour, realism and immersion in games. In *CHI '05 extended abstracts on Human factors in computing systems*, pages 1272–1275. ACM, New York, NY, USA, 2005.
- J. P. Chin, V. A. Diehl, and K. L. Norman. Development of an instrument measuring user satisfaction of the human-computer interface. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 213–218. ACM, New York, NY, USA, 1988.
- J. Coutaz, L. Nigay, D. Salber, A. Blandford, J. May, and R. M. Young. Four easy pieces for assessing the usability of multimodal interaction: the CARE properties. In *Human-Computer Interaction - INTERACT '95*, pages 115–120. Chapman & Hall, London, UK, 1995.
- B. Cowley, D. Charles, M. Black, and R. Hickey. Toward an understanding of flow in video games. *Computers in Entertainment*, 6(2): 20:1–20:27, 2008.
- M. Csíkszentmihályi. *Flow: The Psychology of Optimal Experience*. Harper Perennial, New York, NY, USA, 1990.
- G. D. Dawson. A summation technique for the detection of small evoked potentials. *Electroencephalography and Clinical Neurophysiology*, 6:65–84, 1954.
- H. Desurvire, M. Caplan, and J. A. Toth. Using heuristics to evaluate the playability of games. In *CHI '04 extended abstracts on Human factors in computing systems*, pages 1509–1512. ACM, New York, NY, USA, 2004.
- G. Deuschl and A. Eisen, editors. *Recommendations for the Practice of Clinical Neurophysiology*. Elsevier, Amsterdam, The Netherlands, 2nd edition, 1999.
- K. Devlin. *The Millennium Problems: The Seven Greatest Unsolved Mathematical Puzzles of Our Time*. Basic Books, New York, NY, USA, 2002.

- F. di Russo and D. Spinelli. Effects of sustained, voluntary attention on amplitude and latency of steady-state visual evoked potential: a costs and benefits analysis. *Clinical Neurophysiology*, 113(11):1771–1777, 2002.
- A. C. Dirican and M. Göktürk. Psychophysiological measures of human cognitive states applied in human computer interaction. *Procedia Computer Science*, 3:1361–1367, 2011.
- M. Donnerer and A. Steed. Using a P300 brain-computer interface in an immersive virtual environment. *Presence*, 19(1):12–24, 2010.
- M. Duvinage, T. Castermans, T. Dutoit, M. Petieau, T. Hoellinger, C. D. Saedeleer, K. Seetharaman, and G. Cheron. A P300-based quantitative comparison between the Emotiv Epoc headset and a medical EEG device. In *Biomedical Engineering*, pages 764–071. ACTA Press, Calgary, AB, Canada, 2012.
- G. Edlinger, C. Holzner, C. Groenegress, C. Guger, and M. Slater. Goal-oriented control with brain-computer interface. In *Foundations of Augmented Cognition. Neuroergonomics and Operational Neuroscience*, pages 732–740. Springer-Verlag, Berlin/Heidelberg, Germany, 2009.
- I. Ekman, G. Chanel, S. Järvelä, J. Kivikangas, M. Salminen, and N. Ravaja. Social interaction in games: Measuring physiological linkage and social presence. *Simulation & Gaming*, 43(3):321–338, 2012.
- S. Elliott and P. Nelson. Active noise control. *IEEE Signal Processing Magazine*, 10(4):12–35, 1993.
- L. Ermi and F. Mäyrä. Fundamental components of the gameplay experience: Analyzing immersion. In *Worlds in Play: International Perspectives on Digital Games Research*, pages 37–53. Peter Lang Publishing, New York, NY, USA, 2007.
- M. Fabiani, G. Gratton, and K. D. Federmeier. Event-related brain potentials: Methods, theory, and applications. In *Handbook of Psychophysiology*, pages 85–119. Cambridge University Press, Cambridge, UK, 3rd edition, 2007.
- C. Fabricatore, M. Nussbaum, and R. Rosas. Playability in action videogames: a qualitative design model. *Human-Computer Interaction*, 17(4):311–368, 2002.
- S. H. Fairclough. Physiological computing: Interfacing with the human nervous system. In *Sensing Emotions*, pages 1–20. Springer Science+Business Media B.V., Dordrecht, The Netherlands, 2011.

- L. A. Farwell and E. Donchin. Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. *Electroencephalography and Clinical Neurophysiology*, 70(6):510–523, 1988.
- M. Fatourechi, G. E. Birch, and R. K. Ward. A self-paced brain interface system that uses movement related potentials and changes in the power of brain rhythms. *Journal of Computational Neuroscience*, 23(1): 21–37, 2007.
- A. Febretti and F. Garzotto. Usability, playability, and long-term engagement in computer games. In *Proceedings of the 27th international conference extended abstracts on Human factors in computing systems*, pages 4063–4068. ACM, New York, NY, USA, 2009.
- M. A. Federoff. Heuristics and usability guidelines for the creation and evaluation of fun in video games. Master’s thesis, Indiana University, Bloomington, IN, USA, 2002.
- A. Fernandez. Fun experience with digital games: A model proposition. In *Extending Experiences: Structure, Analysis and Design of Computer Game Player Experience*, pages 181–190. Lapland University Press, Rovaniemi, Finland, 2008.
- P. W. Ferrez and J. del R. Millán. Error-related EEG potentials in brain-computer interfaces. In *Toward Brain-Computer Interfacing*, pages 291–301. The MIT Press, Cambridge, MA, USA, 2007.
- A. Fink, R. H. Grabner, M. Benedek, G. Reishofer, V. Hauswirth, M. Fally, C. Neuper, F. Ebner, and A. C. Neubauer. The creative brain: Investigation of brain activity during creative problem solving by means of EEG and fMRI. *Human Brain Mapping*, 30(3):734–748, 2009.
- A. Finke, A. Lenhardt, and H. Ritter. The MindGame: A P300-based brain-computer interface game. *Neural Networks*, 22(9):1329–1333, 2009.
- K. Finstad. The usability metric for user experience. *Interacting with Computers*, 22(5):323–327, 2010.
- C. Fisher and P. Sanderson. Exploratory sequential data analysis: exploring continuous observational data. *interactions*, 3(2):25–34, 1996.
- J. Forlizzi and K. Battarbee. Understanding experience in interactive systems. In *Proceedings of the 5th conference on Designing interactive systems: processes, practices, methods, and techniques*, pages 261–268. ACM, New York, NY, USA, 2004.
- K. Förster, A. Biasiucci, R. Chavarriaga, J. del R. Millán, D. Roggen, and G. Tröster. On the use of brain decoded signals for online user

