Leveraging human-computer interactions and social presence with physiological computing

Jeremy Frey

THÈSE

Présentée à l'UNIVERSITÉ DE BORDEAUX

École doctorale de Mathématiques et Informatique

Spécialité : INFORMATIQUE

par

Jérémy FREY

Pour obtenir le grade de

Docteur

Leveraging human-computer interactions and social presence with physiological computing

Soutenue le 8 décembre 2015

Jury

Directeur de thèse :	Martin HACHET	Chargé de recherche, Inria
Encadrant de thèse :	Fabien LOTTE	Chargé de recherche, Inria
Président :	Nicolas ROUSSEL	Directeur de recherche, Inria
Rapporteurs :	Stephen H. FAIRCLOUGH	Professor, Liverpool JM University
	Anatole LÉCUYER	Directeur de recherche, Inria
Examinateur :	Thorsten O. ZANDER	Group leader, TU Berlin

université de BORDEAUX

Titre

Améliorer les interactions homme-machine et la présence sociale avec l'informatique physiologique

Résumé

Cette thèse explore comment l'informatique physiologique peut contribuer aux interactions homme-machine (IHM) et encourager l'apparition de nouveaux canaux de communication parmi le grand public. Nous avons examiné comment des capteurs physiologiques, tels que l'électroencéphalographie (EEG), pourraient être utilisés afin d'estimer l'état mental des utilisateurs et comment ils se positionnent par rapport à d'autres méthodes d'évaluation. Nous avons créé la première interface cerveau-ordinateur capable de discriminer le confort visuel pendant le visionnage d'images stéréoscopiques et nous avons esquissé un système qui peux aider à estimer l'expérience utilisateur dans son ensemble, en mesurant charge mentale, attention et reconnaissance d'erreur. Pour abaisser la barrière entre utilisateurs finaux et capteurs physiologiques, nous avons participé à l'intégration logicielle d'un appareil EEG bon marché et libre, nous avons utilisé des webcams du commerce pour mesurer le rythme cardiaque à distance, nous avons confectionné des wearables dont les utilisateurs peuvent rapidement s'équiper afin qu'électrocardiographie, activité électrodermale et EEG puissent être mesurées lors de manifestations publiques. Nous avons imaginé de nouveaux usages pour nos capteurs, qui augmenteraient la présence sociale. Dans une étude autour de l'interaction humain-agent, les participants avaient tendance à préférer les avatars virtuels répliquant leurs propres états internes. Une étude ultérieure s'est concentrée sur l'interaction entre utilisateurs, profitant d'un jeu de plateau pour décrire comment l'examen de la physiologie pourrait changer nos rapports. Des avancées en IHM ont permis d'intégrer de manière transparente du biofeedback au monde physique. Nous avons développé Teegi, une poupée qui permet aux novices d'en découvrir plus sur leur activité cérébrale, par eux-mêmes. Enfin avec Tobe, un toolkit qui comprend plus de capteurs et donne plus de liberté quant à leurs visualisations, nous avons exploré comment un tel proxy décale nos représentations, tant de nous-mêmes que des autres.

Mots clefs

informatique physiologique ; interaction homme-machine ; présence sociale ; électroencéphalographie ; expérience utilisateur

Title

Leveraging human-computer interactions and social presence with physiological computing

Abstract

This thesis explores how physiological computing can contribute to human-computer interaction (HCI) and foster new communication channels among the general public. We investigated how physiological sensors, such as electroencephalography (EEG), could be employed to assess the mental state of the users and how they relate to other evaluation methods. We created the first brain-computer interface that could sense visual comfort during the viewing of stereoscopic images and shaped a framework that could help to assess the overall user experience by monitoring workload, attention and error recognition. To lower the barrier between end users and physiological sensors, we participated in the software integration of a low-cost and open-hardware EEG device; used off-the shelf webcams to measure heart rate remotely, crafted wearables that can quickly equip users so that electrocardiography, electrodermal activity or EEG may be measured during public exhibitions. We envisioned new usages for our sensors, that would increase social presence. In a study about human-agent interaction, participants tended to prefer virtual avatars that were mirroring their own internal state. A follow-up study focused on interactions between users to describe how physiological monitoring could alter our relationships. Advances in HCI enabled us to seamlessly integrate biofeedback to the physical world. We developped Teegi, a puppet that lets novices discover by themselves about their brain activity. Finally, with Tobe, a toolkit that encompasses more sensors and give more freedom about their visualizations, we explored how such proxy shifts our representations, about our selves as well as about the others.

Keywords

physiological computing; human-computer interaction; social presence; electroencephalography; user experience

Unités de recherche

Université de Bordeaux – LaBRI
351 cours de la Libération
33405 Talence, France

Inria Bordeaux Sud-Ouest – Potioc
200 Avenue de la Vieille Tour
33405 Talence, France

Abstract

This thesis explores how physiological computing can contribute to human-computer interaction (HCI) and foster new communication channels among the general public. Our work, which studied physiological sensors at large and electroencephalography (EEG) in particular, covers four aspects.

First, we investigated how those measurements could be employed to assess the mental state of the users and how they relate to other evaluation methods, such as inquiries and behavioral measures. Our review directly led to practical applications, consisting in the continuous evaluation of HCI. We created the first Brain-Computer Interface (BCI) that could discriminate visual comfort from visual discomfort during the viewing of stereoscopic images. In another study, we found evidences that workload could be assessed continuously with EEG during 3D manipulation tasks. When we extended on this work to monitor attention and error recognition besides workload, we compared different interaction techniques through the use of a serious game, shaping a framework that could help to assess the overall user experience.

These various applications necessitate components that are still too difficult to use in the field. We made several technical contributions that permit lowering the barrier between end users and physiological sensors. We participated in the software integration of a low-cost and open-hardware EEG device, OpenBCI. We implemented a signal processing pipeline that uses off-the shelf webcams to measure heart rate remotely, by the mean of photoplethysmography (PPG). We crafted wearables that can quickly equip users so that electrocardiography, electrodermal activity or electroencephalography may be measured during public exhibitions. Those various developments helped us to envision new usages for our sensors beside HCI evaluation.

Indeed, thanks to the steady dissemination in everyday life of devices that sense physiological signals, an additional communication channel for people may come to exist. So as to ground those insights, we elaborated two scenarios involving the most common measure, heart rate. In a study about human-agent interaction, participants tended to prefer virtual avatars that were mirroring their own internal state, meaning that the social presence of artificial beings could be reinforced with little effort. A follow-up study focused on interactions between users, taking advantage of a board game to describe how physiological monitoring could alter our relationships.

This last study profited from advances in HCI to mask computers from the participants, integrating seamlessly digital content – for instance biofeedback – to the physical world. The combination of spatial augmented reality and tangible interface enables the emergence of hybrid objects that can convey a high level of information while

maintaining ease of use and attractiveness. Those tools helped to build a "tangible EEG interface", Teegi. Teegi takes the appearance of a puppet and lets novices discover by themselves about their brain activity. The prototype in our laboratory became a mobile installation that we brought to public festivals. Finally, we pushed forward this project to encompass more physiological sensors and give more freedom about their visualizations, the latter spanning across several layers of abstraction. Using a co-design approach that took place in a museum, we explored how such proxy shifts our representations, about our selves as well as about the others.

Four pictures of a 3-years journey during which I came to think that the purpose of computer science and human-computer interaction is to enhance well-being and facilitate human relationships on the whole – or at least this is the path into which I tried to venture, and this work represents one more step toward this goal I hope. Beyond the possibility for computers to comprehend some of our internal and mental states, physiological computing unveils another mean to mediate oneself, raising our social awareness.

Résumé

Cette thèse explore comment l'informatique physiologique peut contribuer aux interactions homme-machine (IHM) et encourager l'apparition de nouveaux canaux de communication au sein du grand public. Notre travail, qui étudie les capteurs physiologiques en général et l'électroencéphalographie (EEG) en particulier, couvre quatre aspects.

Premièrement, nous avons examiné comment ces mesures pourraient être utilisées afin d'estimer l'état mental des utilisateurs et comment elles se positionnent par rapport à d'autres méthodes d'évaluation, telles que les enquêtes ou les mesures comportementales. Notre revue a directement mené à des applications pratiques, consistant en l'évaluation continue de IHM. Nous avons créé la première interface cerveau-ordinateur (BCI) qui est capable de discriminer le confort visuel de l'inconfort visuel pendant le visionnage d'images stéréoscopiques. Dans une autre étude, nous avons mis en évidence que la charge mentale pouvait être estimée de manière continue avec l'EEG pendant des tâches de manipulation 3D. Lorsque nous avons étendu ce travail pour surveiller attention et reconnaissance d'erreur en plus de la charge mentale, nous avons comparé différentes techniques d'interaction via l'utilisation d'un *serious game*, esquissant un système qui puisse aider à estimer l'expérience utilisateur dans son ensemble.

Ces diverses applications requièrent des composants qui sont encore trop difficiles à utiliser sur le terrain. Nous avons fait plusieurs contributions techniques qui ont permis d'abaisser la barrière entre les utilisateurs finaux et les capteurs physiologiques. Nous avons participé à l'intégration logicielle d'un appareil EEG bon marché et libre, OpenBCI. Nous avons implémenté une suite de traitement de signal qui utilise des webcams du commerce pour mesurer le rythme cardiaque à distance, via photoplethysmographie (PPG). Nous avons confectionné des *wearables* dont les utilisateurs peuvent rapidement s'équiper afin qu'électrocardiographie, activité électrodermale et électroencéphalographie puissent être mesurées lors de manifestations publiques. Ces divers développements nous ont aidé à envisager de nouveaux usages pour nos capteurs à côté de l'évaluation de IHM.

En effet, grâce à la dissémination régulière dans la vie quotidienne d'appareils qui captent les signaux physiologiques, un canal de communication supplémentaire entre les personnes pourrait voir le jour. Dans le but d'ancrer ces idées, nous avons élaboré deux scénarios impliquant la mesure la plus courante, le rythme cardiaque. Dans une étude autour de l'interaction humain-agent, les participants avaient tendance à préférer les avatars virtuels qui répliquaient leur propre état interne, ce qui signifie que la présence sociale d'êtres artificiels pourrait être renforcée avec peu d'effort. Une étude ultérieure s'est concentrée sur l'interaction entre utilisateurs, profitant d'un jeu de plateau pour

décrire comment l'examen de la physiologie pourrait changer nos rapports.

Cette dernière étude a bénéficié de progrès en IHM pour masquer les ordinateurs aux yeux des participants, ceci en intégrant de manière transparente du contenu numérique – en l'occurrence le *biofeedback* – au monde physique. La combinaison de la réalité augmentée spatiale et d'interfaces tangibles permet l'émergence d'objets hybrides qui peuvent transmettre un haut niveau d'information tout en préservant facilité d'utilisation et attrait. Ces outils ont aidé à construire une « interface tangible pour l'EEG », Teegi. Teegi prend la forme d'une poupée et permet aux novices d'en découvrir plus sur leur activité cérébrale, par eux-mêmes. Le prototype dans notre laboratoire est devenu une installation mobile que nous avons amenée dans des festivals publiques. Enfin, nous avons poussé en avant ce projet pour intégrer plus de capteurs physiologiques et donner plus de liberté quant à leurs visualisations, ces dernières couvrant plusieurs couches d'abstraction. En utilisant une approche de co-design qui a pris place dans un musée, nous avons exploré comment un tel proxy décale nos représentations, tant de nous-mêmes que des autres.

Quatre tableaux d'un voyage de trois ans pendant lequel j'en suis venu à penser que le but de l'informatique et de l'interaction homme-machine est d'améliorer le bien-être et de faciliter les relations humaines dans leur ensemble – ou du moins est-ce la voie dans laquelle j'ai essayé de m'aventurer, et ce travail représente un pas de plus dans cette direction je l'espère. Par-delà la possibilité pour les ordinateurs de comprendre certains de nos états internes et mentaux, l'informatique physiologique dévoile un autre moyen de servir de médiateur avec soi-même, élevant notre conscience sociale.

Acknowledgements

A thesis is not the product of a sole person, and this one is no exception. I will start by pointing at the two key persons who made this happen. Martin Hachet and Fabien Lotte combined their expertise and found a research direction which deserved to be studied thoroughly. They discovered me along the road, offered me to hop in and to start this 3 years journey. At a time I was unsure of my future, who knows what I would have become if I had not crossed their path. Before being scientists, Martin and Fabien are good persons, kind, welcoming, open-minded, always eager to listen and to help. While they guided me and channelled my willpower into something (hopefully) valuable, they also let me explore by myself, acting as shields against a world ruled by rigor for the several times I was being too silly – not an easy task, as they would tell you! I am most grateful for the freedom they gave me, a rare treasure that pushed me forward.

Facing the library of life, where each book is a different adventure, it is often difficult to open the right work, even when one strives to enroll in a PhD. Beside mentoring, advisers provide for life changing-opportunities. This statement encompasses those who came by before my thesis; it was under the supervision of André Garenne, during my master degree, that I discovered my calling. Should Research regrets to have me on board, he will be to blame in the first place.

L'argent est le nerf de la guerre, as French people say. Evolving in a world of pure ideas is good, having the means to spread and embody them is better. Bordeaux university, Inria, LaBRI, CNRS, URFIST; different institutions which all gave me access to diverse formations which help the PhD student I was to grow up. Through my scholarship and my teaching activities I was employed by the university, but it was in Inria, among the project-team Potioc that I lived. It was a wonder to be part of this institute, to integrate an environment with a complete and dedicated supporting staff. I was able to focus on the good aspects of my work. And one of the good aspect was the many occasions I had to go abroad. Conferences, summer schools, exhibitions: I've literally travel around the world, to disseminate science, to meet with people. Not the least, the working environment – the building, the open-space where I caused some noise, the equipment at disposal – was just ideal. For my scientific as well as my personal enrichment, I am thankful for that. Finally, as a member of the LaBRI, I had in my possessing during 3 years a proper machine – may be obvious, but in this profession having a reliable computer is appreciated. The LaBRI also provided me with a scholarship so I could visit for two months the MuSAE Lab in Montreal, a laboratory directed by Tiago H. Falk. This event occurred before my defense and I was supposed to wrap up the first version of my manuscript before I left. Of course it did not go as planned, I was late, *weeks* late. Fortunately Tiago was very comprehensive and let me focus on my redaction; one of many providential encounters.

Most of all, I was lucky enough to join an amazing team. For one reason or another, I admired each one of the member of Potioc that I met during my stay. Could be because of their knowledge, their skills, their personality, or, more often than not, a combination of those different factors. I may like to behave a bit rogue, for example eating by my own instead of joining the rest of the group (silly them to go for lunch before 1:30 PM!), but deep in my heart I wish I would have taken the time to get to know my comrades better. When I got truly lucky, I had the occasion to closely work with some of them. I never felt so eager to finish a project than when my brain was only a small part of the process. Really thriving! If I believe my notes and my timestamps, at some point during the birth of Teegi, I've spent 90+ hours in a sole week in the lab. I don't want to brag about it – it was more foolish than anything else –, solely claim that I do not regret a minute of this time. I was just happy to move forward as a team, a revelation in the end: "This is it, not only is this what I want to do, this is what I *like*." Few names among many: Renaud Gervais (**of course**), Jérémy Laviole, with whom I shared more than it looks (like a journey across India), Thomas Hulin, who I've been chatting with about various subjects while he was sitting next to me in my corner of the open-space, Christian Mühl, who gave me valuable advice to write at my best from the start, Thibault Laine, who never refused to share his skills and dedicate his time to some obscure projects of mine, Camille Jeunet, for the hints about the BCI equipment we shared and for her constant cheerfulness, Catherine Mégrat, always supportive and who printed the only physical copy of this manuscript to date, Anke Brock, who sent me a fair amount of tips and links for the PhD afterlife. There were also *many* interns, Léonard Pommereau, Aurélien Appriou, Dennis Wobrock (cognitive science rulez), Maxime Daniel, Maxime Duluc and Alexis Gay; but I will get back to all of them later, in due time, in a proper section – I'd wish to high-five more Potiocs, but I will stay within the boundaries of this PhD, even if it's a struggle to do so. People that I will silence afterwards even though I owe them all my results: the participants that were patient and kind enough to enroll in my experiments, even when sessions would last for hours, with no reward but gel in the hair. If there was no such things as anonymity, know that as a tribute I would put all your names right *here*.

As for friends and family… It seems to be almost mandatory to write out loud how close ones were an invaluable support during a thesis, that it was a particularly difficult time that they endured stoically. Well, in my case I am afraid there was not really a difference: I had been as difficult as usual to bear. My family and especially my parents have always been there for me – providing me with food in my busiest weeks! – and year after year I wonder how my friends still have the patience to get back to me and lighten my life, even when I hardly give news, even if I am stupid enough to silence what I feel, to crack yet another joke during a serious discussion when I could simply tell that I love them. And I do. Be reassured, it will not change anytime soon, they barely read English. This long acknowledgement section is the only reason why I avoided writing in French, no "let's spread my magnificent work around the world" involved, just keeping my true feelings hidden.

Which brings me to you, reader. I do not know why you endured this section – or why you would read this manuscript at all when the world is full of inspiring novels – but while you are here I wish you a pleasant reading experience. And, just between the two of us, do not hesitate to give feedback once you are done, good or bad. Que la fête commence !

Contents

List of abbreviations

- *BCI*: Brain-Computer Interface
- *CSP*: Common Spatial Patterns
- *CSSP*: Common Spatio-Spectral Patterns
- *CV*: Cross-Validation
- *DOF*: Degrees Of Freedom
- *ECG*: ElectroCardioGraphy
- *EDA*: ElectroDermal Activity
- *EEG*: ElectroEncephaloGraphy
- *EMG*: ElectroMyoGraphy
- *EOG*: ElectroOculoGraphy
- *ERP*: Event-Related Potentials
- *ERSP*: Event-Related Spectral Perturbations
- *ErrP*: Error-related Potentials
- *IEQ*: Immersive Experience Questionnaire
- *fNIRS*: functional Near-Infrared Spectroscopy
- *fMRI*: functional Magnetic Resonance Imaging
- *GEQ*: Game Experience Questionnaire
- *GSR*: Galvanic Skin Response (deprecated term for EDA)
- *HCI*: Human-Computer Interaction

- *LDA*: Linear Discriminant Analysis
- *MEG*: MagnetoEncephaloGraphy
- *PPG*: PhotoPlethysmoGraphy
- *REFSF*: Regularized Eigen Fisher Spatial Filters
- *SAM*: Self-Assessment Manikin
- *SAR*: Spatial Augmented Reality
- *sLDA*: shrinkage Linear Discriminant Analysis
- *SPGQ*: Social Presence in Gaming Questionnaire
- *SSCSP*: Stationary Subspace Common Spatial Patterns
- *TUI*: Tangible User Interface
- *UI*: User Interface
- *UX*: User eXperience
- *VR*: Virtual Reality

Introduction

(Mise-en-scène)

You're looking at me, but I hardly look back at you. I'm mostly starring at my screen. You wonder what's going on with me. Are you disturbing me while I'm in the midst of a deadline crisis, with a couple hours left and tons of gibberish data to analyze with über complicated algorithms? Did I even notice you were here? Or am I just faking it, typing randomly so I could pretend I have better things to do than answering to your questions?

You knew this morning you should have been more tactful when you phrased your critics regarding my lengthy e-mails, which obfuscate simple ideas behind convoluted and never-ending sentences. You heard before that I'm touchy on the subject, but you did not mean to be rude. Even if I did not see that back then, you just wanted to help, avoid me facing future disappointments.

I'm mumbling from time to time; I cannot be such a good actor. Maybe I decided to venture once again into this 3D modeller, powerful and free and all, but a pain to use. Hard to tell.

Not sure how to react, you sigh and leave the room. There is this bulky manuscript that awaits on your desk. No escape to that.

This is one staging, fictional only in part. There may be other characters or a different setting involved, yet situations do arise where one is short on clues about those around, about what they endure or what they meant.

I have not the room to write a fantasy novel about how some sort of technological voodoo could create an artifact that fosters empathy and help to bring harmony. I will limit myself to tell how it all started.

Context and motivations

Human-computer interaction

At the Beginning was the Bit. The "computer era" started with simple computations, automatic systems aimed at easing mathematical operations. First machines were no more than abacuses (merely) moving by themselves. One played with switches to input data and obtain in return clicks, flickering light bulbs and some tape. A time mastered by engineers and scientists, fortune-tellers reading holes in cardboard. They were building the system, they knew how to handle it. Then came Progress: computing power increased, new devices appeared as abstraction layers between humans and machines, machines the size of a room shrank to personal computers and, at last, people outside the field started to use those computers. Concurrently, during the eighties, a new discipline appeared: human-computer interaction (HCI).

HCI is concerned with the design and evaluation of such interactive systems [Hewett et al., 1992]. HCI roots in transverse domains – such as psychology or sociology – and shifts from a vision centered around technology to a vision centered around the user [Roussel, 2014, Roussel, 2002]. Indeed, with the increasing complexity of computers, it became a necessity to make the machines comply with human capabilities, *practical* to use. Nowadays we have smartphones, smart TV, smartwatches; computers are literally everywhere, used by everyone. Interactions between human and computers are richer than ever, yet finding the right tools to evaluate the overall user experience and ensure convenience is still an open research question.

Among the variety of evaluation methods, inquiry-based approaches – e.g. questionnaires, think aloud protocol – and behavioral measures – e.g. reaction time, error rate – have been used successfully for decades. They suffer however from a number of limitations. Inquiries are prone to be contaminated by ambiguities [Nisbett and Wilson, 1977] or may be affected by social pressure [Picard, 1995]. It is also very difficult to get continuous insights without disrupting the interaction. Think aloud protocol distracts users and questionnaires can be given only at specific time points, usually at the end of a session – which leads to a bias due to participants' memory limitations [Kivikangas et al., 2010]. Metrics inferred from behavioral measures, on the other hand, can be computed in real time, but they do not provide much information about users' states. For example, a higher reaction time can be induced either by a lower attention level or by a higher cognitive workload [Berka and Levendowski, 2007, Hart and Staveland, 1988].

As such, the conception of new interfaces and interaction techniques could greatly benefit from a framework that would combine continuous recordings and qualitative measures.

Physiological computing

Since a few years, physiological sensors have been studied to improve the ergonomics of HCI [Fairclough, 2009]. "Physiological computing" is the term that coins the use of physiological data to gain a real-time feedback about users' inner state. Physiological sensors cover bodily activity at large, such as heart rate measures or electrodermal activity (EDA, i.e. skin's perspiration). With the proper signal processing, various mental states can be measured, for example emotions in [Villon and Lisetti, 2006] or workload in [Fairclough and Houston, 2004]. Whereas these two works used low-level physiological signals, recently higher-level signals have been considered to deepen the understanding of users, namely brain activity.

A technique alike functional magnetic resonance imaging (fMRI) can determine which regions of the brain are active when users learn new tools [Kitamura et al., 2003]. It can also sense how brain patterns changes while users interact with a virtual environment [Sjölie et al., 2010]. fMRI studies are however difficult to setup and constraining. The equipment is expensive and users are enclosed in the machine during the experiment, far from natural (ecological) scenarios. Fortunately, more affordable and lightweight devices are spreading. Among them, electroencephalography (EEG) and functional near infrared spectroscopy (fNIRS) are particularity well suited for mobile brain imaging [Mehta and Parasuraman, 2013, Cutrell and Tan, 2008].

More about BCI in the first part.

Thanks to EEG and fNIRS, a new type of interface emerged in the last decade. BCI, brain-computer interfaces, are communication devices between humans and machines that rely only on brain activity – i.e. no muscular input [Wolpaw et al., 2002]. While the first systems dealt mostly with commands intentionally issued by users, like moving a cursor, other kinds of BCI applications appeared. Called *passive* BCI, they do not require conscious thoughts from users and are not used as a voluntary input [Zander and Kothe, 2011]. Passive BCI measure instead users' state when they are engaged in another activity. For instance, they could monitor workload so as to adapt automatically and continuously a haptic feedback [George and Lécuyer, 2010].

Beyond adaptive systems, physiological computing is an opportunity to address the current limitation of HCI evaluation methods and assess the overall usability of a system [Ravaja, 2009, Kivikangas et al., 2010, Pike et al., 2012]. Furthermore, not only has physiological computing the potential to alter how we relate to computers, it may also change how we relate to others. While the "affective computing" movement that started twenty years ago focuses mainly on emotions [Picard, 1995], the range of constructs that could be inferred has broadened since then – e.g. attention level, workload.

As technology mingles more and more between humans, the time has passed when it could impede our communications for the sake of being the only viable alternative to link distant people. Now computers are on the verge to *augment* how we exchange information.

Social presence

The notion of "presence" originates from two fields: telecommunications and virtual reality (VR) [Sallnäs et al., 2000]. In VR, *virtual* presence refers to the feeling of belonging to a virtual (or mediated) environment [Slater et al., 2009]. It is the subjective experience of being in one place, even when one is physically situated in another [Witmer and Singer, 1998]. In telecommunications, *social* presence commonly refers to the definition given in [Short et al., 1976], where it relates to the degree of salience of another person. Social presence takes into account interpersonal relationships, it is the feeling that one is socially present with another person.

Social presence appeared at a time when communicating with distant peers was questioned and compared to face-to-face communications. The underlying assumption is that social presence decreases as the medium employed in telecommunications – and, by extension in computer-mediated communications [Gunawardena, 1995] – is deprived of channels such as audio or video. Two years before he came up with a definition with his colleagues, Short showed that negotiations conducted through a loud-speaking audio link were less successful than in face-to-face condition or during videoconferencing – "closed circuit television" back then. Participants were missing various non-verbal cues: eye gaze, posture, facial expression, … [Short, 1974].

Flame wars on bulletin boards were already cited in papers from the eighties.

Despite instantaneous messages from remote locations or a better efficiency, computers were seen as an issue in regard to societal effects. For example it is longer to reach consensus through instant messaging [Kiesler et al., 1984] and computers-mediated communications favour uninhibited behaviors [Siegel et al., 1986]. Part of the reason why may be due to the novelty of medium. Usages have to settle; one of the principle problem was a lack of a shared etiquette [Kiesler et al., 1984, Siegel et al., 1986, Gunawardena, 1995]. This may no longer hold true for people that use computers and Internet daily.

Social and virtual presence are not identical [Sallnäs et al., 2000] but they share the same pitfall: they are bound to technological artifacts. The former could suffer from the fact that "using a keyboard takes time" [Kiesler et al., 1984], the latter necessitates a deep immersion in the virtual environment, thus it requires that the stimuli make people feel included in and able to interact with the environment [Witmer and Singer, 1998]. Yet, the never-ending race toward the perfect device, the perfect simulation, is not the sole path. It has been advocated that beyond *more* presence is possible to be obtain a *different* sense of presence. In [Roussel, 2007] various example are given, such as how filtering and notifications result in a good trade-off between accessibility and privacy – it gives more control to users, as compared to a continuous video feed mimicking face-to-face communications. Sometimes subtle cues are also preferable than direct and explicit communications: an ambient display of colleagues' space improves group coherence and social presence [Roussel, 2007].

All these different works still rely on existing media and senses: text, audio, video, haptic interfaces, ... Social presence may shift to another level if it were for brand new communication channels to become available. Physiological computing can do just that. Interestingly enough, when Sheridan – one of founders of telepresence – wrote about presence, he compared its "subjectivity" to the one of workload, stating that such mental manifestations are "not so amenable to objective physiological definition and measurement" [Sheridan, 1992]. Not anymore, physiological measures are now able to give access to information that were previously covert, hardly available even to conscious thoughts.

More than that, the potential improvements that physiological data could bring to social presence are not limited to distant relationships. It could alter and benefit "real" face-to-face communications as well [Slovák et al., 2012]. Awareness of others is critical for collaboration [Dourish and Bellotti, 1992]; ultimately physiological computing could improve "connectedness", i.e. exchanges that support and augment social relationships among people [Kuwabara et al., 2002, Rettie, 2003].

Computers have proven to be powerful tools. Automatic systems gave to the modern age a considerable boost in productivity; space constrains are more and more irrelevant thanks to telecommunications; the opportunity to duplicate digital information with little effort helps to spread knowledge; that content could be altered and transformed unleashes creativity. Yet, despite those achievements, computers have still a way to go before they could be used seamlessly and blend into society. The same way we sometimes have difficulties to comprehend one another, we lack proper instruments to assess how we interact with machines. By shedding light on covert mental states as they are occurring, physiological computing comes closer to what we experience and endure. A bit of electronics, a hint of signal processing, so the tools around could better suit us. And, maybe, once we know better, they will help in return to improve on our mutual understanding.

Structure of the thesis

The overarching goal of my various contributions is to frame how physiological computing will benefit the general public; whether it is when passive brain-computer interfaces are used beforehand to make human-computer interaction more intuitive, or when unobtrusive physiological sensors and covert computing enhance social interactions in everyday life.

The work I have done during my thesis is divided in 5 parts.

First, I have thoroughly looked at the literature to sense in which aspects electroencephalography could improve the evaluation of HCI. The objective is to propose a complementary method that will ease the conception of better user interfaces. Chapter 1 defines which evaluation methods are investigated; chapter 2 briefly introduces the terms "egocentric" and "exocentric" to overcome the semantic difficulties that arise from the use of "subjective" and "objective" words; chapter 3 lists the constructs that could be measured with physiological sensors and highlights which one are the most promising for HCI evaluation.

Second, I took part in different works which applied these last findings and used EEG to measure users' state. I start in chapter 4 by describing the principles behind BCI and the overall signal processing pipeline. Then in chapter 5 I show how EEG could be used to measure visual comfort during the viewing of stereoscopic images. Chapter 6 put the focus on workload, how EEG bests other physiological sensors and how it is a promising technique to assesses workload during 3D manipulation tasks. These results are strengthened in chapter 7, where EEG is validated as an HCI evaluation method. Workload, attention level and interaction errors are measured in a carefully crafted virtual environment and different interactions techniques are compared.

Third, I participated in various projects which aimed at facilitating the use of EEG and physiological computing. A tool which is not practical, even if proven effective, will not be used. Popularizing complex technologies is one motto of the team I have been evolving in, so no wonder if this is an aspect I was driven to develop. Because this aspect of my thesis is either more technical or still awaits a proper validation, it does not appear in the main body of the manuscript but as appendices. Appendix A summarizes the conditions an "ideal" EEG device should meet, appendix B compares side-by-side a medical grade equipment to a low-cost and open-source amplifier and appendix C gives an example of how such latter device could be quickly integrated to a wearable. In appendix D another physiological measure is considered, heart-rate, using remote sensing through video feeds. The framework works in real time with low-end webcams. Finally, appendix E discusses how the appearance of artifacts in physiological measures may be prevented by giving a subtle feedback to users.

Fourth, I have started to explore how physiological computing, once made easily available and non obtrusive, can shift social interactions. In chapter 8 I suggest that a "similarity-attraction effect" improves the social presence of embodied agents, simply by mirroring users' heart rate. I also investigate, in chapter 9, how sharing such physiological signal in a common space can enriches the user experience of board game players.

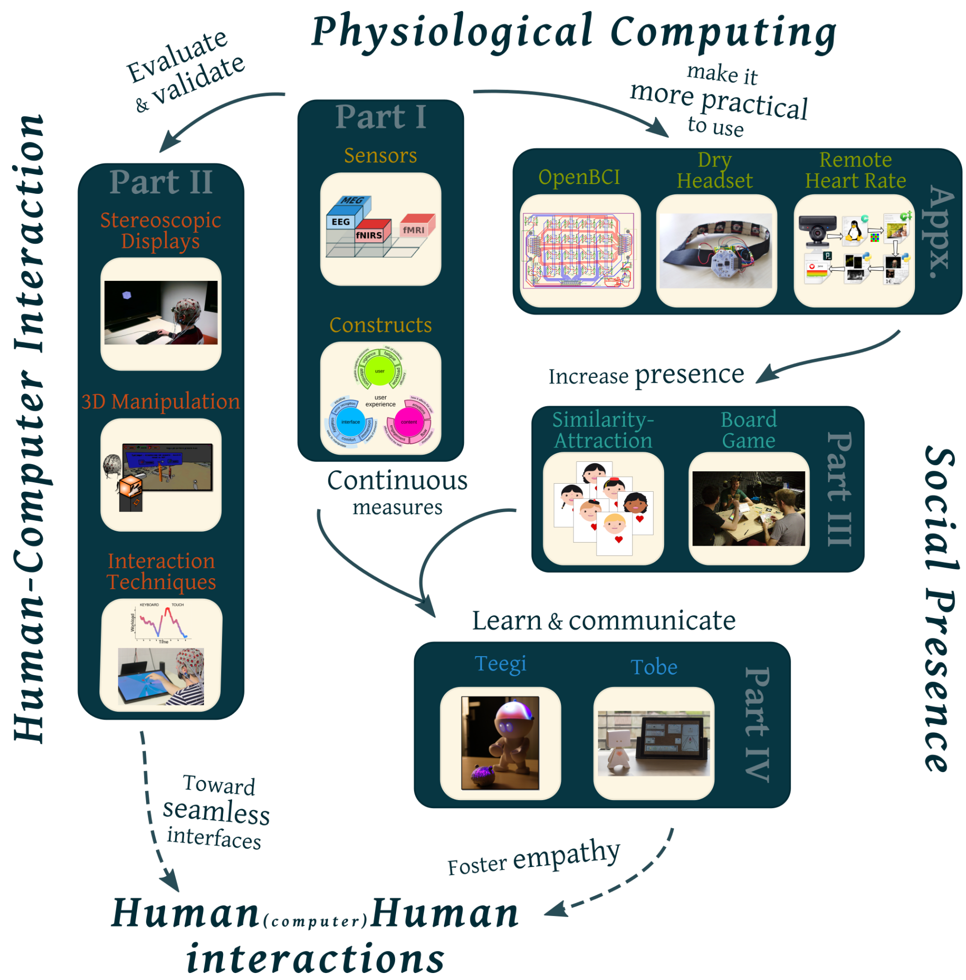

Figure 1 – Outline of the thesis.

Finally, I helped to develop a toolkit that enables the embodiment of physiological signals and mental states in tangible avatars. A first application concerns scientific outreach and let novices to know more about EEG and brain signal – chapter 10. The second implementation, described in chapter 11, encompasses higher level signals, brings customization and ventures into multi-users scenarios. Those proxies enable users to better know themselves and the others.

An outline of the thesis, with the principle chapters, is pictured in Figure 1. To guide readers in a hurry all along these works, each part as well as the longest chapters are preceded by a "takeaway message" – an abstract if you will.

PART I

ASSESSING A USER

In this part we review how one could assess mental states related to HCI. We first describe what tools are available, with their main advantages and their drawbacks. We divided those evaluation methods in 4 categories: behavioral studies (measured from one's interactions); inquiries (what people express); physiological sensors (how the body reacts); and finally neuroimaging (a subcategory of physiological sensors that encompasses brain activity).

A pessimistic defending his thesis would say that traditional evaluation methods are potentially biased or disrupt the interaction. This is not the case for physiological sensors, which are further away from conscious thoughts – i.e. "exocentric" – and which give real-time measures. Neuroimaging techniques are often more accurate and more versatile. EEG for example can account for various affective and cognitive states that we call "constructs". We detail each one of the constructs involved in HCI and frame how the 4 types of evaluation methods can assess them. Namely, we studied workload, attention, vigilance, fatigue, error recognition, emotions, engagement, flow and immersion.

The content of this part is an extended and updated version of the work previously published in [Frey et al., 2014b], at the beginning of my thesis. At that time we cornered the limitations of the neuroimaging techniques, which may hold the most promising applications, but that come at the expense of a more important setup. Along the following years, while we were putting into practice the usages we envisioned – e.g. HCI evaluation in part II – we also helped to lower those technological barriers, as you will *read later on in the Appendices.*

1

Evaluation methods in HCI

Until now: inquiries that disrupt, behavioral measures that are ambiguous.

Along HCI history, various methods aimed at evaluating interactions and user interfaces (UI) prior to their public availability have emerged; there have been behavioral studies (observations of users actions in real time) and inquiries (e.g. questionnaires, interviews, think aloud). Yet those traditional evaluation methods could either be ambiguous, lack continuous recordings, or disrupt the interaction.

Recently, technologies centered around the measure of bodily activity appeared. Physiological sensors help to improve the ergonomics of HCI [Fairclough, 2009], for example with systems that could be tuned to users by monitoring their mental workload in real time [Kohlmorgen et al., 2007]. Physiological sensors add an insightful information channel, that could be used in HCI evaluation. Indeed, while designing a UI it should be acceptable to add the hindrance that comes with some of those sensors – putting on electrodes, calibrating the system, … – to specially enrolled users. Those testers will then help to improve beforehand the UI. Laboratory conditions permit a controlled setup for repeatable measures.

Brain activity helps to understand even more users.

Neuroimaging, a subset of physiological sensors which records brain activity, rely on demanding but sensitive sensors. We consider them as an innovative supplement to conventional evaluation methods. Measuring neural activity during HCI can help us to better understand what occurs in the brain when users are interacting [Parasuraman, 2013]; we will highlight which neuroimaging techniques could be used conveniently within laboratories to overcome the difficulties encountered by traditional evaluation methods alone.

1.1 Behavioral methods

Recording users interactions, such as mouse speed, is one standard way to evaluate a UI. "Behavioral studies" refers to this method: behavior and actions of users inside a software. It represents all the metrics that can be computed from users interactions, not body measures from the "outside" environment (e.g. muscular activity) – these will fall into the "Physiological sensors" section. Behavioral studies are close to performance measures, as seen in human factors. The easiest way to sense if a UI is well designed is to watch users. How fast do they complete the task? How many errors? Are they more accurate with a slower or bigger mouse cursor? Such methods helped to formulate a preeminent law in HCI, Fitts's law, which is all about time to reach a target depending on its distance and size [Fitts, 1954][1].

Behavioral methods seek how users interacts with a system.

Even if behavioral studies take root in the History of computer science, it is worth noting that new approaches are developing. In [Evans and Wobbrock, 2012] a clever combination of text corpus, crowdsourcing and machine learning is used to determine users' *intentions* while typing on keyboard or using a pointing device. In a study more oriented toward content industry, machine learning is used again, this time to build profiles from hundreds of users' behaviors and see how those "archetypes" match designers' intents [Drachen et al., 2009]. It is also possible to predict users' behaviors by applying psychology models [Cowley et al., 2009].

During complex interactions our behavior can give away our intentions.

Although behavioral studies are able to account in real time for users' interactions, they can be hard to interpret: measures may not be specific to one construct. E.g. a high reaction time can be caused either by a low concentration level or a high workload [Berka and Levendowski, 2007, Hart and Staveland, 1988]. Behavioral studies may also not provide much information on the users' state, as on simple tasks little can be computed beside reaction times and a performance metric.

With simpler tasks, though, behavioral methods are limited.