- adaptive gesture recognition systems. In *Pervasive Computing*, pages 427–444. Springer, Berlin/Heidelberg, Germany, 2010.
- D. Friedman, R. Leeb, C. Guger, A. Steed, G. Pfurtscheller, and M. Slater. Navigating virtual reality by thought: what is it like? *Presence*, 16(1): 100–110, 2007.
- E. Frøkjær, M. Hertzum, and K. Hornbæk. Measuring usability: are effectiveness, efficiency, and satisfaction really correlated? In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 345–352. ACM, New York, NY, USA, 2000.
- G. Gediga, K.-C. Hamborg, and I. Düntsch. The IsoMetrics usability inventory: An operationalization of ISO 9241-10 supporting summative and formative evaluation of software systems. *Behaviour & Information Technology*, 18(3):151–164, 1999.
- P. Gloor. *Hans Berger on the Electroencephalogram of Man. The Fourteen Original Reports on the Human Encephalogram*. Elsevier, Amsterdam, The Netherlands, 1969.
- J. L. González Sánchez, N. Padilla Zea, and F. L. Gutiérrez. From usability to playability: Introduction to player-centred video game development process. In *Human Centered Design*, pages 65–74. Springer-Verlag, Berlin/Heidelberg, Germany, 2009.
- R. B. Grady. *Practical software metrics for project management and process improvement*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1992.
- C. Groenegress, C. Holzner, C. Guger, and M. Slater. Effects of P300-based BCI use on reported presence in a virtual environment. *Presence*, 19(1):1–11, 2010.
- G. Hakvoort, B. Reuderink, and M. Obbink. Comparison of PSDA and CCA detection methods in a SSVEP-based BCI-system. Technical Report TR-CTIT-11-03, Centre for Telematics and Information Technology, University of Twente, Enschede, The Netherlands, 2011.
- B. Hamadicharef. International collaborations in brain-computer interface (BCI) research. In *Web Information Systems and Mining*, pages 35–42. Springer-Verlag, Berlin/Heidelberg, Germany, 2011.
- S. G. Hart and L. E. Staveland. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Human Mental Workload*, pages 139–183. North-Holland, Amsterdam, The Netherlands, 1988.
- M. Hassenzahl. The interplay of beauty, goodness, and usability in interactive products. *Human-Computer Interaction*, 19(4):319–349, 2004.

- M. Hassenzahl and A. Monk. The inference of perceived usability from beauty. *Human-Computer Interaction*, 25(3):235–260, 2010.
- M. Hassenzahl and N. Tractinsky. User experience - a research agenda. *Behaviour & Information Technology*, 25(2):91–97, 2006.
- M. Hassenzahl, A. Beu, and M. Burmester. Engineering joy. *IEEE Software*, 18(1):70–76, 2001.
- M. Hassenzahl, M. Burmester, and F. Koller. AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität. In *Mensch & Computer 2003: Interaktion in Bewegung*, pages 187–196. B. G. Teubner, Stuttgart, Germany, 2003.
- M. Hassenzahl, S. Diefenbach, and A. Göritz. Needs, affect, and interactive products - Facets of user experience. *Interacting with Computers*, 22(5):353–362, 2010.
- E. Hatfield, J. Cacioppo, and R. L. Rapson. *Emotional Contagion*. Cambridge University Press, New York, NY, USA, 1994.
- C. Heath, M. Sanchez Svensson, J. Hindmarsh, P. Luff, and D. vom Lehn. Configuring awareness. *Computer Supported Cooperative Work*, 11(3):317–347, 2002.
- A. T. Herdman, O. Lins, P. van Roon, D. R. Stapells, M. Scherg, and T. W. Picton. Intracerebral sources of human auditory steady-state responses. *Brain Topography*, 15(2):69–86, 2002.
- S. I. Hjelm. Research + design: the making of Brainball. *interactions*, 10(1):26–34, 2003.
- C. Holzner, C. Guger, G. Edlinger, C. Groneggress, and M. Slater. Virtual smart home controlled by thoughts. In *18th IEEE International Workshops on Enabling Technologies: Infrastructures for Collaborative Enterprises*, pages 236–239. IEEE, Piscataway, NJ, USA, 2009.
- International Society for Presence Research. The concept of presence: Explication statement, 2000. URL <http://ispr.info/about-presence-2/about-presence/>.
- ISO 9241-11:1998. *Ergonomic requirements for office work with visual display terminals (VDTs) - Part 11: Guidance on usability*. ISO, Geneva, Switzerland, 1998.
- ISO 9241-210:2010. *Ergonomics of human-system interaction - Part 210: Human-centred design for interactive systems*. ISO, Geneva, Switzerland, 2010.

- ISO/IEC 9126-1:2001. *Software engineering - Product quality - Part 1: Quality model*. ISO, Geneva, Switzerland, 2001.
- J. H. Janssen, J. N. Bailenson, W. A. IJsselsteijn, and J. H. D. M. West-erink. Intimate heartbeats: Opportunities for affective communication technology. *IEEE Transactions on Affective Computing*, 1(2):72–80, 2010.
- H. H. Jasper. Report of the committee on methods of clinical examination in electroencephalography: 1957. *Electroencephalography and Clinical Neurophysiology*, 10(2):370–375, 1958.
- C. Jennett, A. L. Cox, P. Cairns, S. Dhoparee, A. Epps, T. Tijs, and A. Walton. Measuring and defining the experience of immersion in games. *International Journal of Human-Computer Studies*, 66(9):641–661, 2008.
- C. I. Jennett. *Investigating Real World Dissociation in Computer Game Immersion: Manipulating Sense of Progression and Measuring Awareness of Distracters*. PhD thesis, UCL Interaction Centre, London, UK, 2009.
- D. Johnson and J. Wiles. Effective affective user interface design in games. *Ergonomics*, 46(13-14):1332–1345, 2003.
- C. Jones, L. McIver, L. Gibson, and P. Gregor. Experiences obtained from designing with children. In *Proceedings of the 2003 conference on Interaction design and children*, pages 69–74. ACM, New York, NY, USA, 2003.
- J. Kagan. Motives and development. *Journal of Personality and Social Psychology*, 22(1):51–66, 1972.
- S. Kelly, E. Lalor, C. Finucane, and R. Reilly. A comparison of covert and overt attention as a control option in a steady-state visual evoked potential-based brain computer interface. In *26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (IEMBS'04)*, pages 4725–4728. IEEE, Piscataway, NJ, USA, 2004.
- L. Kerawalla, D. Pearce, N. Yuill, R. Luckin, and A. Harris. "I'm keeping those there, are you?" The role of a new user interface paradigm - Separate Control of Shared Space (SCOSS) - in the collaborative decision-making process. *Computers & Education*, 50(1):193–206, 2008.
- M. Kipp. ANVIL - a generic annotation tool for multimodal dialogue. In *EUROSPEECH-2001*, pages 1367–1370. ISCA, Baixas, France, 2001.
- J. Kirakowski and M. Corbett. SUMI: the Software Usability Measurement Inventory. *British Journal of Educational Technology*, 24(3):210–212, 1993.

- M. Kivikangas, I. Ekman, G. Chanel, S. Järvelä, B. Cowley, M. Salminen, P. Henttonen, and N. Ravaja. Review on psychophysiological methods in game research. In *Proceedings of DiGRA Nordic 2010: Experiencing Games: Games, Play, and Players*. DIGRA, 2010.
- C. Klimmt, T. Hartmann, and A. Frey. Effectance and control as determinants of video game enjoyment. *CyberPsychology & Behavior*, 10(6): 845–848, 2007.
- C. Koeffel, W. Hochleitner, J. Leitner, M. Haller, A. Geven, and M. Tschelligi. Using heuristics to evaluate the overall user experience of video games and advanced interaction games. In *Evaluating User Experience in Games*, pages 233–256. Springer-Verlag, London, UK, 2010.
- S. Koelstra, C. Mühl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras. DEAP: A database for emotion analysis using physiological signals. *IEEE Transactions on Affective Computing*, 3(1):18–31, 2012.
- H. Korhonen and E. M. I. Koivisto. Playability heuristics for mobile games. In *Proceedings of the 8th conference on Human-computer interaction with mobile devices and services*, pages 9–16. ACM, New York, NY, USA, 2006.
- R. Krepki, B. Blankertz, G. Curio, and K.-R. Müller. The Berlin brain-computer interface (BBCI) - towards a new communication channel for online control in gaming applications. *Multimedia Tools and Applications*, 33(1):73–90, 2007.
- A. Kübler and K.-R. Müller. An introduction to brain-computer interfacing. In *Toward Brain-Computer Interfacing*, pages 1–25. The MIT Press, Cambridge, MA, USA, 2007.
- M. Kubovy. On the pleasures of the mind. In *Well-being: The Foundations of Hedonic Psychology*, pages 134–154. Russell Sage Foundation, New York, NY, USA, 1999.
- S. Kühn, B. C. N. Müller, A. van der Leij, A. Dijksterhuis, M. Brass, and R. B. van Baaren. Neural correlates of emotional synchrony. *Social Cognitive and Affective Neuroscience*, 6(3):368–374, 2011.
- A. Kultima and J. Stenros. Designing games for everyone: the expanded game experience model. In *Proceedings of the International Academic Conference on the Future of Game Design and Technology*, pages 66–73. ACM, New York, NY, USA, 2010.
- R. W. Lansing, E. Schwartz, and D. B. Lindsley. Reaction time and EEG activation under alerted and nonalerted conditions. *Journal of Experimental Psychology*, 58(1):1–7, 1959.