1.2 Inquiries

While it is possible to infer users' thoughts through a behavioral study, it may be simpler to record their opinion. We call this "inquiries". In HCI we are interested in questionnaires related to the use of a UI. Standardized questionnaires have been validated across several studies for various measures: e.g. NASA-TLX for workload [Hart and Staveland, 1988].

Inquiries: if you want to know, just ask.

Unfortunately those "pen and paper" tests are discrete and are not good for real-time assessments. The "think aloud" protocol [Weber, 2007] is a way to circumvent this, yet it could influence the interaction as users still have two different things to do: interact and report their

[1] Fitts's law is still widely used nowadays for evaluation, proof is [Gervais et al., 2015], one side work that needed attention.

experience. It is an example of double task and divided attention [Ogolla, 2011]. "Focus groups" [Bruseberg and McDonagh-Philp, 2002] is the third form of inquiry. It involves experts and advanced users, who exchange about their findings under the control of the designer.

Inquiries give qualitative information that could be biased by many external factors.

Questionnaires, think aloud an focus group are three different forms of inquiry fraught with the same hazards. Resulting measures are prone to be contaminated by ambiguities [Nisbett and Wilson, 1977], social pressure [Picard, 1995] or participants' memory limitations [Kivikangas et al., 2010] – we remember most the first and last items seen and tend to forget what was in-between. If participants figure out what is at stake, answers could also be oriented toward experimenters' expectations (or *against*, depends).

1.3 Physiological sensors

Not only do humans interact with computer using their bodies, but as soon as they act, as soon as they get feedback of their actions, changes occurs inside these very bodies. Spirit and flesh are linked. E.g. respiration rate increases with workload [Karavidas et al., 2010], pupils dilate while experiencing strong emotions [Partala and Surakka, 2003]: many cues a ~~mentalist~~ attentive observer can perceive. Fortunately for less gifted observers, a broadening set of physiological sensors can be used in order to account for such body changes in HCI [Fairclough, 2009, Dirican and Göktürk, 2011] or game [Ravaja, 2009, Nacke et al., 2009] research. Electrodermal activity (EDA, also called "galvanic skin response") is among those sensors, as well as electrocardiography (ECG, the signal modality heart rate is derived from) and electromyography (EMG, caused by muscular activity, including facial expressions).

Physiological sensors combine real-time measures and exocentricity.

Physiological cues are great for the "objectivity" they bring into HCI[2]. They are also more and more present in everyday life, thanks to their integration in smartwatches and to technologies that can sense the body remotely – a hint for what is to come in parts III and IV. Body reactions are sometimes misleading, though: you may record ECG to study attention, whereas an increase in heartbeat can also be caused by strong feelings. Muscles and organs are controlled by the peripheral nervous system. Physiological sensors are a second-order inference about the processing which occurs in the central nervous system. It may then be interesting to go further toward the origin of these signals in order to gain even more accuracy and reliability.

1.4 Neuroimaging

Neuroimaging senses brain activity.

As their name suggest, neuroimaging techniques allows the assessment of brain activity. We classify them apart even if strictly speaking they do belong to physiological sensors. We restrain our overview of the subject

[2] see chapter 2 for a brief discussion about what objectivity is not

to living organisms. Brain slices are valuable to study brain connectivity but useless to HCI. Neither will we examine invasive techniques – i.e. which require surgery, like cortical electrodes. Neuroimaging is a currently rising field used in brain-computer interfaces (BCI) settings [Blankertz et al., 2010], [Hamadicharef, 2010], that we will describe more thoroughly in next part.

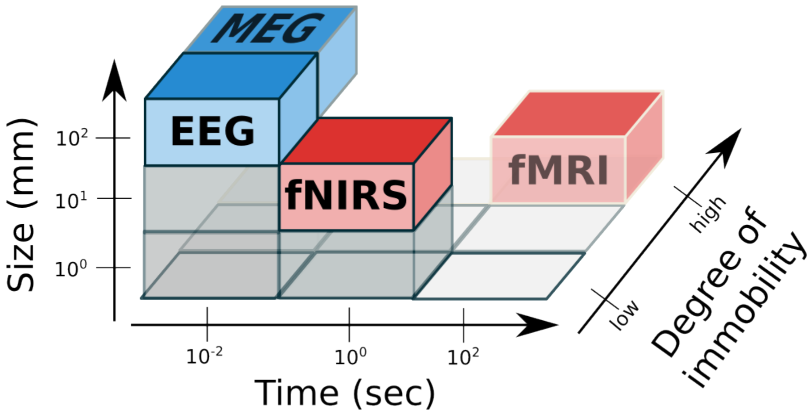

Figure 1.1 – Neuroimaging techniques most often used for ergonomics and in BCI. In blue those relating to electrical currents, in red those relating to blood flow variations. *X-axis*: temporal resolution, *y-axis*: spatial resolution, *z-axis*: degree of immobility. From [Ward, 2006, Mehta and Parasuraman, 2013].

fMRI and fNIRS relate to blood flow variations. The former is heavy, both possess a poor temporal resolution.

Non-invasive neuroimaging techniques are divided into two main families [Mehta and Parasuraman, 2013] (figure 1.1) The first family comprise functional magnetic resonance imaging (fMRI) and functional near-infrared spectroscopy (fNIRS). They indirectly records brain activity through blood flow variations (discharging neurons need more oxygen, hence more blood). fMRI has a very good spatial resolution but is a large device which completely surrounds subjects and costs about one million dollars. fNIRS is a much more lightweight and affordable device. Instead of magnetic fields, it uses direct light, invisible to human eye, to record cervical blood "color" through the skull. Sensors are fixed on a cap, hence subjects are free to interact with a computer while wearing it. Compared to fMRI, the spatial resolution of fNIRS is less detailed. It records only the outer region of the brain due to physical limitations – light is absorbed by tissues. fMRI and fNIRS share a poor temporal resolution. With a latency reaching up to several seconds it is difficult to observe fast and short responses.

MEG and EEG relates to electrical currents, both posses a high temporal resolution.

The second family of neuroimaging uses electrical currents generated by neural activity. Magnetoencephalography (MEG) records magnetic fields. It is less heavy and expensive than fMRI, but still hardly manageable for uses in HCI contexts. MEG has a high temporal resolution, down to the millisecond. Electroencephalography (EEG) also has a high temporal resolution. It is comparable in size to fNIRS. EEG measures electrical current onto the scalp. Electrodes are "dry" – no electrolyte solution – or, more frequently, "wet" – solvent is either water or gel. Despite its poor spatial resolution because of volume conduction effect, it is relatively cheap equipment for a laboratory. Because it is portable and non invasive, it interferes little with HCI setting.

EEG is the best tradeoff between temporal resolution, spatial resolution, affordability and ease of use.

Experimenters must be cautious with the limitations of the device they choose. Is the signal-to-noise ratio sufficient for what they intend to measure? What artifacts could pollute their data? Are they in control of the algorithms producing measures from raw signals? That said, EEG is the most promising candidate to assist inquiries and other physiological sensors in a wide range of evaluation measures. Compared to others neuroimaging devices, EEG offers the best compromise between spatial and temporal resolution, practical use and cost; a versatile technique which benefits a lot from recent progress in signal processing. Properly used, it also gives access to many different mental states. Therefore we focused mostly on this type of brain activity recordings, in this part and during the thesis as a whole.

2

A new continuum for evaluation methods

In many works there is a debate if whether or not some evaluation methods or some particular tools bring different levels of "objectivity" in their measures. The antagonism of the different positions come mainly from the fact that in such context "objective" and "subjective" are scarcely defined in the literature. According to [van de Laar et al., 2013], "the objective methods are based on overt and covert user responses during interaction while the subjective methods rely on user expressions after the interaction". From that perspective, inquiries are "subjective" while behavioral studies, physiological sensors and neuroimaging are "objective".

While we agree such a distinction is required, a more rigorous vocabulary is needed. We also doubt the "time" variable should be involved in the definition. As stated in previous chapter, results of inquiries are prone to social pressure and other self-interpretations, and this is also true for the real-time think aloud. Moreover, when studying emotions, it could be argued that only "subjective" feelings are recorded, as the evaluation is centered on the user. As a matter of fact, this is probably the main argument that is replied when one speaks about objectivity. Hence, without a complex phrasing (i.e. "objective measure of subjective feelings"), employing such words is open to criticisms. As an alternative "direct" and "indirect" could be considered.

But then those concepts are more likely to refer to how measures are reported, not where they originate from (e.g. EMG vs an external observer annotating facial expressions).

As such, we would like to introduce a new nomenclature to name those two aspects and avoid ambiguities: exocentric and egocentric. Those terms are borrowed from spacial navigation research [Brandt et al., 1973] and bring the notion of the self. Exocentric measures are here close to the stimuli, to the source, while egocentric measures are close to the conscious thoughts of the user, to the outcome.

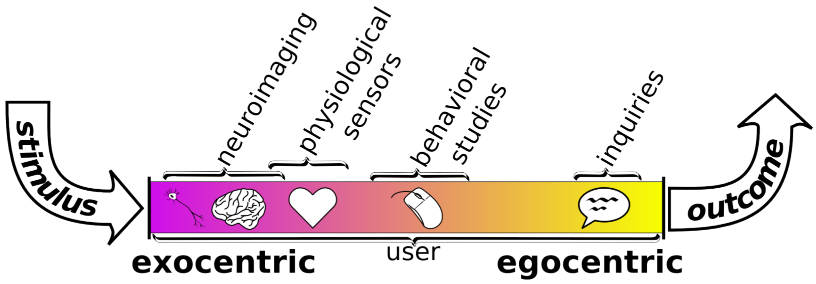

Figure 2.1 – Proposal of an "exocentric / egocentric" scale aimed at classifying evaluation methods for HCI.

We therefore create a continuous space between two extremes (see Figure 2.1). We illustrate this scale with the measurement of pain. The pressure of a needle on a finger would represent a perfect exocentric measure: the stimulus' strength, a value disconnected from human body and perceptions. When the pressure is transmitted to nociceptors in the skin, the measure shifts a little from exocentric to egocentric. As nerves are transmitting signals from the peripheral nervous system to the brain, we go further to the right of the axis. Since we may not be interested in skin's thickness, this neural activity represents the first interesting value from this side of the exo/egocentric scale. Neuroimaging techniques record such activity, hence it is the most exocentric evaluation method. When the signal reaches the central nervous system, autonomic responses are triggered – increase in heart rate, electrodermal activity [Loggia et al., 2011]. Those reactions could be recorded through physiological sensors, a step further from the exocentric extreme.

This scale can be used for various evaluations. Eventually, it is possible to add "objective/subjective" and "direct/indirect" to describe a whole framework. A construct could be objective (usability) or subjective (emotions). A tool could be either direct (sensor) or indirect (observer). A method is more exocentric (neuroimaging) or egocentric (inquiries). E.g. the work of an experimenter assessing workload with ECG can be described as objective/exocentric/direct.

As the pain grows, it will alter behaviors and thoughts. A runner may slow down when experiencing pain in a foot, no matter his willingness. Behavioral studies are able to sense modifications occurring against the will of the subject; that could be placed somewhere in the middle of our scale. Concurrently, most of the time, the person is being aware of the pain and could phrase it if asked to. Many other cognitive processes are involved in such a high level of consciousness (e.g. planning, awareness), thus measures recorded by inquiries are close to the far-end of the scale and are indeed egocentric.

3

Constructs

Now that we are on par with the evaluation methods, we hereby review a repertoire of patterns of users' state which could be used to characterize HCI, and assess how neuroimaging objectively measures them. We call those patterns "constructs", a term which refers to notions as different as workload and the state of "flow".

"Constructs" encompass the various states that could be measured from brain activity.

Previous works already began to sense how neurotechnologies benefit HCI, but they do not cover evaluation [George and Lécuyer, 2010], or if so they do not study many constructs. [Parasuraman, 2013] only discusses workload, vigilance and error recognition. Here we sought to gather from the HCI literature every major construct which relates to the quality of HCI and that could potentially be evaluated with brain activity, with a focus on EEG – as such we do not claim to produce here an exhaustive study, of each evaluation method, for every construct.

We grouped together similar or highly correlated measurements to ease reading. Starting from a literature which on some occasions expresses different viewpoints, we tried to define accurately those constructs whilst remaining brief. We identified that workload, attention, vigilance, fatigue, error recognition, emotions, engagement, flow and immersion are useful for evaluation and can be measured with EEG.

3.1 Workload

In cognitive science it is theorized that humans' possess a limited set of resources to process information [Just et al., 2003]. The workload endured by the brain is determined by the ratio between processing power and data coming from the environment. Workload increases as cognitive resources lessen or as the quantity of data grows – and the other way around.

Workload assessment is critical in sensitive contexts.

Different brain structures handle the information depending on its nature [Just et al., 2003]. If the workload is too high subject's performance decreases, sometimes dramatically. In this thesis we assimilate the term "workload" to mental effort; and although in the fifties the notion of "effort" was mostly constrained to physical activity, since then it has been shown that mental activity as well was source of exhaustion [Fairclough, 2001]. This is why it is critical to check the strain induced on workload by an interaction, moreover in a sensitive context (army, hospital, transportation, ...).

3.1.1 Behavioral studies

Because performance vary with workload, behavioral studies can reveal a too demanding task. However, as in all behavioral studies, it is difficult without any prior knowledge to conclude that the problem lies for sure in workload. On top of that, for the performance to drop significantly, cognitive resources will have to bee overloaded. In [van Drunen et al., 2009] mouse clicks and movements have been recorded along with physiological data in order to validate workload measurement, but authors did not find a strong correlation. Thus, it seems questionable that behavioral studies could stand as a reliable and precise tool to evaluate this construct.

3.1.2 Inquiries

The NASA-TLX (NASA Task Load Index) is a questionnaire which elaboration involved dozens of laboratories and spanned over several years [Hart and Staveland, 1988]. It is nearly systematically used as a validation tool in workload studies. The questionnaire consist in 6 items (mental demand, physical demand, temporal demand, performance, effort and frustration) subjects have to rate on Likert scales.

3.1.3 Physiological sensors

Pupils dilatation is among the physical sensors which has been reported to correlate mental workload [Just et al., 2003]. ECG as well, which in [Mathan et al., 2007] is demonstrated to be even more precise than NASA-TLX. Within ECG signal it is heart-rate variability that denotes mental effort [Fairclough and Houston, 2004]. A more direct measure of metabolic activity, albeit more difficult to put into practice because of the invasive nature of the technique, consists in gauging glucose level in blood [Fairclough and Houston, 2004].

While the technique is still in its infancy, [Tuček et al., 2012] investigate speech characteristics as another mean to asses workload. However, in this particular use of physiological sensors, data is not acquired in real time. Different interaction techniques were used during the completion of a sudoku game. The protocol relies on the capacity unused in task performance, as speech fluency analysis occurred only after the completion of a trial, during a second task in which subjects had to make assumption about a scene presented with picture. This illustrates how the pros and cons that we listed when we categorized evaluation methods do not always apply as is before the variety of their continuum.

3.1.4 Neuroimaging

Using a device with 9 channels [Berka and Levendowski, 2007] correlated EEG with workload. With a better equipment [Mathan et al., 2007] shows how EEG outperforms physiological sensors, with more subtle changes measured compared to ECG.

fNIRS in another well-tried neuroimaging technique which has been studied to assess users' workload, for example to evaluate user interfaces [Hirshfield et al., 2009b]. The relation between oxygen consumption, neuronal activity and workload is pretty straightforward. In [Peck et al., 2013] a task involving both vision and memory is studied. Directly compared to EEG, fNIRS show better results, with 82% of correct classifications between 2 classes (i.e. low *vs* high workload) and 50% with 3 classes (low, medium, high) [Hirshfield et al., 2009a].

fNIRS is very often used to assess workload.

In [Blankertz et al., 2010] EEG online analyses (i.e. in real time) discriminate 2 classes with a 70% accuracy. A 2 minutes time window enables scores from 80% to 90% [Brouwer et al., 2012]. Other reviews report classification between two classes up to 90%, and a score close to 100% if EEG is combined with other physiological sensors [van Erp et al., 2010]. Not every study presents such striking results as follows, but [Grimes et al., 2008] claims 99% success in distinguishing two memory load levels and 88% for 4 levels. Though, as in many BCI experiments, not many participants were involved in this last study (only 8).

We had the opportunity to confirm many of these findings. In the study that we will present in chapter 6, we obtained up to 88% classification accuracy over 2s time windows with EEG, which outperformed both EDA and ECG.

3.2 Attention – Vigilance – Fatigue

Attention, vigilance and fatigue are closely related and regularly measured altogether [Oken et al., 2006].

"Attention" refers to the ability to focus cognitive resources on a particular stimulus [Kivikangas et al., 2010], that is to say to perceive changes from the environment. A correct selective attention allows ignoring distractors (information not relevant to the current task). An insufficient attention level results in a difficulty or an inability to complete the task, whereas too high or narrow attention resources may prevent someone to disengage from a sub-task – e.g. no perception of

A high attention level enables one to focus on a particular stimulus, like this note.

the terminal cue or of alarm signals triggered in the background. In the present definition, the notion of "attention" is similar to the notion of "awareness".

Vigilance lets one to maintain focus over time.

While in the literature, "attention" designates more frequently the ability to perceive changes from the environment, the term "vigilance" then often refers to a broader resource, dependent of both cognitive performance and the arousal level on the sleep–wake spectrum [Oken et al., 2006]. In that sense it refers to a state of sustained attention. One needs to maintain a high degree of vigilance over time in order to focus his attention on something. Hereby "alertness" will be considered as a synonym of "vigilance".

At some point too many notes will exhaust your cognitive resources, that is fatigue.

"Fatigue" is a state in which cognitive resources are exhausted. If the required level of vigilance or attention causes a strain too important on the organism, fatigue arises and performances decrease [Boksem et al., 2005]. Then the task cannot be performed correctly and errors appear [van Erp et al., 2010].

3.2.1 Behavioral measures

An increase in reaction times and in the number of errors can be observed with users who undergo mental fatigue [Lorist et al., 2000]. But it is not systematic and depend on the nature of the task; by coping strategies users can overcome fatigue in their behavior. In [van der Linden et al., 2003] performances of a simple memory task were not affected by fatigue, while subjects displayed impairments in planning process with more complex exercises.

3.2.2 Inquiries

While it could be somewhat biased to ask a subject if he or she is attentive, numerous questionnaires have been developed to assess fatigue. Some scales have been well validated – e.g. [Chalder et al., 1993] reports high sensitivity and specificity coefficients. But experimenters have to be cautious when they select one, because sometimes they measure different things; sleepiness rather than a more typical fatigue experience [Dittner et al., 2004]. On top of that those questionnaires have mainly been created by the medical community and it is important to discriminate normal fatigue from fatigue related to medical disorders [Schwartz et al., 1993].

3.2.3 Physiological sensors

As stated in [Blankertz et al., 2010], sec. 3.1, eye blinks and heart rate are the most widely used physical sensors to detect distracted attention, lapses in vigilance and fatigue. For the latter, eyelid closure is a particularly good sign of drowsiness. See also [Kivikangas et al., 2010] for references about EDA and EMG – allocation of cognitive resources impacts the autonomic nervous system.

3.2.4 Neuroimaging

The alpha band is associated to attention. When eyes are closed, or when fatigue occurs, alpha waves amplitude increases [Shaw, 2003]. This frequency band in the range 8-12Hz is mostly generated by the occipital lobe. It is easily recorded with EEG, even with a single electrode [George et al., 2011]. Alpha band analysis discriminates different attention levels [Klimesch et al., 1998]. Even more, it enables to detect which side of his visual field a subject is paying attention to while his eyes stare in front of him with 70% accuracy [Trachel et al., 2013].