- E. L.-C. Law and P. van Schaik. Modelling user experience - an agenda for research and practice. *Interacting with Computers*, 22(5):313–322, 2010.
- N. Lazzaro. Why we play games: Four keys to more emotion without story, March 2004. URL http://www.xeodesign.com/xeodesign_whyweplaygames.pdf.
- M. A. Lebedev and M. A. Nicolelis. Brain-machine interfaces: past, present and future. *Trends in Neurosciences*, 29(9):536–546, 2006.
- J. E. LeDoux. Emotion: Clues from the brain. *Annual Review of Psychology*, 46:209–235, 1995.
- R. Leeb, D. Friedman, G. R. Müller-Putz, R. Scherer, M. Slater, and G. Pfurtscheller. Self-paced (asynchronous) BCI control of a wheelchair in virtual environments: a case study with a tetraplegic. *Computational Intelligence and Neuroscience*, 2007:1–12, 2007.
- J. Legény, R. Vicianá Abad, and A. Lécuyer. Navigating in virtual worlds using a self-paced SSVEP-based brain-computer interface with integrated stimulation and real-time feedback. *Presence*, 20(6):529–544, 2011.
- O. Leino, H. Wirman, and A. Fernandez, editors. *Extending Experiences: Structure, Analysis and Design of Computer Game Player Experience*. Lapland University Press, Rovaniemi, Finland, 2008.
- J. R. Lewis. IBM computer usability satisfaction questionnaires: Psychometric evaluation and instructions for use. *International Journal of Human-Computer Interaction*, 7(1):57–78, 1995.
- Y. Li, C. S. Nam, B. B. Shadden, and S. L. Johnson. A P300-based brain-computer interface: Effects of interface type and screen size. *International Journal of Human-Computer Interaction*, 27(1):52–68, 2011.
- L.-D. Liao, C.-T. Lin, K. McDowell, A. Wickenden, K. Gramann, T.-P. Jung, L.-W. Ko, and J.-Y. Chang. Biosensor technologies for augmented brain-computer interfaces in the next decades. *Proceedings of the IEEE*, 100:1553–1566, 2012.
- A. Light. Adding method to meaning: A technique for exploring peoples’ experience with technology. *Behaviour & Information Technology*, 25(2):175–187, 2006.
- Z. Lin, C. Zhang, W. Wu, and X. Gao. Frequency recognition based on canonical correlation analysis for SSVEP-based BCIs. *IEEE Transactions on Biomedical Engineering*, 53(12):2610–2614, 2006.

- S. E. Lindley and A. F. Monk. Social enjoyment with electronic photograph displays: Awareness and control. *International Journal of Human-Computer Studies*, 66(8):587–604, 2008.
- S. E. Lindley, J. Le Couteur, and N. Bianchi-Berthouze. Stirring up experience through movement in game play: effects on engagement and social behaviour. In *Proceedings of the twenty-sixth annual SIGCHI conference on Human factors in computing systems*, pages 511–514. ACM, New York, NY, USA, 2008.
- C. Lord, S. Risi, L. Lambrecht, E. H. Cook, B. L. Leventhal, P. C. DiLavore, A. Pickles, and M. Rutter. The Autism Diagnostic Observation Schedule-Generic: A standard measure of social and communication deficits associated with the spectrum of autism. *Journal of Autism and Developmental Disorders*, 30:205–223, 2000.
- S. J. Luck. *An Introduction to Event-Related Potentials and Their Neural Origins*. The MIT Press, Cambridge, MA, USA, 2005.
- R. L. Mandryk, M. S. Atkins, and K. M. Inkpen. A continuous and objective evaluation of emotional experience with interactive play environments. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*, pages 1027–1036. ACM, New York, NY, USA, 2006a.
- R. L. Mandryk, K. M. Inkpen, and T. W. Calvert. Using psychophysiological techniques to measure user experience with entertainment technologies. *Behaviour & Information Technology*, 25(2):141–158, 2006b.
- T. Manninen. Interaction forms and communicative actions in multi-player games. *The International Journal of Computer Game Research*, 3(1), 2003.
- P. Martinez, H. Bakardjian, and A. Cichocki. Fully online multicommand brain-computer interface with visual neurofeedback using SSVEP paradigm. *Computational Intelligence and Neuroscience*, 2007: 94561, 2007.
- C. A. Mateer and M. M. Sohlberg, editors. *Cognitive Rehabilitation: An Integrative Neuropsychological Approach*. The Guilford Press, New York, NY, USA, 2nd edition, 2001.
- E. Mayo. *The human problems of an industrial civilisation*. Macmillan, New York, NY, USA, 1933.
- D. J. McFarland, L. M. McCane, S. V. David, and J. R. Wolpaw. Spatial filter selection for EEG-based communication. *Electroencephalography and Clinical Neurophysiology*, 103(3):386–394, 1997.

- N. McNamara and J. Kirakowski. Functionality, usability, and user experience: three areas of concern. *interactions*, 13:26–28, 2006.
- M. Middendorf, G. McMillan, G. Calhoun, and K. Jones. Brain-computer interfaces based on the steady-state visual-evoked response. *IEEE Transactions on Rehabilitation Engineering*, 8(2):211–214, 2000.
- M. Moore Jackson, R. Mappus, E. Barba, S. Hussein, G. Venkatesh, C. Shastry, and A. Israeli. Continuous control paradigms for direct brain interfaces. In *Human-Computer Interaction. Novel Interaction Methods and Techniques*, pages 588–595. Springer-Verlag, Berlin/Heidelberg, Germany, 2009.
- A. Mouraux and G. Iannetti. Across-trial averaging of event-related EEG responses and beyond. *Magnetic Resonance Imaging*, 26(7):1041–1054, 2008.
- C. Mühl, H. Gürkök, D. Plass-Oude Bos, M. Thurlings, L. Scherffig, M. Duvinage, A. Elbakyan, S. Kang, M. Poel, and D. Heylen. Bacteria hunt: Evaluating multi-paradigm BCI interaction. *Journal on Multimodal User Interfaces*, 4(1):11–25, 2010.
- G. Muller-Putz, R. Scherer, C. Neuper, and G. Pfurtscheller. Steady-state somatosensory evoked potentials: suitable brain signals for brain-computer interfaces? *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 14(1):30–37, 2006.
- L. Nacke and C. A. Lindley. Flow and immersion in first-person shooters: measuring the player’s gameplay experience. In *Proceedings of the 2008 Conference on Future Play: Research, Play, Share*, pages 81–88. ACM, New York, NY, USA, 2008.
- L. Nacke, A. Drachen, and S. Göbel. Methods for evaluating gameplay experience in a serious gaming context. *International Journal of Computer Science in Sport*, 9(2), 2010a.
- L. E. Nacke. *Affective Ludology: Scientific Measurement of User Experience in Interactive Entertainment*. PhD thesis, Blekinge Institute of Technology, Karlskrona, Sweden, 2009.
- L. E. Nacke, M. N. Grimshaw, and C. A. Lindley. More than a feeling: Measurement of sonic user experience and psychophysiology in a first-person shooter game. *Interacting with Computers*, 22(5):336–343, 2010b.
- C. Neuper, R. Scherer, M. Reiner, and G. Pfurtscheller. Imagery of motor actions: Differential effects of kinesthetic and visual-motor mode of imagery in single-trial EEG. *Cognitive Brain Research*, 25(3):668–677, 2005.