Appearance of alpha waves within EEG is a good test to see if actual brain activity is recorded.

Other frequency ranges can be recorded to improve reliability. With alpha, theta (4-8Hz) and beta (13-18Hz) bands combined, [Laurent et al., 2013] detect mental fatigue on 4s time windows with 80% accuracy, 94% over 30s. Other types of brain activity are used, such as delays in event-related potentials (ERP) – e.g. visual selective attention in [Saavedra and Bougrain, 2012]. [Berka and Levendowski, 2007] suggested that EEG is the only sensor which can accurately report attention and vigilance shifts on a second-by-second time frame. Works investigating vigilance measures are reviewed in [Parasuraman, 2013].

Regarding fatigue, if EEG signals are not more accurate than physiological sensors to detect micro sleeps, they offer the possibility to detect preceding inattentive states [Blankertz et al., 2010], sec. 3.1. Mental fatigue has been detected on 4 seconds time windows with 80% accuracy, or 94% over 30 seconds [Laurent et al., 2013]. In order to improve reliability, additional frequency ranges were recorded in this study. For instance alpha, theta (4-8Hz) and beta (13-18Hz) bands have been combined. ERP on the other hand have been used to study how fatigue impairs differently cognitive processes [Lorist et al., 2000].

The construct evolving around attention could be one of the main beneficiaries of neuroimaging. To distinguish clearly in their measurements vigilance and fatigue would be one point. On the other hand EEG studies showed that visual artifacts in images or videos are detected by subjects beyond consciousness [Scholler et al., 2012], whenever it is conscious perception or attention [Mustafa et al., 2012]. This would suggest that ERP could be used to anticipate how much information users are able to process, before even considering their attention level.

Various cues hidden within sensory modalities in order to elicit evoked potentials would even create a “human bandwidth” assessment, upstream from vigilance and attention. Such application has been studied with audio cues while participants were playing a video-game [Burns and Fairclough, 2015].

We will put into practice these types probes in the next part.

3.3 Error Recognition

Error recognition, or when the external world behaves poorly.

We call "error recognition" the situation which occurs when users detect *by themselves* an outcome different from what is expected [Nieuwenhuis et al., 2001]. It can be something users genuinely trigger but then they realize they did a mistake, e.g. a wrong turn while driving a car. Or it can happen due to commands erroneously interpreted by the machine. E.g. a user manipulates a 3D object in a modelling software and presses a key combination on the keyboard in order to rotate it, but instead its scale is modified.

It is important to notice that here "error recognition" do not account when a negative feedback is given *per se*. Instead, it is a matter of recognition *by the user* of a faulty event. A driver should have turned left but turns right and realizes the mistake as he does it, without any road sign involved. The theoretical modelling software would simply execute *another* command upon error, not show a red dialog box written "bad input". In UI evaluation, error recognition could be an objective measure of subjective (mis)representations, an objective assessment of how intuitive an HCI is.

Four types of errors could be distinguished both conceptually and in practice [Ferrez and Millan, 2008]:

- The *response error* is detected with an operator realizing that he or she made a mistake.
- The *interaction error* arises when a system reacts in an unexpected way.
- The *observation error* is produced when a subject sees a third person committing a mistake.
- The *feedback error* is measured when a feedback (reward or punishment) differs from what is expected.

If the mechanism *producing* the error is not studied here, the ability of a person to detect a non congruent event depends directly on his or her vigilance state and his or her level of attention. As soon as an error is detected, either the operator will try to rectify it, if he or she can modify the system, either he or she will change his or her inner representations to accommodate the system, as it is often the case in a learning context.

3.3.1 Behavioral studies

In software which enables to go back in command history, like with the "undo" command in office suites, the number of corrections should be linked to error production. However, command history can also be used as a way to explore the different options proposed by the software, or as a step in the creative process of the user. It is then more a usability issue [Akers et al., 2012] than error recognition.

3.3.2 Physiological sensors

It is somewhat natural to think about physiological sensors to assess error recognition. After all, who doesn't swear out-loud when he misses something or shout at a reluctant computer? Alas, those reflexes – which could well be analyzed through a sound sensor – are not *that* systematic, hence proceed to an experimental protocol seems slippery at the least. Muscular activity such as facial expressions (e.g. frowns) could be recorded over the course of the task [Mirza-Babaei et al., 2013]. But then it would be hard to discriminate emotions related solely to error recognition from emotions triggered by other events.

3.3.3 Neuroimaging

Event-related potentials (ERP) are "peaks" and "valleys" in averaged EEG recordings associated with an external event. ERP differ in their "shapes", place on the scalp and latency depending on the source of the stimuli or on the underlying cognitive mechanism. High-level processing (e.g. planning, memory) takes longer than low-level (e.g. sensory information, motor reflex).

One particular kind of ERP has been discovered: error-related potentials (ErrP) [Schalk et al., 2000]. They are triggered when an "error" occurs, as previously defined. All the four different types of errors that we described (response, interaction, observation, feedback) have distinguishable features, hence they could separately be measured with EEG. During the completion of a simple task, like target selection, a negative component occurring 80ms after a trial characterizes a response error [Falkenstein et al., 2000], while a latency of 250ms is associated to a feedback error [Holroyd and Coles, 2002]. The appearance of a delayed positive component denotes an interaction feedback [Ferrez and Millan, 2008]. The amplitudes of the components relates to the frequency at which the errors appear [Ferrez and Millan, 2008]. ErrP are closely related to decision-making process [Fedota and Parasuraman, 2010] and dopaminergic activity [Holroyd and Coles, 2002].

Brain signals are elicited even when users are not consciously aware of errors [Nieuwenhuis et al., 2001]. ErrP have been used to discriminate between incorrect and correct users decision. In [Chavarriaga and Millan, 2010], respectively 76% and 63% classification accuracy with observed ErrP. In "single trial", that is to say in detecting ErrP for each user's action, these are scores commonly found in literature. Other studies expose similar recognition rates: 79% and 84% in a task involving interaction ErrP [Ferrez and Millan, 2008]; 79% and 83% with observed ErrP [Iturrate et al., 2010]. Accuracy relates to the quality of EEG devices. It can vary from 70% with an entry-level headset and non gel-based electrodes [Vi and Subramanian, 2012] up to 90% with a more expansive device [Schmidt et al., 2012]. While ErrP detection does not yet reach 100% (chance is 50%), those scores are already sufficient to improve HCI reliability in various task, such as visual discrimination [Parra et al., 2003]. There are successful examples in a target acquisition task with two different interaction techniques, touch-based in [Vi and Subramanian, 2012] and mid-air gestures in [Chavarriaga et al., 2010]. When EEG recordings already takes place, as in BCI settings with mental typewriter, ErrP could be used to improve the accuracy of the system with no additional cost [Schmidt et al., 2012] – see [Chavarriaga et al., 2014] for a broader review of the possible applications of ErrP in a BCI context.

[Sobolewski et al., 2013] recorded EEG while subjects use a mouse and have to reach different targets. In one-fourth of the trials the hand-to-cursor mapping is randomly off-set by several degrees. Users do not expect these shifts and the analysis gives first insights that the amplitudes of elicited ErrP could relate to the degree of error. If this result is confirmed we may link error recognition to "intuitivity" evaluation.

Currently only a binary measure and poorly detailed data – "an ErrP is detected or not" – is reliably obtained. Fortunately it seems possible to measure a modulated ErrP [Sobolewski et al., 2013], thus sensing by how much an operation in the UI has perturbed users. If it is to be confirmed, this would enable a quantitative and qualitative data assessment. We saw how single trial detection can be achieved with EEG. Promising work reported ErrP detection as the movement is occurring, within a 400ms time frame [Milekovic et al., 2013]. At the moment this near continuous detection uses an invasive technique.

In chapter 7 we will investigate how interaction errors could be used to compare a keyboard-based interaction technique with a touch-based interaction technique.

3.4 Emotions

Emotions are more than just a matter of passive perceptions. Psychology and neuroscience showed that contrary to previous beliefs, emotions are not disconnected from high-level reasoning. They are tightly linked to decision-making processes [Damasio, 1994]. As emotions both arise from subjective feelings and impact our interactions, their study is getting more and more attention from the HCI community [Picard, 1995, Hancock et al., 2005].

The valence/arousal model is the most commonly used paradigm to categorize emotions [Picard, 1995, Posner et al., 2005, Kivikangas et al., 2010]. In this two-dimensional representation, valence is related to hedonic tone and varies from "negative" to "positive" (e.g. frustrated *vs* pleasant); arousal is related to bodily activation and varies from "calm" to "excited" (e.g. satisfied *vs* happy). While the valence/arousal model is a useful tool to categorize emotions, it must be applied with caution with some population. Children, for example, hardly make distinction between different arousal levels [Posner et al., 2005].

3.4.1 Inquiries

The self-Assessment Manikin (SAM) is an effective and quick way for subjects to report their affective response [Bradley and Lang, 1994]. Two scales are presented to them, each one aimed at rating valence or arousal. A sketchy character is associated with each measure, taking different traits along the axis in order to guide subjects in their evaluation. The SAM is widely used in emotions studies, for example to validate other evaluation methods – e.g. [Nijboer et al., 2009, Soleymani et al., 2009]. Emotions, and how we deal with them, are a sensitive subject though, and a potential bias in the assessment of emotional states through inquiries lies maybe in our tendency to conceal negative feelings.

Sometimes a third component is investigated with SAM, "dominance".

3.4.2 Behavioral studies

How emotions affect behavior has been described through approach-avoidance motivation. Our reaction time varies depending on the valence associated to a stimulus; we are quicker to dodge something that represent a danger (avoidance motivation) and more prompt to seize an object that appears to bring a reward (approach motivation) [Chen and Bargh, 1999].

Even though those findings may be combined with the fact that cursor trajectory is affected by distracting elements [Hurtienne et al., 2014], it was another feature derived from the mouse that was measured by [Sun et al., 2014] so as to sense the emotional state of a user. In this study, the mouse patterns associated to muscle stiffness were used to differentiate between a calm and a stressful condition. Authors concluded that this indicator was more sensitive and more robust than ECG, although a classification score was computed only for the former (70%).

3.4.3 Physiological sensors

Three physiological sensors are mainly used to assess emotions. ECG is among the most widely employed, both for valence and arousal. Yet, since heart is regulated by many different bodily processes, it is difficult to ensure signal reliability in complex situations [Kivikangas et al., 2010], especially if subject is moving.

EMG, through facial expressions, is used as a valence indicator. Sensors are usually placed on the cheek and on the eyebrow to record smiles (positive valence) and frowns (negative valence) [Mandryk and

Atkins, 2007]. Because of these positions, EMG is prone to noise and could be affected by social communication [Kivikangas et al., 2010].

EDA is associated with arousal; while in general EDA is less sensible to noise than ECG and EMG and very easy to setup, responses are delayed between one and four seconds [Kivikangas et al., 2010].

EDA relates to the sympathetic branch of the autonomic nervous system.

There is not a sensor on its own which is able to record adequately emotions. This is why a combination of sensors is often employed. Either EMG and EDA to account for both dimensions of the arousal/valence model [Soleymani et al., 2009], or more frequently – for even better results – the three mentioned sensor at the same time [Ravaja et al., 2006, Mandryk and Atkins, 2007]. For example, using EDA, ECG and temperature [Lisetti and Nasoz, 2004] report a 84% accuracy in detecting 6 emotions; [Picard et al., 2001] used EDA, EMG, blood pressure and respiration and reached 81% in accuracy for 8 emotions.

3.4.4 Neuroimaging

While different neuroimaging techniques have been used to study how the brain responds to emotions, technologies with the highest temporal resolution, such as MEG or EEG, are more indicated when a dynamic content is involved [Vecchiato et al., 2011].

An asymmetry of the alpha band power in the frontal brain correlates with the emotional valence. A negative valence is associated with a power decrease in the left lobe. On the opposite a positive valence is linked to a decrease in the right lobe [Molina et al., 2009]. The arousal level of a stimulus is more easily perceived through the theta band, or by studying the amplitudes of ERP [Molina et al., 2009]. Still, EEG is not yet a reliable sensor to assess emotions. In [Chanel et al., 2011] even if EEG was better than the other studied physiological sensors on short period of times, a 56% accuracy barely suffices for the differentiation of three emotions (chance level is 33%).

Some Papers report high classifications rates. In [Liu et al., 2011] 7 emotions are categorized. Authors state a 85% accuracy for arousal and 90% for valence. This using only three channels of an EEG headset which is known to be sensitive to EMG artifacts. In pure EEG studies it is important to control for facial expressions (i.e. EMG signals), because they can be easily recorder by electrodes. This is even more problematic when emotions are involved. Although we have to be cautious when assessing EEG reliability, there is nothing wrong in combining EEG and EMG (or other sensors) to improve overall performance or to build proof of concepts – we have done so when we tested a device in a public exhibition, see chapter 11.

Despite the lack of clear indicators of affect in EEG, neuroimaging is nevertheless a good lead for novel research in this topic. For example different patterns of EEG signals have been observed depending on the sense (sight or hearing) which induces an emotion [Mühl et al., 2011]. It could then be speculated that neuroimaging one day will be able to discriminate which emotion is elicited by which input modality, or which information channel leads to positive and which to negative user experience.

3.5 Usability – Comfort

The term usability groups together the notions of "ease of use" and "usefulness" [Bowman et al., 2002]. It relates to speed, accuracy and error rates in task completion, hence it depends on UI. The learnability of UI, that is to say how fast a user learn to use an interface, is also a key point of usability. As such a good affordance of UI elements – how perceptions of objects induce a proper use – will improve overall usability. Usability suffers from UI nature and constrains. E.g. a gesture-based input device such as the Microsoft Kinect is likely to be more tiring than a joystick in the long run. Usability is inextricably bound to users' comfort because of that.

3.5.1 Behavioral studies

By definition, usability on its own can be partly assessed from speed, accuracy and error rate measures when two different UI are compared. Events logging, such as navigation in commands history by users, allows restraining which parts of the interaction need improvements [Akers et al., 2012].

3.5.2 Inquiries

It is possible to foresee how different designing choices in an UI will impact overall usability. Focus groups of experts originating from various fields such as ergonomics, cognitive science or physiology help to quicken tedious development process based on trial-and-error approach and will avoid dead-ends – e.g. [Hix et al., 1999, Stanney et al., 2003].

A poor usability brings confusion to users, who do not understand instructions. Some studies use a think aloud protocol to record such events [Pike et al., 2012]. More commonly questionnaires are given to users, where for example they have to rate items such as "ease" or "feels good" [Bos et al., 2011].

The System Usability Scale tries to standardize this approach around 10 criteria (e.g. "I found the system to be simple", "I felt very confident using the system", …). This scale has been correlated to behavioral measures such as speed or number of errors [Sauro and Lewis, 2009]. Even a subset of 2 of its items can asses usability [Lewis et al., 2013]. However, it has been shown that such questionnaires hardly account impartially for a whole session. Instead, the psychological "recency effect" emphasizes last experiences and influences results [Hassenzahl and Sandweg, 2004].

3.5.3 Physiological sensors & neuroimaging

To our knowledge there is no study involving physiological sensors or neuroimaging which account solely for usability. Because the notion of usability involves various parameters, those measures has been used instead as an *indicator*. For example workload through fNIRS [Hirshfield et al., 2009b] or frustration through EDA [Gilleade and Dix, 2004]. Eye-tracking is also a promising type of input [Vrzakova, 2013]. In conjunction with other evaluation methods, continuous recordings from physiological sensors and neuroimaging give additional insights and help to contextualize data [Pike et al., 2012].

Prior to the work that led to the completion of this thesis, there was as well hardly any study that had dealt with the evaluation of users comfort along an interaction using neuroimaging. Therefore, we tried to fill the gap with other constructs while we were evaluating a specific aspect of users comfort: physical hindrance during the viewing of 3D scenes through stereoscopic displays [Frey et al., 2014c, Frey et al., 2015]. See chapter 5 for a description of this work.

3.6 Engagement – Flow – Immersion

There is no real consensus in literature to define exactly what "engagement", "immersion", nor "flow" overlap. From [Matthews et al., 2002], task engagement is defined as an "effortful striving towards task goals". Authors add that task engagement increases during a demanding cognitive task and decreases when participants perform a sustained and monotonous vigilance task – see also [Fairclough, 2009]. In [Chanel et al., 2011] "engagement" is one particular emotion, expressed as "positive excited" in the valence/arousal model. As we can see, engagement is at a crossroads between several concepts studied in this chapter: workload, attention and emotions.

"Flow" originates from psychological studies involving challenge and/or creativity, such as sport, art or chess. It is a state in which someone is totally involved in what he is doing. Flow happens when the skills of the person meet a sufficient amount of challenge. A too important challenge brings anxiety, for too many skills it is boredom, and too few of both results in apathy [Nacke and Lindley, 2009]. Here again, several measures are involved. Challenge relates to workload and the resulting state to emotions. By definition, flow implies engagement.

"Immersion" is studied mainly in virtual reality (VR) literature. In [Slater et al., 2009] immersion stands for the modalities hardware gives to users, how well devices can preserve fidelity in VR compared to reality (e.g. display's size, input's degrees of freedom, etc.). Then the subjective feeling of belonging to VR is called "presence" – here in the sense of *virtual* presence, see [Sallnäs et al., 2000]. Unfortunately this useful distinction is less clear-cut in other papers – see [Nacke and Lindley, 2009]. If not a prerequisite to it, immersion is sometimes regarded as a synonymous for flow, e.g. [Nijholt et al., 2009].

3.6.1 Inquiries

The Dundee Stress State Questionnaire assess factors linked emotion, motivation and cognition and has been used to gauge engagement [Matthews et al., 2002]. Flow and immersion are among the seven factors assessed by the Game Experience Questionnaire (GEQ) [Ravaja, 2009]. Correlation has been observed between flow assessed via GEQ and intended level design of a first-person shooter video-game [Nacke and Lindley, 2009]. Further analysis on items split from GEQ also correlated the questionnaire to "immersion" condition and "boredom" condition (seen as the counterpart of engagement in the paper) [Nacke et al., 2010].

Specifically dedicated to immersion, the Immersive Experience Questionnaire (IEQ) [Jennett et al., 2008] measures how much users are sensible to external stimuli – an approach similar to the report of breaks in presence, when a user disengages from the task and shift his attention from virtual environment to reality [Slater et al., 2009]. Even though IEQ could differentiate between two conditions that involved a different content – 3D video-game and a 2D task in [Jennett et al., 2008] – it failed to differentiate a screen and a head-mounted display when the immersive variable related to the interface [Burns and Fairclough, 2015].

3.6.2 Behavioral studies

How users engage in an interaction reflects directly on how they perform – this is in fact the very reason why for many the quest is to seek the Holy Flow. As such, performance metrics, such as completion time, can be used to discriminate different levels of immersion [Jennett et al., 2008].

3.6.3 Physiological Sensors

Emotional states relating to frustration, stress or anxiety are linked to flow and engagement. It is hypothesized in [Gilleade and Dix, 2004] that among physiological sensors, interactions with the physical controller can help to assess frustration. In another study superfluous mouse clicking is being considered [Scheirer et al., 2002] – how users interact with the mouse or the gamepad could be assimilated to EMG measures. In the same paper up to 82% classification for frustration is achieved with a combination of EDA and blood pressure.