- J. Nielsen. *Usability Engineering*. Academic Press, San Diego, CA, USA, 1993.
- J. Nielsen. Usability inspection methods. In *Conference companion on Human factors in computing systems*, pages 413–414. ACM, New York, NY, USA, 1994a.
- J. Nielsen. Heuristic evaluation. In *Usability Inspection Methods*, pages 25–62. John Wiley & Sons, New York, NY, USA, 1994b.
- F. Nijboer, J. Clausen, B. Allison, and P. Haselager. The Asilomar survey: Stakeholders’ opinions on ethical issues related to brain-computer interfacing. *Neuroethics*, 2012. To appear.
- A. Nijholt, B. Reuderink, and D. Oude Bos. Turning shortcomings into challenges: Brain-computer interfaces for games. In *Intelligent Technologies for Interactive Entertainment*, pages 153–168. Springer-Verlag, Berlin/Heidelberg, Germany, 2009.
- A. R. Nikolaev and A. P. Anokhin. EEG frequency ranges during perception and mental rotation of two- and three-dimensional objects. *Neuroscience and Behavioral Physiology*, 28(6):670–677, 1998.
- S. Ogletree and R. Drake. College students’ video game participation and perceptions: Gender differences and implications. *Sex Roles*, 56(7):537–542, 2007.
- K. O’Hara, A. Sellen, and R. Harper. Embodiment in brain-computer interaction. In *Proceedings of the 2011 annual conference on Human factors in computing systems*, pages 353–362. ACM, New York, NY, USA, 2011.
- K. Oum, H. Ayaz, P. Shewokis, and P. Diefenbach. MindTactics: A brain computer interface gaming platform. In *2010 International IEEE Consumer Electronics Society’s Games Innovations Conference (ICE-GIC)*. IEEE, Piscataway, NJ, USA, 2010.
- K. Overbeeke, T. Djajadiningrat, C. Hummels, and S. Wensveen. Beauty in usability: Forget about ease of use! In *Pleasure with Products, Beyond Usability*, pages 9–18. Taylor & Francis Ltd., London, UK, 2002.
- K. Overbeeke, T. Djajadiningrat, C. Hummels, S. Wensveen, and J. Prens. Let’s make things engaging. In *Funology: From Usability to Enjoyment*, pages 7–17. Kluwer Academic Publishers, Dordrecht, The Netherlands, 2003.
- S. Oviatt. Taming recognition errors with a multimodal interface. *Communications of the ACM*, 43:45–51, 2000.

- S. Oviatt, P. Cohen, L. Wu, J. Vergo, L. Duncan, B. Suhm, J. Bers, T. Holzman, T. Winograd, J. Landay, J. Larson, and D. Ferro. Designing the user interface for multimodal speech and pen-based gesture applications: state-of-the-art systems and future research directions. *Human-Computer Interaction*, 15(4):263–322, 2000.
- G. Pfurtscheller and F. H. Lopes da Silva. Event-related EEG/MEG synchronization and desynchronization: basic principles. *Clinical Neurophysiology*, 110(11):1842–1857, 1999.
- G. Pfurtscheller, C. Brunner, A. Schlögl, and F. H. Lopes da Silva. Mu rhythm (de)synchronization and EEG single-trial classification of different motor imagery tasks. *NeuroImage*, 31(1):153–159, 2006.
- T. W. Picton, M. S. John, A. Dimitrijevic, and D. Purcell. Human auditory steady-state responses. *International Journal of Audiology*, 42(4):177–219, 2003.
- D. Plass-Oude Bos, B. Reuderink, B. van de Laar, H. Gürkök, C. Mühl, M. Poel, A. Nijholt, and D. Heylen. Brain-computer interfacing and games. In *Brain-Computer Interfaces: Applying our Minds to Human-Computer Interaction*, pages 149–178. Springer-Verlag London Ltd., London, UK, 2010.
- D. Plass-Oude Bos, M. Poel, and A. Nijholt. A study in user-centered design and evaluation of mental tasks for BCI. In *Advances in Multimedia Modeling*, pages 122–134. Springer-Verlag, Berlin/Heidelberg, Germany, 2011.
- K. Poels, W. IJsselsteijn, Y. de Kort, and B. Van Iersel. Digital games, the aftermath: Qualitative insights into postgame experiences. In *Evaluating User Experience in Games*, pages 149–163. Springer-Verlag, London, UK, 2010.
- G. Prasad, P. Herman, D. Coyle, S. McDonough, and J. Crosbie. Applying a brain-computer interface to support motor imagery practice in people with stroke for upper limb recovery: a feasibility study. *Journal of NeuroEngineering and Rehabilitation*, 7:60, 2010.
- D. Regan. Steady-state evoked potentials. *Journal of the Optical Society of America*, 67(11):1475–1489, 1977.
- D. Reidsma. *Annotations and Subjective Machines: Of Annotators, Embodied Agents, Users, and Other Humans*. PhD thesis, Centre for Telematics and Information Technology, University of Twente, Enschede, The Netherlands, 2008.
- G. Riva. Is presence a technology issue? Some insights from cognitive sciences. *Virtual Reality*, 13(3):159–169, 2009.

- S. R. Rochester. The significance of pauses in spontaneous speech. *Journal of Psycholinguistic Research*, 2(1):51–81, 1973.
- R. Rouse. *Game Design: Theory & Practice*. Wordware, Sudbury, MA, USA, 2nd edition, 2005.
- J. A. Russell and L. F. Barrett. Core affect, prototypical emotional episodes, and other things called emotion: Dissecting the elephant. *Journal of Personality and Social Psychology*, 76(5):805–819, 1999.
- D. Salber and J. Coutaz. Applying the Wizard of Oz technique to the study of multimodal systems. In *Human-Computer Interaction*, pages 219–230. Springer-Verlag, Berlin/Heidelberg, Germany, 1993.
- M. Salvaris and F. Sepulveda. Visual modifications on the P300 speller BCI paradigm. *Journal of Neural Engineering*, 6(4):046011, 2009.
- R. Scherer, A. Schloegl, F. Lee, H. Bischof, J. Janša, and G. Pfurtscheller. The self-paced Graz brain-computer interface: Methods and applications. *Computational Intelligence and Neuroscience*, 2007:79826, 2007.
- A. Schlögl, J. Kronegg, J. E. Huggins, and S. G. Mason. Evaluation criteria for BCI research. In *Toward Brain-Computer Interfacing*, pages 327–342. The MIT Press, Cambridge, MA, USA, 2007.
- E. W. Sellers and E. Donchin. A P300-based brain-computer interface: initial tests by ALS patients. *Clinical Neurophysiology*, 117(3):538–548, 2006.
- H. Serby, E. Yom-Tov, and G. Inbar. An improved P300-based brain-computer interface. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 13(1):89–98, 2005.
- R. Sharma, V. Pavlovic, and T. Huang. Toward multimodal human-computer interface. *Proceedings of the IEEE*, 86(5):853–869, 1998.
- K. M. Sheldon, A. J. Elliot, Y. Kim, and T. Kasser. What is satisfying about satisfying events? Testing 10 candidate psychological needs. *Journal of Personality and Social Psychology*, 80(2):325–339, 2001.
- H. Shibasaki and M. Hallett. What is the Bereitschaftspotential? *Clinical Neurophysiology*, 117(11):2341–2356, 2006.
- N. Sobell and M. Trivich. Brainwave drawing game. In *Delicate Balance: Technics, Culture and Consequences*, 1989, pages 360–362. Los Angeles Chapter IEEE SSIT-30, Torrance, CA, USA, 1989.
- I. Soute, P. Markopoulos, and R. Magielse. Head Up Games: combining the best of both worlds by merging traditional and digital play. *Personal and Ubiquitous Computing*, 14(5):435–444, 2010.