In [Nacke and Lindley, 2009] a more extensive study has been using facial EMG and EDA during video-game sessions with three different conditions (level design): boredom, immersion and flow. EDA measures correlated to the "flow" condition. Depending on the recorded muscle, EMG activity was significantly different between the three conditions. While studied sample is small and homogeneous, since users also filled the GEQ it is possible assert that physiological sensors are able to account for flow.

3.6.4 Neuroimaging

In neuroimaging literature [Fairclough, 2009, George and Lécuyer, 2010] engagement assessment studies are mentioned, but they often relate only to sub-components such as workload or attention. Their purpose is really to adapt the challenge level of the interaction.

Engagement is seen as a process related to information gathering, visual scanning, and sustained attention in [Berka and Levendowski, 2007]. This study managed to discriminate workload and engagement by using EEG and correlations to an engagement index measures with egocentric tools. However, the tasks involved (mental additions, recalls) are close to what is seen elsewhere in attention/vigilance protocols. Engagement is often left entangled with other states in a "performance" measure, see [Blankertz et al., 2010], sec. 3.2. Complex experiments are needed for engagement to emerge on its own, maybe by adding to its "monad" [Latour et al., 2012] emotions assessments. This situation is similar with immersion, when ERP associated to external stimuli – hence, to attention level – is explicitly used [Burns and Fairclough, 2015].

Experiments conducted during the FUGA project showed that flow could be related to fMRI measures [Ravaja, 2009]. The analysis with EEG of band frequencies shows different pattern across three conditions of interaction: boredom (i.e. not engaged), flow and immersion in a pilot study [Nacke et al., 2010, Berta et al., 2013] improved on this work and achieved a 66% classification accuracy.

3.7 User Experience

We dedicated a proper section to user experience (UX) as it is the subject of many HCI papers, but this notion entirely sits atop previously seen measures. For [Mandryk et al., 2006], UX is a shift from usability analysis, and by bringing emotions into the equation, users' entertainment is involved. UX embeds "usability / comfort", "emotions" and "engagement / flow / immersion". UX is a higher level of comprehension of what users undergo during interactions.

The project FUGA, "fun of gaming", compiled various evaluation methods in order to measure media enjoyment [Ravaja, 2009]. It is also possible to refer to UX when one intends to study the social aspect of interactions – EDA is different if the opponent in a sport game is played by a friend or a computer [Mandryk et al., 2006]. Assessing UX every time new technologies are used could guide the HCI community in its choices, e.g. with BCI [van de Laar et al., 2013].

3.8 Conclusion and pointers to subsequent work

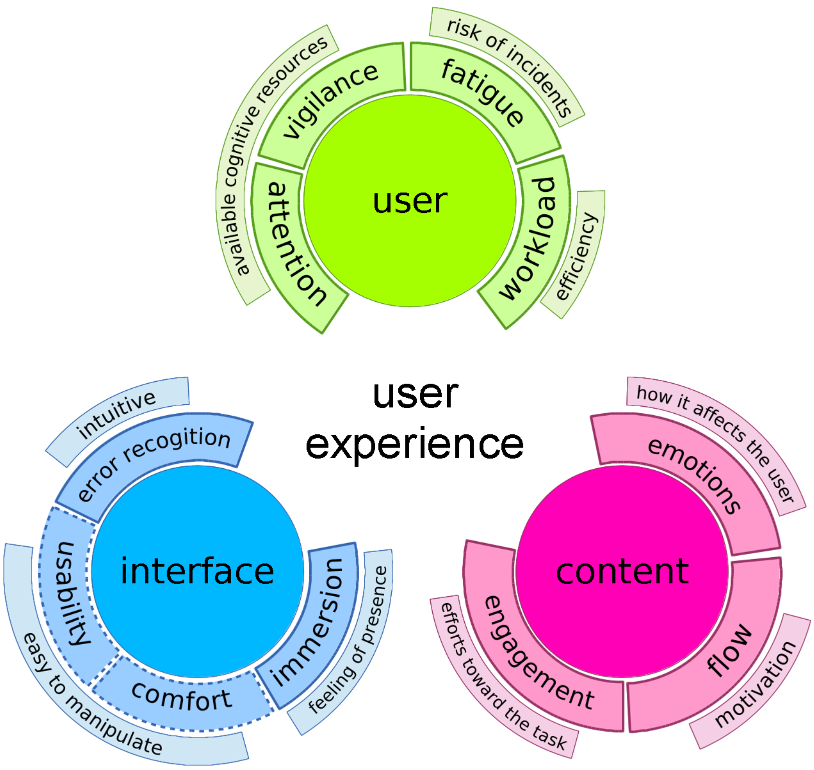

Figure 3.1 – One possible view of a simplified characterization of the constructs. In the middle circles are the constructs (dotted = not yet measurable with EEG). The inner circles represent the HCI components the most closely related to the constructs, or on which it would be easier to leverage. The outer circles give a hint about what an evaluation would be useful for.

We reviewed how neuroimaging techniques could assess constructs relevant for HCI evaluation. Between the four categories of evaluation methods, inquiries could deliver more qualitative data, while physiological sensors and neuroimaging are exocentric measures (the most "objective" measures of subjectively perceived stimuli). It is particularly interesting to combine those methods for constructs otherwise difficult to assess with exactitude, as investigated in many studies [Ravaja, 2009, Nacke and Lindley, 2009, van Erp et al., 2010, Chanel et al., 2011].

Our analysis of neuroimaging techniques focused on EEG as it promises a good trade-off between cost, time resolution and ease of installation. We advocate that neurotechnologies can bring useful insights to HCI evaluation. EEG devices are not yet perfectly reliable and practical to use; hardware and software processing are still evolving. However, their cumbersomeness is partially avoided if they are used during a dedicated evaluation phase in the HCI development process, with specially enrolled users (testers).

This review enabled us to shed light on certain issues that needed to be addressed in order to push forward the evaluation of HCI with passive BCI. These issues guided our subsequent work. Altogether with pointers to the dedicated sections relating their story, they can be summarized as follows:

3.8.1 Evaluating constructs with a comprehensive methodology

We studied workload, attention, vigilance, fatigue, error recognition, emotions, engagement, flow and immersion. Figure 3.1 stimulates thoughts about their relationships with HCI components. Some constructs should benefit more than the others from EEG measures: 1) workload, EEG being more sensible to changes compared to other methods [Mathan et al., 2007]; 2) attention, because event related potentials could help to anticipate how many details users register [Mustafa et al., 2012]; 3) emotions, with an arousal/valence state measured over a short time-frame [Chanel et al., 2011]. Error recognition could hardly be assessed precisely with anything but neuroimaging. Such construct highlights how innovative this evaluation method is. Among the outlined challenges, a continuous and modulated error recognition would greatly help to assess usability and comfort.

Next studies should start to combine the various constructs, along with a comprehensive framework which gathers every evaluation method, one's advantages preventing others' drawbacks. This should lead to an increase of the overall user experience.

As such, we first used a combination of physiological sensors and EEG to study workload during the evaluation of a novel input device, the CubTile. The CubTile multiplies the number of degrees of freedom of a classical touch surface by using 5 different sides of a cube. It has been used with a 3D manipulation task. This work is described in chapter 6.

In a second study, we dug into the evaluation of a construct that was yet to be investigated through neuroimaging techniques. Focusing this time on output device, we assessed visual comfort of users while they were watching at a stereoscopic display. In chapter 5 we demonstrate that in this situation EEG was able to monitor in near real time users' state. Besides HCI evaluation, this study also serves as a proof of concept of an adaptive system that could circumvent one of the major drawback of a promising technology.

A third study takes HCI evaluation through passive BCI to another level by giving an example of the use of our methodology during the conception from scratch of a software. In chapter 7 we described how the elaboration of a 3D maze could be the opportunity to verify various hypothesis concerning different kinds of interaction technique with neuroimaging techniques. We recorded workload, attention and error recognition and match the dynamic of these measures with changes we provoked during the interaction.

Over the course of this thesis we describe extensively the different "bricks" that we used, developed and/or improved. We hope that the description of such a backbone will help to disseminate the use of passive BCI, either for HCI evaluation of in other settings, as later parts of the thesis bring EEG outside the lab, in public or social settings.

3.8.2 Two directions to make signal processing more reliable

Improvements in signal processing, either in features extraction or classification, could benefit every technology. Constructs, such as emotions, are not yet accurately assessed with pure EEG signals. When too many classes (e.g. emotions and workload levels) are assessed altogether, the classifier performance drops – e.g. see how the "curse-of-dimensionality" relates to classifiers' complexity [Friedman, 1997]. Improvements in mathematical analysis and machine learning algorithms, as well as a better understanding of brain activity, would increase the reliability of the whole system by a great amount and favour every construct.

This was for the *machine* learning part. We went for another direction, opposite in some aspect, by investigating in appendix E how we would guide users while they interact with a BCI in order to get more reliable EEG signals, giving them a feedback about how tensed their muscles were. The more relaxed the better.

3.8.3 Jumping into open hardware movement to craft affordable EEG headsets

Some limitations observed in EEG research are yet to be resolved to make EEG-based evaluation of HCI more operable. EEG devices, while practical compared to other neuroimaging techniques, take long to setup. Hence, experiments can be tedious both for the experimenter and for the participant. This is why there are often only few participants during EEG or BCI experiments, which is a problem for the reliability of the results. EEG signals contain many potential artifacts (e.g. muscular activity and electrical parasites); the quality of the device is essential. EEG signals must be calibrated, processed and interpreted carefully.

Since a few years new EEG devices have appeared, oriented toward a larger public. Their electrodes use no conductive solution, or water as solvent. These electrodes are faster to set-up – no more gel to be put on each one after the device has been installed – but may be less sensitive, see [Blankertz et al., 2010], sec. 2.1. Hence, some companies, while transforming EEG into a mass-product, bring less reliable technology to the market. Those devices often possess fewer electrodes. Without a cap the electrodes are difficult to place in a standardized position on the scalp. Finally, they are often packaged with software development kits which hide the signal processing from the users. Constructs like attention or emotions are then claimed to be directly measured, without further justification or muscular artifact control, see [Heingartner, 2009]. Nevertheless, while experimenters must be aware of such limits if their intent is to rely solely on brain activity, this increasing appeal in favor of cheap EEG devices is a great opportunity to push forward the use of neuroimaging in HCI.

Enthusiasts coming from the DIY (do it yourself) movement came to the building of an EEG headset based an Arduino board: OpenBCI. It occurred during the achievement of this thesis and we took this opportunity to be among the first to test such solution, participating in its software integration. We have studied this hardware and compared it to medical grade amplifiers. In the Appendices we describe how a cheap alternative, costing a fraction of the price the products seen in the marked up until then, compete to what is sold by established actors within EEG suppliers. It may well be the perfect trade-off between cost and reliability that we were hoping for at the beginning of our work in order to see the HCI community grasp neuroimaging.

Not only did we test and compare hardware that were already available, but we also crafted our own sensors. Indeed, we saw an opportunity to finish the job and obtain a EEG headset practical to use, that we used to deploy quickly an installation in a public setting. We gave back to the open-hardware community by describing all the process and giving away every details and files that led to our solution. Of course the performances that we could achieve at this moment are not as good as with medical grade devices, far from it, but this proof of concept helped to ease the acceptability of EEG devices. This story takes place in part IV.

PART II

BRINGING EEG AS AN EVALUATION METHOD

Now that we reviewed what mental states could be assessed and which tools are at our disposal to do so, we will put into practice this knowledge, first by travelling into the realm of human-computer interaction (HCI). In this part, we use brain-computer interfaces (BCI) by the mean of electroencephalography (EEG) in order to evaluate beforehand various HCI components.

In chapter 5, we study how users' comfort vary depending on the extent of stereoscopic effects when "3D" displays are used. This a proof of concept of the first BCI that could discriminate in near real time between different visual comfort conditions – an revised version of works previously published in [Frey et al., 2014c, Frey et al., 2015, Frey et al., 2016a].

We also describe how a BCI could monitor the workload induced during a 3D manipulation task in chapter 6 – appeared in [Wobrock et al., 2015]. We improved on this work in chapter 7, where we propose protocols and tools to assess workload, attention and error recognition. We compared two interaction techniques (keyboard vs *touch) during a 3D navigation task that we built from scratch. We tailored the virtual environment so as to validate the use of EEG as an evaluation method for HCI.*

In order to fully grasp the pipelines that will be utilized in those three chapters, we start by detailing further the grand principles behind BCI.

4

BCI 101: THE BASICS

BCI, brain-computer interfaces, are communication devices between humans and machines that rely only on brain activity (i.e. no muscular input) to issue commands or monitor states [Wolpaw et al., 2002]. BCI is an emerging research area in Human-Computer Interaction that offers new opportunities for interaction, beyond standard input devices [Tan and Nijholt, 2010].

The "interface" term covers many different areas of applications. As soon as a command originating from recordings of brain activity is issued to the computer, the process could be called a BCI. Sometimes the term "brain-machine interface" (BMI) is used. Nowadays, BCI and BMI both designate the same thing . Each originate from a different community. BMI comes from the neuroscience field and the people using invasive techniques, where brain interfaces where historically studied to command prosthesis. BCI comes from computer science, and at the beginning it was bound to non invasive interfaces, that did not required surgery. Fortunately the technical skills and the knowledge involved have merged years ago and all scientists or so work happily together under the "BCI" flag.

The different technologies that could be used to sense brain activity have been described in section 1.4. Basically, EEG and – to a less extend – fNIRS are used as non-invasive techniques. In few research projects that deals with invasive technologies applied to humans, electrodes arrays (ECOG, electrocorticography) are used in some specific applications with humans, e.g. motor tasks [Leuthardt et al., 2006], that demonstrates how invasive technologies enable better performances. There is but very few examples of humans implanted with deep brain electrodes, even though this is a promising technique for people with motor impairments, such as tetraplegia [Hochberg et al., 2006]. Deep brain electrodes are at the moment mostly employed with animals to study neurological diseases or to prepare the ground for human BCI. Electrodes implanted in human brain may be used to stimulate neurons but then there is no BCI involved, "just" a therapy, e.g. to suppress motor tremor in Parkinson's disease [Benabid et al., 2009].

Most of the BCI studies seen in the literature and described in this thesis use EEG, but the other way round is not systematically true: it is not because a system uses EEG that it is a BCI. EEG could just be used as a diagnostic tool, without issuing commands, with no feature extraction or classification. It may be a little difficult to pin-down an exact definition of what a "BCI" is and what it is not – as it goes with all trendy terms, it is employed more often that it should be [Allison, 2011]. The task is difficult, but we will try to face the challenge.

First, a BCI needs to rely on brain activity. It is not that obvious to everybody, really, even if the "brain" word is comprised in the acronym. Brain activity does not mean muscle activity, frown your eyebrow activity, eye blinks activity, be very focus and clench your teeth activity. It is *brain* activity, messages that are inferred directly (may it be current, oxygen consumption, electromagnetism) from firing neurons. OK, this part may be polemical, we never know really what we record. Even if we are very precocious, something more than brain activity is quick to slip through signal processing, and many confounding factors can deter what is truly measured with neuroimaging [Brouwer et al., 2015].

Let's talk about the second requirement of a BCI system: a feedback. That is to say a command issued to the computer from the brain activity. At least that was the case at first, when BCI were all *active*, when brain activity was processed in real time (minus delays imposed by computations) and that the users were issuing consciously commands, for example *imagining* hand movements to move inside a 3D environment [Lotte et al., 2010]... or feet movements to lift a (virtual, alas) space ship [Lotte et al., 2008].

More recently, other kinds of BCI applications emerged, that do not required users to consciously issue command but that toggle commands or adapt software depending on the basal state of users. These are called *passive* BCI, BCI not used as input to HCI [Zander and Kothe, 2011, George and Lécuyer, 2010]. Passive BCI could be used to build adaptive systems, for example by adding details to an air traffic control software if brain recordings show that users can handle more information [Abbass et al., 2014]; or by supporting users with an interface easier to manipulate when the workload is too important [Afergan et al., 2014]. BCI "users" could even be part of a signal processing pipeline, as for projects that show rapidly numerous satellite images and seek within EEG the event-related potentials that arise when operators see a missile silos [Sajda et al., 2010] – kind of subliminal reactions. To think about users as computational units does not exactly please me, but in a sense it is a way to create a hybrid between flesh and silicon. Not necessarily a bad thing once the tool is given to people a bit more creative – designers or architects that construct new shape by listening to their sub-conscious [Cutellic and Lotte, 2013]; or when it acts as an explicit communication channel between people, as I tried to investigate ultimately in part IV with physiological sensors at large.

Between passive and active BCI there is an available slot: reactive BCI [Zander and Kothe, 2011]. In reactive BCI, the brain activity triggered by external stimuli is indirectly modulated by users so as to control the application. One of the most famous BCI belongs to this category: in the "P300 speller" letters that randomly flash on the screen can be used to spell words with the sole brain activity. When the letter that the user wants to spell flashes, a particular event-related potential arises within the EEG, which possess a positive "peak" around t=300 ms after the stimulus onset – this is commonly referred to as the "oddball paradigm". A promising application, but hold your horses before you get rid of your keyboard, even with online correction the speed is in the range of 1 word per minute [Schmidt et al., 2012].

It is sometimes difficult to separate active, passive and reactive BCI. The level of awareness or control of users is not always obvious. For example, in the area of affective computing, games have been modified to adapt the players avatar to the player emotional states. In a lab version of World of Warcraft, the nice elves shifted in a terrible bear when the users get stressed [Nijholt et al., 2009]. Although this can be described as a passive BCI, users learn how to control their activity, and during combats they could want to deliberately enrage so that their character could fight back.

An active BCI is traditionally represented by the full loop of Figure 4.1, with users commands detected by brain imagery and signal processing and a feedback given. A passive BCI will only lack one step in the process, the feedack that may be absent (when the BCI occurs afterwards) or unconscious (with adaptive systems). A BCI is what sounds BCI and looks BCI – and uses brain activity. In the end a BCI could be seen as a tool to comprehend users brain activity.

Machine learning is widely used with BCI because these is an important variability between people's brain and brain patterns, and many external factors that could influence brain recordings (amplifier's specifications, electrodes exact location, ...). As such, it is difficult to make a strong assumption between a set of features and a given mental state, one that could carry on between sessions and between participants. With this approach, a calibration phase occurs so that the system could learn which features are associated to a specific individual. In this part we will make use of machine learning to select features that come from EEG signals. There is three main types of information:

- Frequency domain: oscillations that occur when large groups of neurons fire altogether at a similar frequency

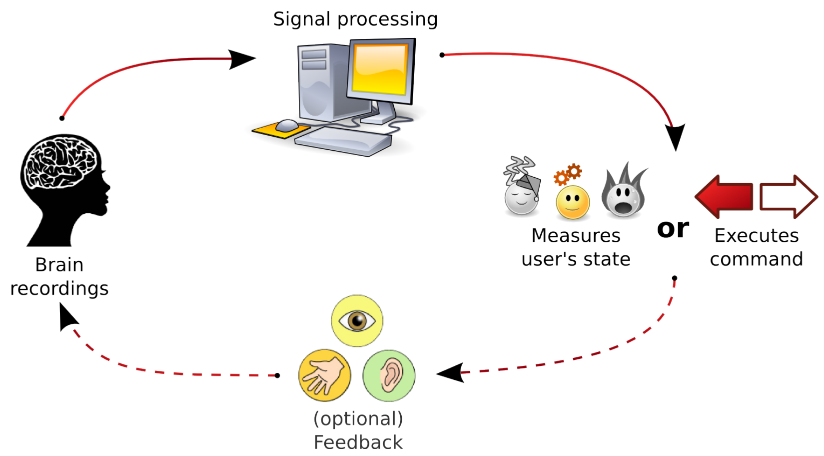

Figure 4.1 – The classic active BCI loop. 1: user's brain activity is recorded, 2: signal processing detects features, 3: user's mental state is detected or a command is issued, 4: a feedback is given to the user. Passive BCI: there is no feedback given to users, yet the system monitors continuously brain activity.