- P. Sweetser and P. Wyeth. Gameflow: a model for evaluating player enjoyment in games. *Computers in Entertainment*, 3(3):1–24, 2005.
- E. S. Tan and J. Jansz. The game experience. In *Product Experience*, pages 531–556. Elsevier, San Diego, CA, USA, 2008.
- M. Tangermann, M. Krauledat, K. Grzeska, M. Sagebaum, B. Blankertz, C. Vidaurre, and K.-R. Müller. Playing Pinball with non-invasive BCI. In *Advances in Neural Information Processing Systems 21*, pages 1641–1648. The MIT Press, Cambridge, MA, USA, 2009.
- E. L. Thorndike. A constant error in psychological ratings. *Journal of Applied Psychology*, 4(1):25–29, 1920.
- M. Turunen, J. Hakulinen, A. Melto, T. Heimonen, T. Laivo, and J. Hella. SUXES - user experience evaluation method for spoken and multimodal interaction. In *Proceedings of INTERSPEECH 2009*, pages 2567–2570. ISCA, Baixas, France, 2009.
- J. van Baren and W. IJsselstein. Measuring presence: A guide to current measurement approaches. OmniPres Deliverable 5, 2004.
- M. van Gerven, J. Farquhar, R. Schaefer, R. Vlek, J. Geuze, A. Nijholt, N. Ramsey, P. Haselager, L. Vuurpijl, S. Gielen, and P. Desain. The brain-computer interface cycle. *Journal of Neural Engineering*, 6(4):041001, 2009.
- J. A. Veltman and A. W. K. Gaillard. Physiological workload reactions to increasing levels of task difficulty. *Ergonomics*, 41(5):656–669, 1998.
- J. J. Vidal. Toward direct brain-computer communication. *Annual Review of Biophysics and Bioengineering*, 2(1):157–180, 1973.
- E. M. Voorhees. The TREC-8 question answering track report. In *Proceedings of the 8th Text Retrieval Conference*, pages 77–82. NIST, Gaithersburg, MD, USA, 1999.
- W. Walker, P. Lamere, P. Kwok, B. Raj, R. Singh, E. Gouvea, P. Wolf, and J. Woelfel. Sphinx-4: A flexible open source framework for speech recognition. Technical Report SMLI TR-2004-139, Sun Microsystems Inc., Menlo Park, CA, USA, 2004.
- S. Walter, C. Quigley, S. K. Andersen, and M. M. Mueller. Effects of overt and covert attention on the steady-state visual evoked potential. *Neuroscience Letters*, 519(1):37–41, 2012.
- M. W. Watkins and M. Pacheco. Interobserver agreement in behavioral research: Importance and calculation. *Journal of Behavioral Education*, 10(4):205–212, 2000.

- F. J. Wertz, K. Charmaz, L. M. McMullen, R. Josselson, R. Anderson, and E. McSpadden. *Five Ways of Doing Qualitative Analysis: Phenomenological Psychology, Grounded Theory, Discourse Analysis, Narrative Research, and Intuitive Inquiry*. The Guildford Press, New York, NY, USA, 2011.
- R. W. White. Motivation reconsidered: the concept of competence. *Psychological Review*, 66(5):297–333, 1959.
- S. S. Wiltermuth and C. Heath. Synchrony and cooperation. *Psychological Science*, 20(1):1–5, 2009.
- B. G. Witmer and M. J. Singer. Measuring presence in virtual environments: A presence questionnaire. *Presence*, 7(3):225–240, 1998.
- J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan. Brain-computer interfaces for communication and control. *Clinical Neurophysiology*, 113(6):767–791, 2002.
- P. Wright, J. McCarthy, and L. Meekison. Making sense of experiences. In *Funology: From Usability to Enjoyment*, pages 43–53. Kluwer Academic Publishers, Dordrecht, The Netherlands, 2003.
- N. Yuill and Y. Rogers. Mechanisms for collaboration: A design and evaluation framework for multi-user interfaces. *ACM Transactions on Computer-Human Interaction*, 19(1):1:1–1:24, 2012.
- T. O. Zander, C. Kothe, S. Jatzev, and M. Gaertner. Enhancing human-computer interaction with input from active and passive brain-computer interfaces. In *Brain-Computer Interfaces*, pages 181–199. Springer-Verlag London Ltd., London, UK, 2010.
- Z. Zeng, M. Pantic, G. Roisman, and T. Huang. A survey of affect recognition methods: Audio, visual, and spontaneous expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(1):39–58, 2009.

ABSTRACT

A brain-computer interface (BCI) infers our actions (e.g. a movement), intentions (e.g. preparation for a movement) and psychological states (e.g. emotion, attention) by interpreting our brain signals. It uses the inferences it makes to manipulate a computer. Although BCIs have long been used exclusively to support disabled people (e.g. through brain-controlled wheelchairs, spellers), with the emerging low-cost and portable hardware, they have started to be considered for a variety of human-computer interaction (HCI) applications for non-disabled people as well. Among these, games have been receiving the interest of researchers and practitioners from both the BCI and HCI communities.

In BCI research, games have long been used solely to demonstrate the performance of signal processing and analysis methods. Therefore, they have been evaluated only for their performance (e.g. recognition accuracy, information transfer rate). However, games are not meant to satisfy our practical needs. They satisfy our hedonic needs. They challenge us, let us make our fantasies true, evoke our memories, and so on. We look for these experiences while playing games. Thus, rather than the performance of the controller used, the user experience (UX) of the game is essential.

UX of a game is a consequence of the player's internal state, the game characteristics and the context. Evaluating such a complex phenomenon is non-trivial. Often, UX is measured in terms of other, measurable concepts such as flow, immersion, presence, social behaviour and so on. Methods to evaluate UX also vary and include questionnaire, interviewing, and observation analysis. Evaluating UX of BCI games is even harder because UX evaluations may be biased due to the low recognition performance of the BCI. But this should not keep us from investigating UX of BCI games and identifying the good and bad practices, independent of performance. Because, ignoring UX while trying to improve performance might lead to games that are perfectly functional, but not enjoyed or played by anyone.

In this work, we investigated how the BCI control can influence the UX of a computer game. We considered a futuristic scenario in which BCI functioned as perfectly as other modalities. To simulate this scenario, we proposed and followed an approach called equalised comparative evaluation (ECE). For this, we equalised the perceived performance of BCI and several other modalities. We did not simply introduce artificial errors on the modalities as this could reduce player effectance and, thus, enjoyment. Instead, we manipulated the challenge

of the tasks the players performed. Then, we evaluated and compared the UX while playing with BCI and with the other modalities.

Our work consisted of three studies in each of which we evaluated different UX related concepts and used different data collection methods. In all the studies, participants played an experimental multimodal game that we had developed, called Mind the Sheep! (MTS!). They controlled 3 dogs using different modalities in order to herd 10 sheep across a meadow. The goal was to pen all the sheep as quickly as possible.

In Study 1, we showed the effectiveness of our ECE approach. Pairs of participants played a collaborative, multi-player version of MTS! once using a BCI that relied on the steady-state visually evoked potential (SSVEP), once by simple mouse pointing and clicking (non-ECE approach) and once using a visuomotor control mechanism that was as challenging to control as BCI (ECE approach). We relied on observation analysis, interviewing and questionnaires to evaluate UX in terms of social interaction. We found that challenging control dampened collaborative social interaction but it improved emotional social interaction.

In Study 2, participants played single-player MTS! once using BCI and once using the visuomotor control mechanism we used in Study 1. They indicated their UX in terms of affect and immersion using questionnaires. We found that the BCI selection method was more immersive and that the participants were more indulgent towards BCI control. One question that arose from our findings was whether the positive UX of BCI control was due to a novelty effect. This was what we investigated in Study 3.

In Study 3, we compared UX of BCI control to that of automatic speech recogniser (ASR) control, under the assumption that both ASR and BCI were novel game input modalities. Participants played single-player MTS! once with BCI, once with ASR and once with the option of switching between the two. Using questionnaires, they rated their expectations, engagement and workload levels as well as perceived game/controller quality. We also conducted interviews and analysed game logs. The participants rated BCI control higher in hedonic quality but lower in pragmatic quality than ASR control. The challenge and novelty of BCI influenced their modality switching behaviour.

The contributions of our work are manifold. The ECE approach we proposed allows evaluating UX of BCI games (or applications in general) independent of their performance and investigating the unique capabilities of BCI. The UX evaluation results demonstrate the ways the challenge, cognitive involvement and novelty offered by BCI can influence the UX of a game. Our discussion on the preferable (and non-preferable) measures and methods for evaluating the UX of BCI games provides guidelines to other researchers. Furthermore, the BCI and physiological computing (PC) frameworks we proposed allows developers to situate their applications among other BCI or PC systems.

SIKS DISSERTATION SERIES

2009

2009-01 Rasa Jurgelenaite (RUN), *Symmetric Causal Independence Models*

2009-02 Willem Robert van Hage (VU), *Evaluating Ontology-Alignment Techniques*

2009-03 Hans Stol (UvT), *A Framework for Evidence-based Policy Making Using IT*

2009-04 Josephine Nabukenya (RUN), *Improving the Quality of Organisational Policy Making using Collaboration Engineering*

2009-05 Sietse Overbeek (RUN), *Bridging Supply and Demand for Knowledge Intensive Tasks - Based on Knowledge, Cognition, and Quality*

2009-06 Muhammad Subianto (UU), *Understanding Classification*

2009-07 Ronald Poppe (UT), *Discriminative Vision-Based Recovery and Recognition of Human Motion*

2009-08 Volker Nannen (VU), *Evolutionary Agent-Based Policy Analysis in Dynamic Environments*

2009-09 Benjamin Kanagwa (RUN), *Design, Discovery and Construction of Service-oriented Systems*

2009-10 Jan Wielemaker (UVA), *Logic programming for knowledge-intensive interactive applications*

2009-11 Alexander Boer (UVA), *Legal Theory, Sources of Law & the Semantic Web*

2009-12 Peter Massuthe (TUE, HU Berlin), *Operating Guidelines for Services*

2009-13 Steven de Jong (UM), *Fairness in Multi-Agent Systems*

2009-14 Maksym Korotkiy (VU), *From ontology-enabled services to service-enabled ontologies (making ontologies work in e-science with ONTO-SOA)*

2009-15 Rinke Hoekstra (UVA), *Ontology Representation - Design Patterns and Ontologies that Make Sense*

2009-16 Fritz Reul (UvT), *New Architectures in Computer Chess*

2009-17 Laurens van der Maaten (UvT), *Feature Extraction from Visual Data*

2009-18 Fabian Groffen (CWI), *Armada, An Evolving Database System*

2009-19 Valentin Robu (CWI), *Modeling Preferences, Strategic Reasoning and Collaboration in Agent-Mediated Electronic Markets*

2009-20 Bob van der Vecht (UU), *Adjustable Autonomy: Controlling Influences on Decision Making*

2009-21 Stijn Vanderlooy (UM), *Ranking and Reliable Classification*

2009-22 Pavel Serdyukov (UT), *Search For Expertise: Going beyond direct evidence*

2009-23 Peter Hofgesang (VU), *Modelling Web Usage in a Changing Environment*

2009-24 Annerieke Heuvelink (VUA), *Cognitive Models for Training Simulations*

2009-25 Alex van Ballegooij (CWI), *"RAM: Array Database Management through Relational Mapping"*

2009-26 Fernando Koch (UU), *An Agent-Based Model for the Development of Intelligent Mobile Services*

2009-27 Christian Glahn (OU), *Contextual Support of social Engagement and Reflection on the Web*

2009-28 Sander Evers (UT), *Sensor Data Management with Probabilistic Models*

2009-29 Stanislav Pokraev (UT), *Model-Driven Semantic Integration of Service-Oriented Applications*

2009-30 Marcin Zukowski (CWI), *Balancing vectorized query execution with bandwidth-optimized storage*

2009-31 Sofiya Katrenko (UVA), *A Closer Look at Learning Relations from Text*

2009-32 Rik Farenhorst (VU) and Remco de Boer (VU), *Architectural Knowledge Management: Supporting Architects and Auditors*

2009-33 Khiet Truong (UT), *How Does Real Affect Affect Recognition In Speech?*

2009-34 Inge van de Weerd (UU), *Advancing in Software Product Management: An Incremental Method Engineering Approach*

2009-35 Wouter Koelewijn (UL), *Privacy en Politiegegevens; Over geautomatiseerde normatieve informatie-uitwisseling*

2009-36 Marco Kalz (OUN), *Placement Support for Learners in Learning Networks*

2009-37 Hendrik Drachslers (OUN), *Navigation Support for Learners in Informal Learning Networks*

2009-38 Riina Vuorikari (OU), *Tags and self-organisation: a metadata ecology for learning resources in a multilingual context*

2009-39 Christian Stahl (TUE, HU Berlin), *Service Substitution – A Behavioral Approach Based on Petri Nets*

2009-40 Stephan Raaijmakers (UvT), *Multinomial Language Learning: Investigations into the Geometry of Language*

2009-41 Igor Bereznyy (UvT), *Digital Analysis of Paintings*

2009-42 Toine Bogers (UvT), *Recommender Systems for Social Bookmarking*

2009-43 Virginia Nunes Leal Franqueira (UT), *Finding Multi-step Attacks in Computer Networks using Heuristic Search and Mobile Ambients*

2009-44 Roberto Santana Tapia (UT), *Assessing Business-IT Alignment in Networked Organizations*