- Temporal information: event-related potentials (ERP) posses temporal features; positive and negative "peaks" with varying amplitudes and delays. Note that while ERP cover self-induced signals (e.g. motor preparation), the term "evoked potentials" (EP), which is also found in the literature, is restrained to activities that arise from external stimuli.
- Spatial domain: position of the electrodes on the scalp that record a specific brain activity. It is possible to reconstruct the source of the signals thanks to an inverse-model, a mathematical model that describe how the signal is diffused. The connectivity between sets of electrodes could be studied as well.

A BCI *must* use brain activity, but if we do not focus on fundamental research, could as well use everything that is recorded by EEG (EOG and EMG included) to extend the possibilities and improve the performances – when muscular artifacts help the system and do not disrupt the signals. Machine learning could be used the same way with other physiological signals, the resulting system being a hybrid BCI [Zander, 2011]. We described it at several occasions in section 3 and this is what we studied while we evaluated a novel interaction technique in 6. Well, it is not always beneficial to combine physiological sensors, but contrary to the impressions we may give until now by focusing as much on clean EEG signals, we welcome every churches in the happy world of physiological computing.

It is implicit that BCI work in real time. In this part, EEG activity is used in order to evaluate HCI. But the evaluation has nothing to do with the interaction when it occurs. It is a tool that will be used afterwards to correct poor choices. EEG enables continuous measures that do not need to take place at the same time of the interaction. It is a "time-shifted live recording", that utilizes almost all the workflow employed in BCI, except for the feedback given to users.

The studies that follow put intro practice the signal processing pipeline described in this chapter. If we consider BCI applications where users are not aware of the ongoing modification of the system, the works described in this part lie at the frontier of passive BCI's definition. For instance, the results gathered during the evaluation of visual comfort pave the way for an adaptive system that could ease users' experience with stereoscopic displays.

5 Visual comfort with stereoscopic displays

With stereoscopic displays a sensation of depth that is too strong could impede visual comfort and may result in fatigue or pain. We used Electroencephalography (EEG) to develop a novel brain-computer interface that monitors users' states in order to reduce visual strain. We present the first system that discriminates comfortable conditions from uncomfortable ones during stereoscopic vision using EEG. In particular, we show that either changes in Event Related Potentials (ERP) amplitudes or changes in EEG oscillations power following stereoscopic objects presentation can be used to estimate visual comfort. Our system reacts within 1s to depth variations, achieving 63% accuracy on average (up to 76%) and 74% on average when 7 consecutive variations are measured (up to 93%). Performances are stable (≈ 62.5%) when a simplified signal processing is used to simulate online analyses or when the number of EEG channels is lessened. Even though at the moment only very specific stimuli are considered, this study could lead to adaptive systems that automatically suit stereoscopic displays to users and viewing conditions. For example it could be possible to match the stereoscopic effect with users' state by modifying the overlap of left and right images according to the classifier output.

This work first appeared in [Frey et al., 2015] in a shorter format. The present chapter contains most of the additional analyses that were published afterwards in [Frey et al., 2016a], as well as some details that appeared in a previous pilot study [Frey et al., 2014c].

I thank Léonard Pommereau and Aurélien Appriou for their help – more in the appendices, section Credits.1.

5.1 Stereoscopy at risk

Stereoscopic displays have been developed and used for years in computer science, for example to improve data visualization [Frohlich et al., 1999, Drossis et al., 2013], to ease collaboration between operators [Salzmann et al., 2009] or to better manipulate virtual objects [Hachet et al., 2011]. However, it is only during the past decade that this technology began to reach users beyond experts. Notably, movie theaters – and the entertainment industry in general – helped to popularize so-called "3D" contents. Nowadays stereoscopic displays are used at home. "3D" television sets gain in popularity and game devices started to use this technology. Yet, whenever devices use shutter or polarized glasses[1], parallax barrier[2] (e.g. Nintendo 3DS) or head-mounted displays (as with the Oculus Rift) to produce pairs of images, visual discomfort could occur when the stereoscopic effect is too strong. Some viewers could even feel pain [Lambooij et al., 2009].

In order to mitigate those symptoms and adapt the viewing experience to each user, we propose an innovative method that can discriminate uncomfortable situations from comfortable ones. It reacts quickly (within 1s), without calling upon users, so it does not disrupt the viewing.

Our solution is versatile because all stereoscopic displays use the same mechanism to give the illusion of depth. They send a different image to the left and right eyes. As with natural vision, the visual fields of our eyes overlap and the difference between the two images helps our brain to estimate objects' distance.

To facilitate images merge, observers rely on two mechanisms. First, they need to maintain the point of interest at the same place on both their retinas. This is why the closer an object gets, the more eyeballs rotate inward. This is called "vergence", and it also happens with stereoscopic displays. Second, in a way similar to how camera lenses operate, crystalline lenses need to focus light beams. They deform accordingly to objects' position in order to obtain a clear picture. This other physiological phenomenon is called "accommodation" and is *not* replicated with stereoscopic displays.

Besides stereoscopy, other mechanisms help to asses depth, such as shadows or parralax.

In a natural environment, vergence and accommodation are locked to objects' positions and occur altogether. However, since the focal

[1] respectively "active" and "passive" displays
[2] "autoscopic" displays

plane in stereoscopic displays is fixed, accommodation will not change. No matter how far or how close virtual objects appear to be, physical screens remain at the same place. The discrepancy between vergence and accommodation is called the “vergence-accommodation conflict” (VAC, see Figure 5.1). It causes stress on users [Lambooij et al., 2009]. The closer or further a virtual object gets compared to the display plane, the stronger this conflict is. When it is too important or lasts too long, visual discomfort occurs.

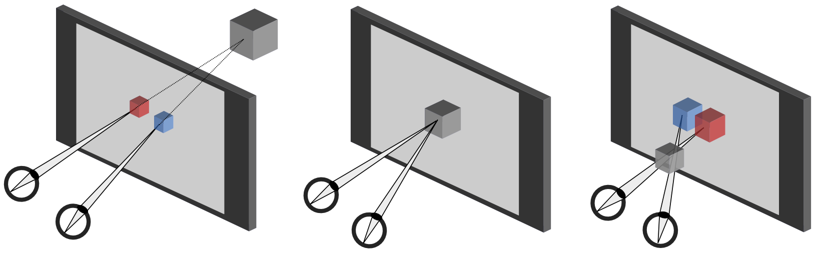

Figure 5.1 – The vergence-accommodation conflict (VAC). *Left*: object “behind” the screen, negative VAC. *Middle*: the object appears flat, no VAC. *Right*: object “in front”, positive VAC.

VAC is one of the major causes of the symptoms associated to visual fatigue in stereoscopic displays [Hoffman et al., 2008, Lambooij et al., 2009]. Guidelines exist to limit the VAC and prevent such negative effects. In particular, Shibata et al. [Shibata et al., 2011] established a “zone of comfort” using questionnaires, a zone within which the apparent depth of objects should remain to avoid discomfort (see Figure 5.2). It takes into account the distance between viewers and displays. Unfortunately individual differences [Lambooij et al., 2009] make it hard to generalize such recommendations and use them as is. Besides, viewing conditions vary. Ambient light, screen settings, viewing angle and stereoscopic techniques are parameters among others that influence the rendering and as such alter visual strain [Bangor, 2001]. It may then be interesting to back up the VAC with another type of measure than sporadic and disruptive questionnaires.

As seen in part I, new investigation techniques record users’ physiology. Complementary to qualitative questionnaires, as used in [Shibata et al., 2011], brain activity recordings enable the monitoring of users states. Not only such technology could give continuous insights, it also makes measures without interrupting or disrupting the interaction, here the viewing. In [Gaeblerlabel et al., 2014], authors demonstrate with functional magnetic brain resonance imaging (fMRI) how stereoscopy increases inter-subject correlation of several neural networks, overlapping data with the time course of a movie, and how it reflects immersive tendencies reported via questionnaires.

Electroencephalography (EEG) is among the cheapest and most lightweight devices that sense brain signals. Even though EEG has been used to investigate visual fatigue induced by stereoscopic display

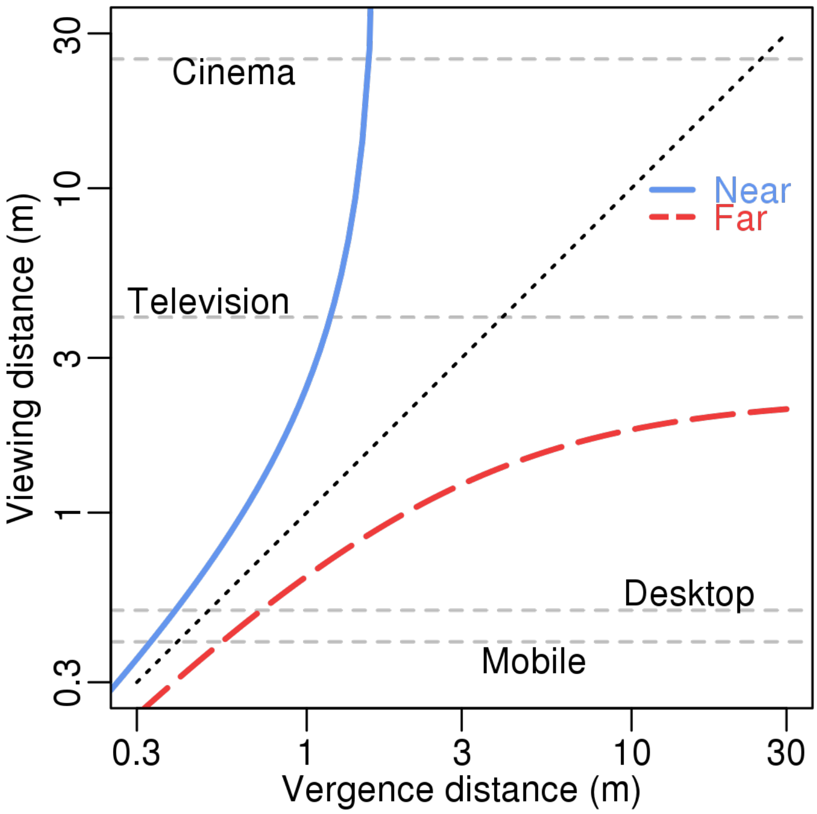

Figure 5.2 – The acceptable zone of comfort depending on viewing distance and vergence distance – i.e. the apparent depth of contents. From [Shibata et al., 2011]

[Li et al., 2008, Cho et al., 2012, Chen et al., 2013, Bang et al., 2014], those studies only compared flat images with stereoscopy. They do not control for objects virtual positions, hence they cannot account for different comfort conditions. Furthermore, most of the EEG studies related to stereoscopic display and fatigue analyzed stimuli which last several minutes (e.g. from 3 to 40 min in [Li et al., 2008, Chen et al., 2013, Bang et al., 2014]). Such protocols could not lead to adaptive systems that react quickly.

In a first pilot we studied the zone of comfort continuum through EEG. In [Frey et al., 2014c], we conducted a preliminary investigation that compared short appearances of virtual objects. Results tended to show that the brain activity induced by stereoscopic displays was different whether objects were presented within the zone of comfort or not. There ware significant differences both in event-related potentials (ERP) and in frequency bands power. An uncomfortable stereoscopy correlates with a weaker negative component and a delayed positive component in ERP. It also induces a power decrease in the alpha band and increases in theta and beta bands.

Following this work, we tested the accuracy of a system that classifies EEG data to measure visual comfort. Our main contribution is to prove the feasibility of an EEG system that could estimate in near real time (1s delay) the visual comfort viewers are experiencing as they watch stereoscopic displays. It could be adapted to real-case scenarios by controlling the discrepancy between left and right images depending on the output of the classifier. Then it could be employed in different settings to improve HCI by easing users' comfort, for example when they manipulate 3D contents during prolonged periods of time – such as remote design or video games – or when people are watching 3D movies – especially when there are many rapid depth variations, as seen in action sequences.

5.2 Experiment

5.2.1 Overview

We studied the appearance of virtual objects. They were presented to participants at different apparent depths for a few seconds (see Figure 5.3). We studied two conditions: objects appeared at a comfortable position ("C" condition) or at an uncomfortable position ("NC" condition).

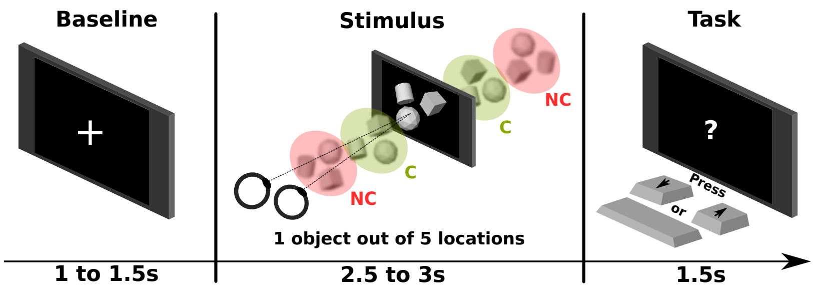

Figure 5.3 – One trial: cross (baseline), object at random depth, task.

We displayed simple grey objects over a black background. Three kinds of primitives were employed: cube, cylinder (32 vertices) and icosphere (80 faces). The primitives varied in shapes as curves and surfaces size are important for objects comprehension [Champion et al., 2004]. Objects orientation were randomized along the three axes, rotations producing more stimuli [Norman et al., 2009]. Rotations were controlled so as cube and cylinder faces couldn't be orthogonal to the camera plan, thus preventing the appearance of artificial 2D shapes. The resulting 3D scenes were kept simple enough to ensure that there were no distracting elements and that no variables beside the VAC were manipulated. We deprived the depth cues to control for VAC. For example casting shadows would have helped to differentiate close objects from far objects without the need of binocular fusion [Mikkola et al., 2010].

We defined ranges inside and outside the zone of comfort that match the equations established by [Shibata et al., 2011]. Related to the location of participants sitting 1m away from the display, in "C" condition virtual objects were positioned within [0.75m; 0.85m] (comfortable close) or within [1.3m; 1.6m] (comfortable far). In "NC" conditions, ranges were [0.35m; 0.45m] (uncomfortable close) or [4m; 6m] (uncomfortable far). During one third of the trials, objects appeared "flat" (no stereoscopic effect, 1m apparent depth, as far as the screen).

In order to assess their capacity to situate virtual objects in space and to maintain their vigilance high during the whole experiment, participants had to perform a task. When a question mark was shown on screen, "down" arrow, "space" bar or "up" arrow were pressed to indicate whether objects appeared "in front of", "as far as" (flat images) or "behind" the screen. With both hands on the keyboard, choosing those keys to answer ensured that participants' gaze was not leaving the screen and that participants' movements would not pollute EEG signals.

The participants were also motivated to complete the experiment and reach the best score (the number of correct responses). They were told that the one with the best score during the task would won a bottle of wine[3].

A trial started with a neutral stimulus, a 2D cross appearing on-screen for 1 to 1.5s. Then the virtual object appeared for 2.5 to 3s. Finally, a question mark appeared for 1.5s, a period during which participants had to perform the task. After that, a new trial began. This sequence is illustrated in Figure 5.3. The first two time intervals, that randomly varied by 0.5s, prevented participants to anticipate objects appearance and the moment they had to respond to the task. On average a trial took 5.5s. All in all there were 160 trials per C and NC conditions, randomly distributed. Trials were equally split across 4 sub-sessions to let participants rest during the investigation and avoid a too tedious experiment.

5.2.2 Apparatus

Stereoscopic images were shown in full HD resolution (1080p) on a 65 inches Panasonic TX-P65VT20E, an active display – participants wore shuttered glasses. The software that rendered the virtual objects was programmed with Processing framework, version 2.2.1. Objects were dynamically created. No matter their apparent depths, primitives sizes on screen remained identical: they were scaled within the virtual scene. In combination with a diffuse illumination of the scene, this made it impossible to discriminate conditions without stereoscopy. The interpupillary distance used to compose stereoscopic images was set at 6cm, an average value across population [Dodgson, 2004].

EEG signals were acquired at a 512Hz sampling rate with 2 g.tec g.USBamp amplifiers. This medical-grade equipment handles 32 elec-

[3] Bordeaux city makes us do this kind of promises, even if in the end it ended up with a beer fest for everyone.

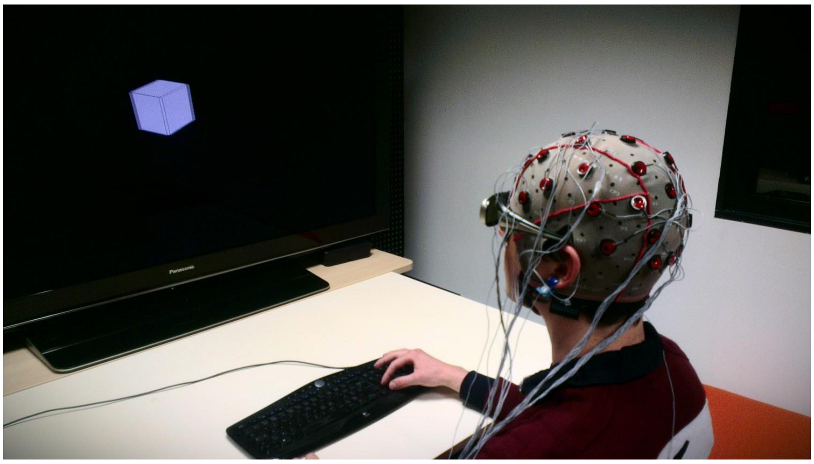

Figure 5.4 – Setup of the experiment, with a participant being presented with stereoscopic images while his EEG signals are being recorded.

trodes. We used 4 electrodes to record specifically electrooculographic (EOG) activity and 28 to record EEG. In the international 10-20 system, EOG electrodes were placed at LO1, LO2, IO1 and FP1 sites; EEG electrodes were placed at AF3, AF4, F7, F3, Fz, F4, F8, FC5, FC1, FC2, FC6, C3, Cz, C4, CP5, CP1, CP2, CP6, P7, P3, Pz, P4, P8, PO3, PO4, O1, Oz and O2 sites. OpenViBE 0.17 recorded both electrodes signals and key strokes of the task.

OpenViBE was also used to trigger images appearance in Processing. To do so, TCP messages were sent from OpenViBE to Processing. The same machine ran both programs, thus TCP latency was negligible (less than 1ms). 3D rendering on Processing side could necessitate some CPU cycles, though, and event-related potentials (ERP) analyses suffer from bad synchronizations. This is why we took extra precautions to accommodate rendering delays and ensure a reliable synchronization between objects appearance and EEG recordings. Processing framerate was reduced down to 25 FPS and a 60ms interval was set between TCP messages interception and the appearance of a new image – a sufficient time for the machine to make the virtual rendering and avoid lags. Overall, this mechanism ensured a constant 100ms delay between sent messages and images appearance[4]. The whole setup can be seen in Figure 5.4.

5.2.3 Participants

12 participants took part in the experiment: 5 females, 7 males, mean age 22.33 (SD=1.15). They reported little use of stereoscopic displays: 1.91 (SD=0.54) on a 5-point Likert scale (1: never; 2, 3, 4, 5: respectively several times a year, month, week or day). If applicable, participants

[4]Here we dealt with static images, we took another approach regarding ERP synchronization in chapter 7

wore their optical corrections – there was enough space beneath the shutter glasses for regular glasses not to disrupt user experience.