2009-45 Jilles Vreeken (UU), *Making Pattern Mining Useful*

2009-46 Loredana Afanasiev (UvA), *Querying XML: Benchmarks and Recursion*

2010

2010-01 Matthijs van Leeuwen (UU), *Patterns that Matter*

- 2010-02** Ingo Wassink (UT), *Work flows in Life Science*
- 2010-03** Joost Geurts (CWI), *A Document Engineering Model and Processing Framework for Multimedia documents*
- 2010-04** Olga Kulyk (UT), *Do You Know What I Know? Situational Awareness of Co-located Teams in Multidisplay Environments*
- 2010-05** Claudia Hauff (UT), *Predicting the Effectiveness of Queries and Retrieval Systems*
- 2010-06** Sander Bakkes (UvT), *Rapid Adaptation of Video Game AI*
- 2010-07** Wim Fikkert (UT), *Gesture interaction at a Distance*
- 2010-08** Krzysztof Siewicz (UL), *Towards an Improved Regulatory Framework of Free Software. Protecting user freedoms in a world of software communities and eGovernments*
- 2010-09** Hugo Kielman (UL), *A Politiele gegevensverwerking en Privacy, Naar een effectieve waarborging*
- 2010-10** Rebecca Ong (UL), *Mobile Communication and Protection of Children*
- 2010-11** Adriaan Ter Mors (TUD), *The world according to MARP: Multi-Agent Route Planning*
- 2010-12** Susan van den Braak (UU), *Sensemaking software for crime analysis*
- 2010-13** Gianluigi Folino (RUN), *High Performance Data Mining using Bio-inspired techniques*
- 2010-14** Sander van Splunter (VU), *Automated Web Service Reconfiguration*
- 2010-15** Lianne Bodestaff (UT), *Managing Dependency Relations in Inter-Organizational Models*
- 2010-16** Sicco Verwer (TUD), *Efficient Identification of Timed Automata, theory and practice*
- 2010-17** Spyros Kotoulas (VU), *Scalable Discovery of Networked Resources: Algorithms, Infrastructure, Applications*
- 2010-18** Charlotte Gerritsen (VU), *Caught in the Act: Investigating Crime by Agent-Based Simulation*
- 2010-19** Henriette Cramer (UvA), *People's Responses to Autonomous and Adaptive Systems*
- 2010-20** Ivo Swartjes (UT), *Whose Story Is It Anyway? How Improv Informs Agency and Authorship of Emergent Narrative*
- 2010-21** Harold van Heerde (UT), *Privacy-aware data management by means of data degradation*
- 2010-22** Michiel Hildebrand (CWI), *End-user Support for Access to Heterogeneous Linked Data*
- 2010-23** Bas Steunebrink (UU), *The Logical Structure of Emotions*
- 2010-24** Dmytro Tykhonov (TUD), *Designing Generic and Efficient Negotiation Strategies*
- 2010-25** Zulfiqar Ali Memon (VU), *Modelling Human-Awareness for Ambient Agents: A Human Mindreading Perspective*
- 2010-26** Ying Zhang (CWI), *XRPC: Efficient Distributed Query Processing on Heterogeneous XQuery Engines*
- 2010-27** Marten Voulon (UL), *Automatisch contracteren*
- 2010-28** Arne Koopman (UU), *Characteristic Relational Patterns*
- 2010-29** Stratos Idreos(CWI), *Database Cracking: Towards Auto-tuning Database Kernels*
- 2010-30** Marieke van Erp (UvT), *Accessing Natural History - Discoveries in data cleaning, structuring, and retrieval*
- 2010-31** Victor de Boer (UVA), *Ontology Enrichment from Heterogeneous Sources on the Web*
- 2010-32** Marcel Hiel (UvT), *An Adaptive Service Oriented Architecture: Automatically solving Interoperability Problems*
- 2010-33** Robin Aly (UT), *Modeling Representation Uncertainty in Concept-Based Multimedia Retrieval*
- 2010-34** Teduh Dirgahayu (UT), *Interaction Design in Service Compositions*
- 2010-35** Dolf Trieschnigg (UT), *Proof of Concept: Concept-based Biomedical Information Retrieval*
- 2010-36** Jose Janssen (OU), *Paving the Way for Lifelong Learning: Facilitating competence development through a learning path specification*
- 2010-37** Niels Lohmann (TUE), *Correctness of services and their composition*
- 2010-38** Dirk Fahland (TUE), *From Scenarios to components*
- 2010-39** Ghazanfar Farooq Siddiqui (VU), *Integrative modeling of emotions in virtual agents*
- 2010-40** Mark van Assem (VU), *Converting and Integrating Vocabularies for the Semantic Web*
- 2010-41** Guillaume Chaslot (UM), *Monte-Carlo Tree Search*
- 2010-42** Sybren de Kinderen (VU), *Needs-driven service bundling in a multi-supplier setting - the computational e3-service approach*
- 2010-43** Peter van Kranenburg (UU), *A Computational Approach to Content-Based Retrieval of Folk Song Melodies*
- 2010-44** Pieter Bellekens (TUE), *An Approach towards Context-sensitive and User-adapted Access to Heterogeneous Data Sources, Illustrated in the Television Domain*
- 2010-45** Vasilios Andrikopoulos (UvT), *A theory and model for the evolution of software services*
- 2010-46** Vincent Pijpers (VU), *e3alignment: Exploring Inter-Organizational Business-ICT Alignment*
- 2010-47** Chen Li (UT), *Mining Process Model Variants: Challenges, Techniques, Examples*
- 2010-48** Milan Lovric (EUR), *Behavioral Finance and Agent-Based Artificial Markets*
- 2010-49** Jahn-Takeshi Saito (UM), *Solving difficult game positions*
- 2010-50** Bouke Huurnink (UVA), *Search in Audio-visual Broadcast Archives*
- 2010-51** Alia Khairia Amin (CWI), *Understanding and supporting information seeking tasks in multiple sources*
- 2010-52** Peter-Paul van Maanen (VU), *Adaptive Support for Human-Computer Teams: Exploring the Use of Cognitive Models of Trust and Attention*
- 2010-53** Edgar Meij (UVA), *Combining Concepts and Language Models for Information Access*
- 2011**
- 2011-01** Botond Cseke (RUN), *Variational Algorithms for Bayesian Inference in Latent Gaussian Models*
- 2011-02** Nick Tinnemeier(UU), *Organizing Agent Organizations. Syntax and Operational Semantics of an Organization-Oriented Programming Language*