We made sure that no participant suffered from stereo blindness by using a TNO test [Momeni-Moghadam et al., 2012]. We created a computerized version of this test to ensure that their ability to perceive stereoscopic images was on par with our equipment, as advised in [Gadia et al., 2014].

5.2.4 Measures

Beside EEG measures, task scores were computed from participants' assessment of objects' virtual position in space – whether they appeared "in front of", "as far as" or "behind" the screen. During the 1.5s time window question marks appeared, the first pressed key, if any, was taken into account. A correct answer resulted in 1 point, an incorrect in -1 point and none in 0 point. Final scores were normalized from [-480;480] to [-1;1] intervals.

A questionnaire inquiring the symptoms associated with the different apparent depths preceded first trials and followed each sub-session. There were 2 items, one asking about participants vision clarity and the other about eyes tiredness. The corresponding 5-point Likert scales were adapted from [Shibata et al., 2011] and translated to French, "1" representing no negative symptoms and "5" severe symptoms. We measured respectively how well participants saw the stereoscopic images and how comfortable they felt; to do so we averaged the answers (10 values per item and per C/NC conditions).

5.2.5 Procedure

The experiment occurred in a quiet environment, isolated from the outside, with a dimmed ambient light. The whole experiment was approximately 90 minutes long and comprised the following steps:

1. Participants entered the room. They were seated 1m away from the stereoscopic screen (distance from their eyes), next to a table. They read and signed an inform consent form and filled a demographic questionnaire.

2. The stereoscopic display was switched on and participants stereoscopic vision was assessed with a TNO test.

3. EEG cap was installed onto participants' heads and we ensured reliable EEG signals recordings.

4. The "symptoms" questionnaire was given orally, experimenter manually triggering objects appearances. There was 1 object per virtual depth range (C close/far, NC close/far) and 2 flat objects; making 6 randomized objects per questionnaire.

5. A training session occurred. During this session participants had the opportunity to get familiar with the trials and with the task. We waited until participants felt confident enough and were ready to proceed with the main part of the experiment.
6. The 4 sub-sessions, described previously, occurred. When a sub-session ended, participants were given again the questionnaire of step 4 before they could rest, drink and eat. Once they felt ready, we pursued with the next sub-session.

We made sure that the participants were not suffering for any sickness before they left the experimental room.

5.3 Analyses

Because we want to increase fundamental knowledge on brain activity, we were particularly cautious to base our analyses on "clean" EEG signals, that is to say on EEG signals not polluted by artifacts such as eye movements [Fatourechi et al., 2007]. The signal processing that we present in this section, uses state-of-the-art tools to remove such artifacts. In the results section we will explain how the use of a simplified pipeline – one that could be easily applied online in real-life scenarios – has little impact on performance

5.3.1 EEG signal processing

We used EEGLAB 13.3.2b and Matlab R2014a to process EEG signals offline. Data gathered from the 4 sub-sessions were concatenated. We applied a 0.5Hz high-pass filter to correct DC drift and a 25Hz low-pass filter to remove from our study signal frequencies that were more likely to be polluted by muscle activity. We extracted the 320 epochs – "slices" of EEG – around C and NC stimuli onsets, from -1s to +2.5s.

Due to the important amount of data (3840 trials across our 12 participants), we chose automated methods to clean the signals. The EEGLAB function $\text{pop}_{\text{autorej}}$ removed epochs that contained muscular artifacts. Following the results obtained in [Ghaderi et al., 2013], EOG activity was suppressed from the signal using the ADJUST toolbox 1.1 [Mognon et al., 2010]. After an Infomax independent component analysis, we rejected components that ADJUST labelled as eye blinks or eye movements (vertical and horizontal).

An event-related potential (ERP) corresponds to one or more "peaks" in EEG recordings, associated with an event – in our case the appearance of stereoscopic images. Averaged ERPs across participants indicated that ERPs had a higher positive peak in C (see Figure 5.5).

There were some differences in EEG oscillations – event-related spectral perturbations (ERSP), depicted in Figure 5.6. Overall, there may be notably both a decrease of signal power within the alpha band ($\approx$

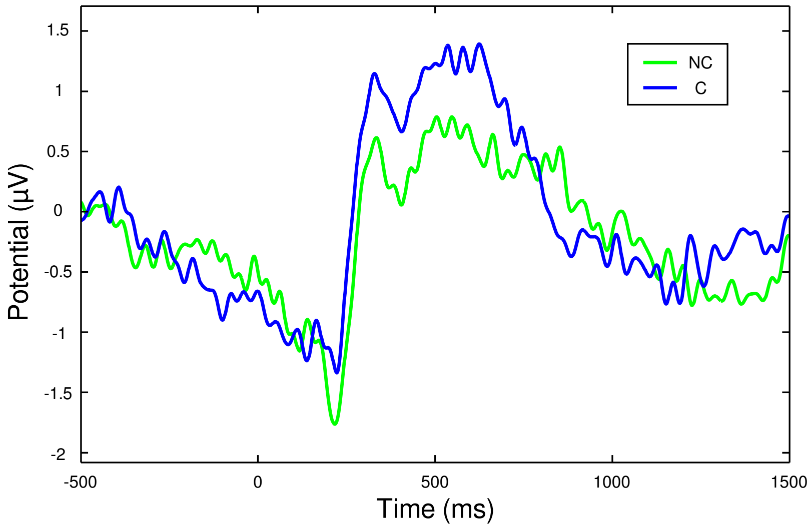

Figure 5.5 – Average ERP across 28 EEG electrodes and 12 participants. Blue: Comfort condition; green: No-Comfort condition (≈ 160 trials each). The stereoscopic object appears at t=0ms.

7Hz - 13Hz) and an increase within the theta band (≈ 4Hz - 6Hz) in non-comfort condition. Based on these findings over averaged trials, we employed spectral domain information with different features extraction methods and different classifiers for single trial classification. The benefits derived from the combination of temporal (ERP) and spectral (band power) characteristics were minor compared to the growing complexity of the underlying signal processing. For the sake of the argument, we preferred to detail a more intelligible framework in this section and to relegate a brief description of the combination of features in the Results section. This is why our classification strategy solely relies on temporal information when we compared different pipelines – e.g. Monte Carlo simulations, pseudo-online and reduced number of electrodes, see below.

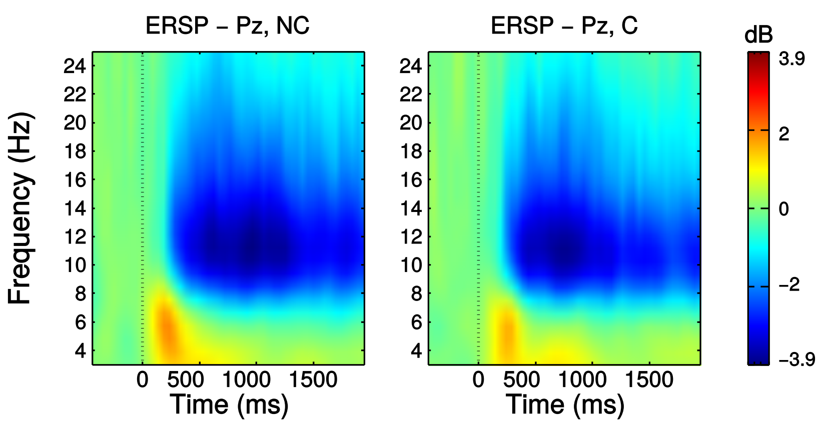

Figure 5.6 – Average ERSP in Pz (medial parietal region). *Left*: No-Comfort condition; *right*: Comfort condition (≈ 160 trials each, 12 participants).

II.5

5.3.2 Classification

We used a common pipeline to classify EEG signals. Basically, it consists in extracting relevant signal features, training the classifier on a certain set of data – it corresponds to a "calibration" phase – and then testing the classifier performances on unseen data, which simulates a real-case application.

We split in two the EEG dataset of each participant. The first half of the trials was used as a training set, and the second half as a testing set. This distribution facilitates the comparison between offline and online signal processing. In order to utilize temporal information, features extraction relied on regularized Eigen Fisher spatial filters (REFSF) method [Hoffmann et al., 2006]. This spatial filter, specifically designed for ERPs classification, reduced signals dimension from 28 EEG channels to 5 "virtual" channels whose signal is more discriminant between conditions. Note that we did not include in our study the 4 channels that were specifically recording EOG activity.

II.5

We selected a time window of 1s, starting at t=100ms to accommodate the fixed delay with objects appearances (see Apparatus). In order to reduce the number of features, we decimated the signal by a factor 16. As a result, there was 160 features by epoch (5 channels $\times$ 512Hz $\times$ 1s / 16). We used shrinkage LDA (linear discriminant analysis) as a classifier [Ledoit and Wolf, 2004]. Shrinkage LDA algorithm is more efficient compared to regular LDA when it comes to a high number of features [Blankertz et al., 2011].

5.3.3 Simulating longer stimuli with Monte Carlo

Although we used 1s time windows as a basis for our analyses, we wanted to go beyond and test longer stimuli by aggregating trials.

We could not use directly the data we gathered because in our experimental protocol conditions were randomized. So we had to simulate. We used Monte Carlo simulations to cluster trials. The principle is as follows: studying 3 presentations, we cluster 3 similar trials drawn from the testing set (e.g. "no-comfort", 3xNC). Then we look at individual classifications from the system (e.g. NC-NC-C) and keep the label which has the majority – in this case NC, the resulting classification is correct for this cluster. Had the classifier labelled trials as "C-C-C", "NC-C-C", "C-NC-C" or "C-C-NC", the cluster would have been erroneously labelled as "C".

Different combinations of trials were drawn from the testing set to compute the scores for n=3,5,7. Monte Carlo simulations served two purposes. On the one hand, it simulates the behavior of the classifier over a longer sequence of identical stimuli. On the other hand – and reciprocally – it allows the experimenter to suit the stimuli to the performance she or he wants to obtain for the desired use-case. Indeed, with a “n” as big as one want, the trade-off between accuracy and exposure time could be freely chosen.

5.4 Results

5.4.1 Task & symptoms questionnaires

We used a Wilcoxon Signed-rank test to compare task scores between C and NC conditions (means: 0.45 *vs* 0.40). There was no significant effect ($p = 0.78$).

A Wilcoxon Signed-rank test showed a significant effect of the C/NC conditions on both symptoms items ($p < 0.01$). Participants reported more eye comfort (means: 2.41 *vs* 3.46) and more vision clarity (means: 2.10 *vs* 3.13) in C than in NC.

II.5

5.4.2 Classification

Table 5.1 – Classifier accuracy (in percentage) for every participant. Mean: 63.30%, SD: 7.64.

Participant	1	2	3	4	5	6
Accuracy	54.17	59.23	58.22	70.32	60.53	64.19
Participant	7	8	9	10	11	12
Accuracy	62.91	76.06	72.46	71.52	53.24	56.74

We were able to predict with an average accuracy of 63.30% (SD=7.64) the visual comfort experienced by viewers (see Table 5.1). We studied further this first result on 3 different aspects: we used Monte Carlo simulations to improve performances over longer stimuli; we investigated how the classifier behave when only half of the EEG electrodes are employed and finally we simulated an online analysis to assess performance in a real-life scenario. Those results are detailed below and summarized in Table 5.2.

5.4.2.1 *Monte Carlo simulations*

With Monte Carlo simulations, we investigated how the system would perform with the appearance of several images from the same condition. Classifier accuracy reached 68.91% (SD=10.32) over 3 trials. Over 5 trials the classification reached 90% for some users, resulting in a 71.83% average (SD=12.28). With n=7, one-third of the participants reached 90% or more (74.08\ on average, SD=13.39). See Figure 5.7.

5.4.2.2 *Channels' contribution – accuracy over 14 channels*

EEG device that possesses fewer electrodes would be more comfortable to wear, faster to setup – i.e. more practical – and less expensive.

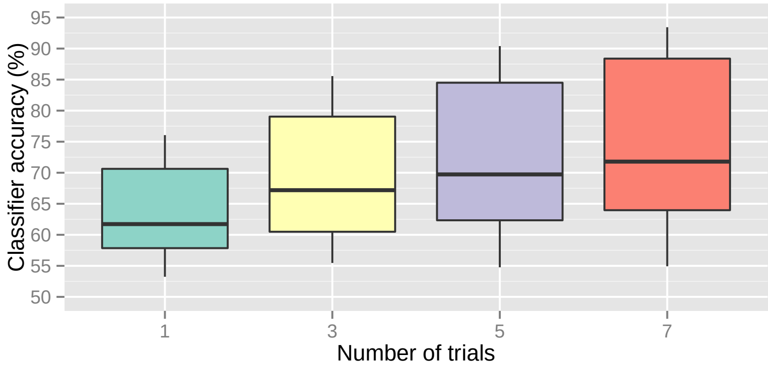

Figure 5.7 – Classifier accuracy depending on the size of trials clusters

We studied which channels contributed the most and which contributed the least to the classifier output. For each channel, we averaged across participants the absolute value of the spatial filter's coefficients that were computed by the REFSF extraction method. We arbitrarily normalized the data between -1 and 1 for more clarity (see Figure 5.8).

To assess the performance of a BCI system that would use less EEG electrodes, we retained the upper half of the channels that contributed the most to features extraction using these computations – i.e. F4, PO4, CP1, FC1, FC2, CP2, P3, Oz, FC6, P4, Fz, AF4, PO3 and Pz. With the reduced set of 14 EEG channels, the classifier resulted in a 62.77% accuracy (SD=7.47), which is close to the configuration that includes all channels.

5.4.2.3 *Online scenario*

The pipeline that we presented in Section 5.3 would be difficult to apply in real-life scenarios – online analyses prevent the use of advanced signal processing, such as ICA for artifact removal, because it requires heavy computations and often necessitates the entire EEG trace to be effective. Fortunately, artifacts had little incidence on the performance of our system. We simulated an online pipeline by skipping several steps – we removed ICA decomposition and did not use neither the ADJUST toolbox nor the `pop_autorej` function from eeglab – and still obtained similar results, with an accuracy of 62.40% (SD=4.80).

5.4.2.4 *Combining with frequency bands*

Although the purpose of this chapter is not to focus on signal processing, we will briefly describe how we managed to improve the performance of our system by combining temporal features (i.e. ERPs) with spectral features ("frequency bands").

Besides REFSF for temporal features, we used common spatio-spectral patterns (CSSP) to extract spectral features [Lemm et al., 2005]. 4 frequency bands were extracted: delta (1Hz - 3Hz), theta

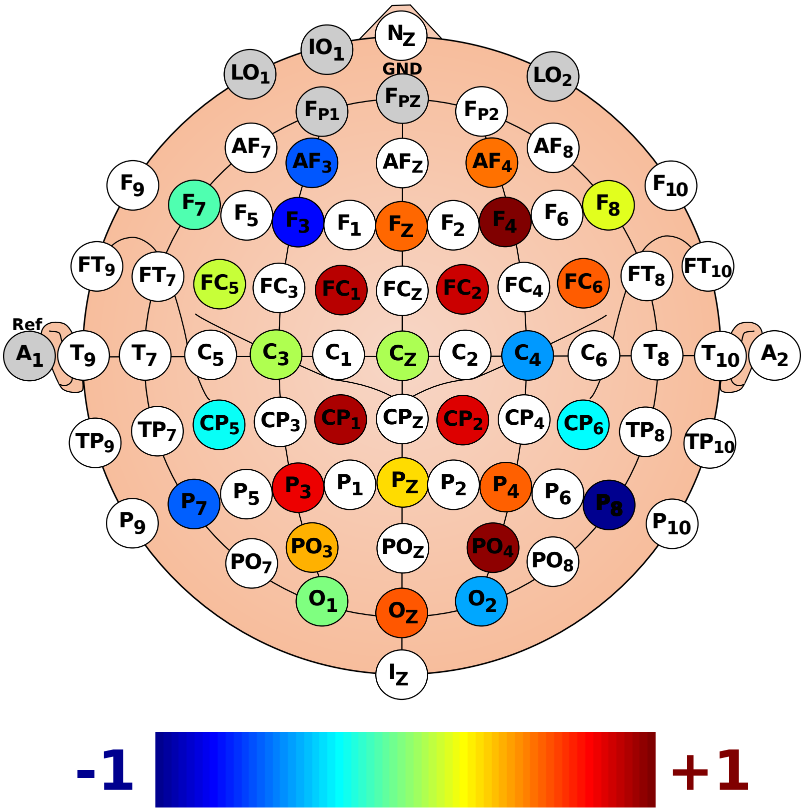

Figure 5.8 – EEG channels' contribution to the spatial filter used by the classifier, averaged across participants. The unit of the scale is arbitrary, from "-1" (least important) to "+1" (most important).

(4Hz - 6Hz), alpha (7Hz - 13Hz) and beta (14Hz - 25Hz). Concurring with ERSP analyses and the time course in Figure 5.6 – that depicts differences between C and NC conditions – the best results were reached by extracting spectral features over a 1s time window that started at t=1100ms (1000ms + 100ms for image appearance delay). REFSF and CSSP features were concatenated and normalized (z-score). Using a feature selection method based on the ratio of features' medians [Guyon, 2003], we reduce the number of features passed to the classifier from 184 to 50 – there were at the beginning 160 features from REFSF + 24 features from CSSP, 3 pairs from 4 bands.

In the end we obtained a 64.66% mean accuracy (SD=5.79). Note, however, that this 64.66% classification accuracy is not statistically different from the 63.30% accuracy obtained using only ERP (Wilcoxon test, $p > 0.05$). This therefore suggests that although the spectral features do contain relevant information for classification, this information might not be different from the one contained in ERP. Alternatively, maybe the approach we used to combine these two kinds of information was not optimal.

Table 5.2 – Overview of the classifier performance for the various methods we investigated.

Method	Accuracy	SD
Offline pipeline (ERP)	63.30%	7.64
Monte Carlo (ERP, n=3)	68.91%	10.32
Monte Carlo (ERP, n=5)	71.83%	12.28
Monte Carlo (ERP, n=7)	74.08%	13.39
ERP + spectral	64.66%	5.79
14 EEG channels (ERP)	62.77%	7.47
Simulated online pipeline (ERP)	62.40%	4.80

5.4.3 Factors influencing classification

We investigated which personal factors could influence the results of our classifier. Outside EEG recordings, the data that reflected most participants inter-variability was concealed among the task's scores and the symptoms associated to stereoscopy. We used Spearman's rank correlation to test between, on the one hand, classifier accuracy and, on the other hand, the difference between NC/C scores and NC/C answers to symptoms questionnaires.

There was no significant association. Neither with the performance task ($p = 0.44$), with eye comfort ($p = 0.81$) nor with vision clarity ($p = 0.57$).

5.5 Discussion

During short exposures to images, participants reported worse vision clarity and less visual comfort in NC condition, thereby validating a clear distinction between the two zones of comfort of our protocol. Participants performed equally well in both conditions during the task, suggesting that even if severe, a VAC does not alter their ability to make rough estimations of virtual depths. In this context, it also highlights the limits of behavioral methods in measuring participants' comfort. A neuroimaging technique, on the other hand, did manage to discriminate two comfort conditions.

EEG signals reflected the disparities in visual comfort. We mainly focus our computations on ERPs, as temporal features led to a signal processing pipeline that was both comprehensible and effective. Using an offline analysis, it was possible to build a classifier that achieved an accuracy greater than 63%, with several participants exceeding 70%. The system scored above chance level in all our analyses ($p < 0.01$) [Müller-Putz et al., 2008]. The performance of the classifier was not influenced by participants' ability to perceive depth nor by the strain that induced the presentation of stereoscopic images.