- 2011-03** Jan Martijn van der Werf (TUE), *Compositional Design and Verification of Component-Based Information Systems*
- 2011-04** Hado van Hasselt (UU), *Insights in Reinforcement Learning; Formal analysis and empirical evaluation of temporal-difference learning algorithms*
- 2011-05** Base van der Raadt (VU), *Enterprise Architecture Coming of Age - Increasing the Performance of an Emerging Discipline*
- 2011-06** Yiwen Wang (TUE), *Semantically-Enhanced Recommendations in Cultural Heritage*
- 2011-07** Yujia Cao (UT), *Multimodal Information Presentation for High Load Human Computer Interaction*
- 2011-08** Nieske Vergunst (UU), *BDI-based Generation of Robust Task-Oriented Dialogues*
- 2011-09** Tim de Jong (OU), *Contextualised Mobile Media for Learning*
- 2011-10** Bart Bogaert (UvT), *Cloud Content Contention*
- 2011-11** Dhaval Vyas (UT), *Designing for Awareness: An Experience-focused HCI Perspective*
- 2011-12** Carmen Bratosin (TUE), *Grid Architecture for Distributed Process Mining*
- 2011-13** Xiaoyu Mao (UvT), *Airport under Control. Multiagent Scheduling for Airport Ground Handling*
- 2011-14** Milan Lovric (EUR), *Behavioral Finance and Agent-Based Artificial Markets*
- 2011-15** Marijn Koolen (UvA), *The Meaning of Structure: the Value of Link Evidence for Information Retrieval*
- 2011-16** Maarten Schadd (UM), *Selective Search in Games of Different Complexity*
- 2011-17** Jiyin He (UVA), *Exploring Topic Structure: Coherence, Diversity and Relatedness*
- 2011-18** Mark Ponsen (UM), *Strategic Decision-Making in complex games*
- 2011-19** Ellen Rusman (OU), *The Mind 's Eye on Personal Profiles*
- 2011-20** Qing Gu (VU), *Guiding service-oriented software engineering - A view-based approach*
- 2011-21** Linda Terlouw (TUD), *Modularization and Specification of Service-Oriented Systems*
- 2011-22** Junte Zhang (UVA), *System Evaluation of Archival Description and Access*
- 2011-23** Wouter Weerkamp (UVA), *Finding People and their Utterances in Social Media*
- 2011-24** Herwin van Welbergen (UT), *Behavior Generation for Interpersonal Coordination with Virtual Humans: On Specifying, Scheduling and Realizing Multimodal Virtual Human Behavior*
- 2011-25** Syed Waqar ul Qounain Jaffry (VU), *Analysis and Validation of Models for Trust Dynamics*
- 2011-26** Matthijs Aart Pontier (VU), *Virtual Agents for Human Communication: Emotion Regulation and Involvement-Distance Trade-Offs in Embodied Conversational Agents and Robots*
- 2011-27** Aniel Bhulai (VU), *Dynamic website optimization through autonomous management of design patterns*
- 2011-28** Rianne Kaptein (UVA), *Effective Focused Retrieval by Exploiting Query Context and Document Structure*
- 2011-29** Faisal Kamiran (TUE), *Discrimination-aware Classification*
- 2011-30** Egon van den Broek (UT), *Affective Signal Processing (ASP): Unraveling the mystery of emotions*
- 2011-31** Ludo Waltman (EUR), *Computational and Game-Theoretic Approaches for Modeling Bounded Rationality*
- 2011-32** Nees-Jan van Eck (EUR), *Methodological Advances in Bibliometric Mapping of Science*
- 2011-33** Tom van der Weide (UU), *Arguing to Motivate Decisions*
- 2011-34** Paolo Turrini (UU), *Strategic Reasoning in Interdependence: Logical and Game-theoretical Investigations*
- 2011-35** Maaïke Harbers (UU), *Explaining Agent Behavior in Virtual Training*
- 2011-36** Erik van der Spek (UU), *Experiments in serious game design: a cognitive approach*
- 2011-37** Adriana Burlutiu (RUN), *Machine Learning for Pairwise Data, Applications for Preference Learning and Supervised Network Inference*
- 2011-38** Nyree Lemmens (UM), *Bee-inspired Distributed Optimization*
- 2011-39** Joost Westra (UU), *Organizing Adaptation using Agents in Serious Games*
- 2011-40** Viktor Clerc (VU), *Architectural Knowledge Management in Global Software Development*
- 2011-41** Luan Ibraimi (UT), *Cryptographically Enforced Distributed Data Access Control*
- 2011-42** Michal Sindlar (UU), *Explaining Behavior through Mental State Attribution*
- 2011-43** Henk van der Schuur (UU), *Process Improvement through Software Operation Knowledge*
- 2011-44** Boris Reuderink (UT), *Robust Brain-Computer Interfaces*
- 2011-45** Herman Stehouwer (UvT), *Statistical Language Models for Alternative Sequence Selection*
- 2011-46** Beibei Hu (TUD), *Towards Contextualized Information Delivery: A Rule-based Architecture for the Domain of Mobile Police Work*
- 2011-47** Azizi Bin Ab Aziz (VU), *Exploring Computational Models for Intelligent Support of Persons with Depression*
- 2011-48** Mark Ter Maat (UT), *Response Selection and Turn-taking for a Sensitive Artificial Listening Agent*
- 2011-49** Andreea Niculescu (UT), *Conversational interfaces for task-oriented spoken dialogues: design aspects influencing interaction quality*
- 2012**
- 2012-01** Terry Kakeeto (UvT), *Relationship Marketing for SMEs in Uganda*
- 2012-02** Muhammad Umair (VU), *Adaptivity, emotion, and Rationality in Human and Ambient Agent Models*
- 2012-03** Adam Vanya (VU), *Supporting Architecture Evolution by Mining Software Repositories*
- 2012-04** Jurriaan Souer (UU), *Development of Content Management System-based Web Applications*
- 2012-05** Marijn Plomp (UU), *Maturing Interorganizational Information Systems*
- 2012-06** Wolfgang Reinhardt (OU), *Awareness Support for Knowledge Workers in Research Networks*

- 2012-07** Rianne van Lambalgen (VU), *When the Going Gets Tough: Exploring Agent-based Models of Human Performance under Demanding Conditions*
- 2012-08** Gerben de Vries (UVA), *Kernel Methods for Vessel Traject*
- 2012-09** Ricardo Neisse (UT), *Trust and Privacy Management Support for Context-Aware Service Platforms*
- 2012-10** David Smits (TUE), *Towards a Generic Distributed Adaptive Hypermedia Environment*
- 2012-11** J.C.B. Rantham Prabhakara (TUE), *Process Mining in the Large: Preprocessing, Discovery, and Diagnostics*
- 2012-12** Kees van der Sluijs (TUE), *Model Driven Design and Data Integration in Semantic Web Information Systems*
- 2012-13** Suleman Shahid (UvT), *Fun and Face: Exploring non-verbal expressions of emotion during playful interactions*
- 2012-14** Evgeny Knutov (TUE), *Generic Adaptation Framework for Unifying Adaptive Web-based Systems*
- 2012-15** Natalie van der Wal (VU), *Social Agents. Agent-Based Modelling of Integrated Internal and Social Dynamics of Cognitive and Affective Processes*
- 2012-16** Fiemke Both (VU), *Helping people by understanding them - Ambient Agents supporting task execution and depression treatment*
- 2012-17** Amal Elgammal (UvT), *Towards a Comprehensive Framework for Business Process Compliance*
- 2012-18** Eltjo Poort (VU), *Improving Solution Architecting Practices*
- 2012-19** Helen Schonenberg (TUE), *What's Next? Operational Support for Business Process Execution*
- 2012-20** Ali Bahramisharif (RUN), *Covert Visual Spatial Attention, a Robust Paradigm for Brain-Computer Interfacing*
- 2012-21** Roberto Cornacchia (TUD), *Querying Sparse Matrices for Information Retrieval*
- 2012-22** Thijs Vis (UvT), *Intelligence, politie en veiligheidsdienst: verenigbare grootheden?*
- 2012-23** Christian Muehl (UT), *Toward Affective Brain-Computer Interfaces: Exploring the Neurophysiology of Affect during Human Media Interaction*
- 2012-24** Laurens van der Werff (UT), *Evaluation of Noisy Transcripts for Spoken Document Retrieval*
- 2012-25** Silja Eckartz (UT), *Managing the Business Case Development in Inter-Organizational IT Projects: A Methodology and its Application*
- 2012-26** Emile de Maat (UvA), *Making Sense of Legal Texts*