This score of 63% accuracy, while not as high as some other established BCI systems, may be already sufficient to improve users' comfort.

Indeed, on-the-fly correction of uncomfortable images can be seen as error correction, and in such settings detection rates from 65% are acceptable to improve interactions [Vi and Subramanian, 2012]. These findings depend on the nature of the task, of course. This is why we proposed a mechanism to increase the performance of the classifier.

By taking into account more than one object appearance, or by increasing the duration of viewing sessions, the classifier should become more reliable. The system score improved by 6 points when we clustered trials by 3. During our simulations, the accuracy went around 90% for some users with 5 trials, and for one-third of the participants over 7 trials. It is possible to use this method to simulate an arbitrary number of consecutive trials. Therefore, this tool can estimate how many presentations are needed to reach a specific accuracy and suit the desired application.

During our study we found differences among frequency bands power. While those differences were not large and did not significantly improve the accuracy, spectral features may present an opportunity to strengthen classifier's performance with further investigations.

We were able to replicate our results with a simplified pipeline that could be applied online, paving the way for real-life applications. Furthermore, we were able to select the EEG channels that contributed the most to classifier performance and to halve their number with little loss in accuracy. Even though we used a medical grade EEG equipment to set the basis of a new adaptive system, it seems to indicate that our system could remain functional with entry-level devices. As a matter of fact, the reduced number of channels that we used – 14 – correspond to the number of EEG electrodes found on the Emotiv EPOC[5]. With the EPOC the positioning of the electrodes is constrained by the manufacturer, but other initiatives, such as the 16-channels OpenBCI system[6], may combine affordability, flexibility, reliability and ease-of use, as we investigate in the Appendices.

5.6 Conclusion

We described an innovative system that can distinguish uncomfortable stereoscopic viewing conditions from comfortable ones by relying on EEG signals. We controlled the experimental conditions with questionnaires, founding significant differences in visual comfort between short exposures of images. Visual *comfort* was assessed, whereas existing studies focused on visual *fatigue* – a component that appears on the long term and that we propose to prevent beforehand.

A passive stereoscopic comfort detector could potentially be useful for multiple applications, as a tool to: 1) compare with exocentric measures (possibly offline) different stereoscopic displays, 2) dynamically

[5] https://emotiv.com/
[6] http://www.openbci.com/

enhance stereoscopic effects, by increasing discrepancy without causing discomfort, 3) quickly calibrate stereoscopic displays, 4) dynamically adapt discrepancy to avoid discomfort (e.g. during 3D movies) or voluntarily cause discomfort (e.g. for basic science studies about perception), among many others.

Using short time windows (features were extracted over 1s), and minimalist stimuli (still objects) we set the basics of a tool capable of monitoring user experience with stereoscopic displays in near real time. Our offline analysis used the state of the art in signal processing to demonstrate the feasibility of such a method with clean EEG signals. We obtained a similar classification accuracy without computationally demanding artifacts filtering, demonstrating also that the work presented here could perfectly be applied online. A framework like OpenViBE could ease the creation of an online scenario. Even though some BCI applications are biased by artifacts non originating from brain activity – e.g. emotion recognition by facial expression [Friedman and Thayer, 1991] – during our investigations we discovered that eye artifacts did nothing but adding slight noise to the system. Either an automatic removal method could be employed to clean the signal online [Schlögl et al., 2007] or the EEG electrodes could be positioned over the parietal and occipital regions.

More complex signal processing can increase classification rate. We gave insights on how the addition of spectral features to the temporal information may improve the accuracy of the system. We also described a method that can assess how many stimuli are needed to reach a particular accuracy – i.e. Monte Carlo simulations.

Although it is not deniable that it is currently easier to calibrate displays without EEG, a passive BCI can adapt the parameters to users' state throughout the viewing. It is complementary to other methods that aimed at improving users' comfort. It is possible to integrate EEG measures with other physiological sensors, as hinted by other systems [Bang et al., 2014] and as see when we reviewed HCI evaluation methods in part II.

At the same time that a passive BCI that could adapt viewing conditions to users is built, experimental protocol should be enhanced to integrate richer stimuli. Colors, shadows, relative positions or movements: many cues participate in the comprehension of depth. Besides, real-life scenarios involve virtual scenes that are more complex than grey images of primitive shapes. It will also lead to a broader VAC spectrum. Even if the "curse-of-dimensionality" [Friedman, 1997] will prevent a classifier to possess many classes, the more VAC are taken into account, the more refined adaptive systems will be.

In order to make the adaptive system reliable and useful for the many, differences between individuals that influence classifier performance need to be studied. Physiological characteristics – e.g. interpupillary distance –, past experience with stereoscopic displays – some people may need more time to acclimate to such technology – should be weighed against. Users' states also have to be taken into account; e.g. mental fatigue likely relates to visual fatigue. This study will hereby lead to promising work in many fields: human factors to understand brain patterns disparities, signal processing to improve accuracy, design to create adaptive interfaces, entertainment to integrate comfort measures, and manufacturers, to create more accessible hardware solutions and popularize the use of EEG. By combining those different areas of expertise, passive BCIs should become a viable option for increasing users' comfort, a solution that does not disrupt work or the narrative.

The transition toward more practical settings should be seamless, as classifier performance remains stable even when half the EEG electrodes are used. Next step would consist in conceiving an analogous online system that monitors more complex virtual scenes. A real-world application could consist in a gamified version of our task that smoothly corrects depth range upon classifier output. Such smooth alteration could be applied to animation movies as well. The discrepancy between left and right images would be gradually reduced while discomfort is detected – e.g. when several presentations of objects that are virtually close to the users trigger such label within the classifier. On the contrary, the discrepancy could be increased gradually to enhance the stereoscopic effect as long as no discomfort is detected. This requires only the tuning of one parameter of the display, which is accessible for example through devices such as the Nintendo 3DS or the Nvidia 3D vision system. When the content is dynamically generated – i.e. video games – the control over the virtual scene is even more substantial. In this case one could differently adapt objects' position, according to whether they seem to appear in front of the screen or behind it.

We documented a novel solution to a famous issue – i.e. estimating stereoscopic discomfort – thus increasing fundamental knowledge. Besides 3D scenes control, by giving access in real time to users' inner states, EEG will help to modulate more closely the viewing experience according to the effect one wants to achieve. In a HCI context, this tool could also be extended with the measure of other constructs, such as workload – see next chapters.

6

Workload during 3D manipulation tasks

As stated in part I, designing User Interfaces (UI) requires adequate evaluation tools to ensure good usability and user experience. While many evaluation tools are already available and widely used, existing approaches generally cannot provide continuous and exocentric measures of usability qualities during interaction without interrupting the user. In this chapter, we propose to use brain (with electroencephalography) and physiological (electrocardiography, electrodermal activity) signals to continuously assess the mental effort made by the user to perform 3D object manipulation tasks. We first show how this mental effort (a.k.a., mental workload) can be estimated from such signals, and then measure it on 8 participants during an actual 3D object manipulation task with an input device known as the CubTile. Our results suggest that monitoring workload enables us to continuously assess the 3DUI and/or interaction technique ease-of-use. This was the first application of the framework outlined in previous part, suggesting that EEG could become a useful addition to the repertoire of available evaluation tools, enabling a finer grain assessment of the ergonomic qualities of a given 3D user interface.

This work was published in [Wobrock et al., 2015].

I thank Dennis Wobrock for his involvement – more in the appendices, section Credits.2.

6.1 Introduction

3D User Interfaces (UI) and systems are increasingly used in a number of applications including industrial design, education, art or entertainment [Bowman et al., 2004, Jankowski and Hachet, 2013]. As such, 3DUI and interaction techniques can be used by many different users with many varying skills and profiles. Therefore, designing them requires adequate evaluation tools to ensure good usability and user experience for most targeted users [Bowman et al., 2002, Jankowski and Hachet, 2013]. To do so, a number of evaluation methods has been developed including behavioral studies, testbeds, questionnaires and inquiries, among others [Bowman et al., 2004, Jankowski and Hachet, 2013] – see also chapter 1. This resulted in the design of more relevant, efficient and easy-to-use 3DUI.

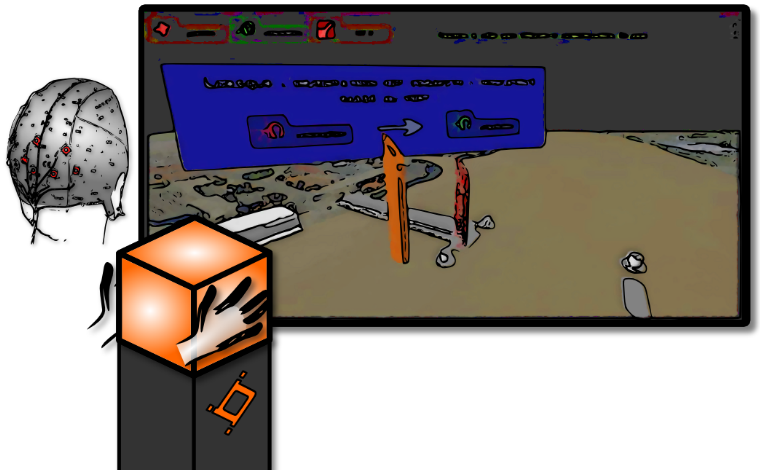

Figure 6.1 – Schematic view of a user performing 3D manipulations tasks with the CubTile input device. His/her mental effort are monitored based on brain signals (ElectroEncephaloGraphy).

Nevertheless, there is still a lot of room for improvements due to limitations of traditional evaluation methods – e.g. measures' ambiguity and discontinuity, bias due to social pressure and so on, see part I. A useful UI evaluation measure is the user's mental workload, i.e. the pressure on the user's working memory, which is typically measured using the NASA-TLX post-hoc questionnaire [Hart and Staveland, 1988]. Even though it can be used to assess users' preferences regarding UI [Karnik et al., 2013], NASA-TLX being a post-experiment measure, this is only a egocentric and global measure that cannot inform on where and when the user experienced higher or lower workload. There is therefore a need for more exocentric and continuous measures of the usability qualities of 3DUI that do not interrupt the user during interaction. We described previously how physiological computing could help to obtain such measures of the user's inner-state during interaction.

Indeed, there are increasing evidence that the mental states that can be relevant for 3DUI evaluation, like mental workload [Mühl et al., 2014], can be estimated from brain and physiological signals [Nourbakhsh et al., 2013]. Interestingly enough, some recent works have started to use brain signal based measures of workload to compare 2D visual information displays [Peck et al., 2013, Anderson et al., 2011]. However, to the best of our knowledge, estimating mental workload from both brain and physiological signals has never been explored to evaluate 3DUI, although it could provide relevant evaluation metrics to complement the already used ones. Indeed, previous works were focused on evaluating workload levels based on brain signals during 2D visualizations, thus with more passive users [Peck et al., 2013, Anderson et al., 2011]. 3D interaction tasks are more complex for the user since 1) the user is actively interacting with the application, and not as passively observing it, and as such should decide what to do and how to do so, and 2) perceiving and interacting with a 3D environment is more cognitively demanding than perceiving and interacting with a 2D one, since it required the user to perform 3D mental rotation tasks to successfully manipulate 3D objects or to orientate him/herself in the 3D environment. Therefore, as compared to existing works which only explored passive 2D visualizations, monitoring mental workload seems more relevant during 3D manipulation tasks, since the user is more likely to experience pressure on his/her cognitive resources. Therefore, evaluating the resulting changes in workload levels seems even more necessary to ensure the design of usable 3DUI. Moreover, the active role of the user during 3D interaction tasks (as compared to more passive visualizations) and the higher cognitive demand as well as the richer visual feedback resulting from the use of a 3D environment means that EEG and physiological signals will be substantially different and more variable as compared to those measured during 2D visualization tasks. Finding out whether they can still be used to estimate workload levels in this context is therefore a challenging and relevant question to explore.

Therefore, in this chapter, we propose to assess users' workload (i.e. their mental effort) during 3D object manipulation tasks, based on brain (EEG) and other physiological signals. We notably propose a method to estimate workload levels from EEG, ECG and EDA signals, and we study mental workload levels during a 3D docking task in a pilot study (see Figure 6.1). Our results show that this approach can provide useful information about how users learn to use the 3DUI and how easy-to-use it is.

6.2 Experiment

6.2.1 Interaction device and 3D task

Participants had to completed 3D manipulation tasks using the CubTile. The CubTile is an input device developed by Immersion, made of 5 orthogonal touch surfaces (see Figure 6.2). Several webcams are embedded inside the "cube" and monitor the shadow that are cast off the outer surface in order to detect touch – thus gloves or tools can be used to interact with the CubTile. Since the surface are white and opaque, only close fingers or objects are registered, i.e. when a contact is made. Finally, because the surfaces are illuminated from the inside, the ambient light do not interfere with the images recorded by the webcams. A computer is comprised in the main body of the CubTile, which acquires webcams' optical flow and detects commands. Three different commands are handled:

- Zoom with a "pinch" movement. Fingers getting closer: zoom-out, fingers spreading on the surface: zoom-in.
- Rotation, when at least 2 fingers are detected and a rotation motion is detected
- Translation: sliding movement motion detected

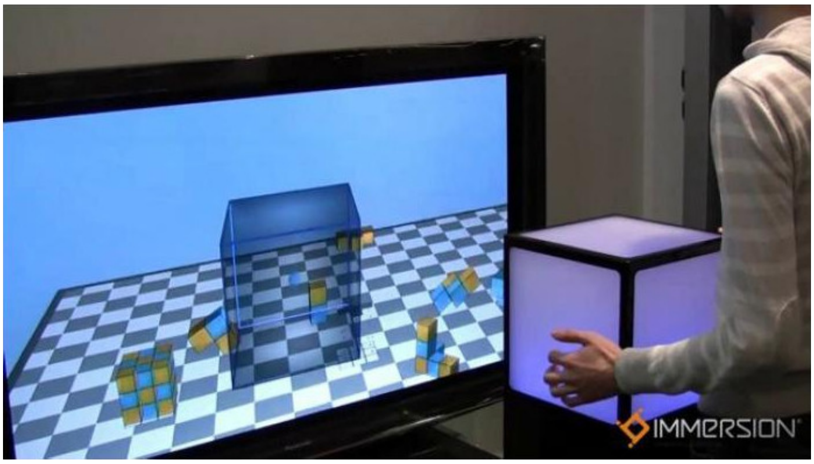

Figure 6.2 – The CubTile device, 5 touch surfaces assembled orthogonally.

The axis onto which are applied those transformations depends of the face of the cube that is used. E.g. a rotation gesture on the sides of the box will produce a rotation of the Y axis in a point of reference with Z facing upward, a sliding gesture on face situated on top of the cube that goes further away from the user will produce a translation along the X axis of the virtual environment, toward the origin.

The CubTile is connected to the main computer with an ethernet cable, using VRPN (virtual reality peripheral network) protocol to send information. Each transformation and axis is encoded in one VRPN channel (9 in total).

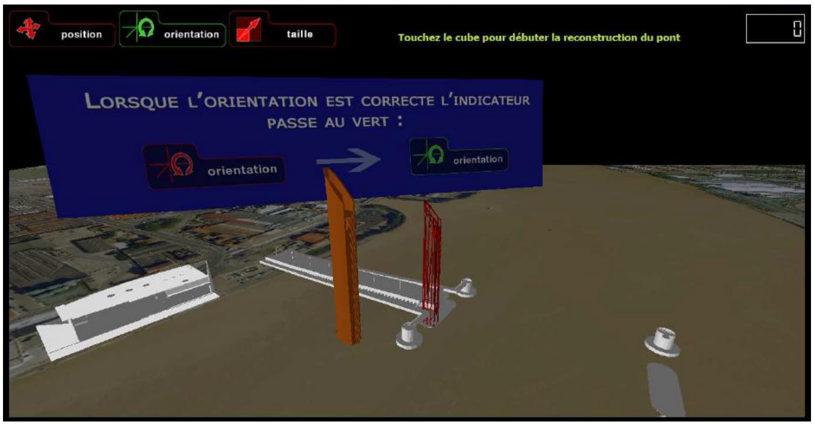

Figure 6.3 – Building of a bridge using the CubTile in a virtual environment.

The CubTile is particularly well suited to manipulate 3D objects, as the transformations could be directly applied to them. The task used in our experiment was previously developed to demonstrate such capabilities, where users had to assemble one by one parts of a 3D bridge – 4 supporting pillars and the road (see Figure 6.3). In particular, users had to perform docking tasks, by translating, rotating and scaling the bridge parts, in order to put them at the correct location. The correct location was indicated to users with proper 3D feedback, integrated to the 3D scene, in the form of text and color indicating how close he/she was from the correct position, scale and orientation. All the translations, rotations and scaling were controlled by the CubTile. The participant had to perform a set of 7 docking tasks:

Virtually building the bridge "Chaban Delmas" was initially a public exhibition that took place in the scientific museum Cap Sciences.

1. Positioning the 1st pillar, by controlling rotation, translation and scaling. Repeated 3 times for different angles, sizes and positions.
2. Positioning the 2nd pillar, by controlling 2 translations, 1 rotation and scaling, while the pillar was being continuously and automatically translated along the vertical axe. Repeated 4 times for different angles, sizes and positions.
3. Positioning the 3rd pillar lower half by controlling a crane, carrying the pillar part, along 1 rotation and 1 translation (up/down). Repeated 3 times for different angles and heights.
4. Positioning the 3rd pillar upper half by controlling a crane along 1 rotation and 1 translation, seen from a different angle as above. Repeated 3 times for different angles and heights.
5. Positioning the 4th pillar by controlling 2 translations and 1 rotation. Without warning the users, the gestures for rotation and translation were inverted, e.g. moving symmetrically two fingers

on opposite sides of the cube triggered a rotation instead of the usual translation. Controls were only inverted for this task.

6. Positioning the road joining the first two pillars to the river bank with 2 translations and 1 rotation. Repeated 3 times for different angles and starting position.

7. Positioning the road joining the four pillars with 1 translation, 3 rotations and scaling.

These different tasks enable us to observe how users get to learn how to use the CubTile for 3D objects manipulation tasks. Task number 5, with inverted control commands, enables us to observe mental workload while using voluntarily difficult and counter-intuitive interaction techniques.

6.2.2 Physiological recordings

During the experiment, the physiological states of the users were recorded with EEG, ECG and EDA.

The EEG device was, once again, composed by two g.USBAmp amplifiers made by g.tec. This system can record up to 32 electrodes. In this montage 30 electrodes were installed, but we had to discard two of them due to technical difficulties. The active electrodes were connected to two g.GAMMAsys boxes and were positioned according to the 10-20 international system. In the end the following locations were used during our analysis: C6, CP4, CPz, CP3, P5, P3, P1, Pz, P2, P4, P6, PO7, PO8, Oz, Fz, F4, FT8, FC4, FCz, FC3, FC5, FT7, C5, C3, C1, Cz, C2 and C4.

The other physiological sensors were acquired separately, through the use of a Bitalino acquisition card, an Arduino compatible board dedicated to physiological recording [Placido da Silva et al., 2014]. We used the "plugged" version of the Bitalino board. The sensors of the plugged version are remotely connected to the board through a molex wire and not directly attached to the board, so that users movements are less hindered by the device.

3 ECG sensors were placed on the user's torso, and 2 EDA sensors on the user's index and middle fingers from the non active hand. EMG electrodes were also positioned on the forearm corresponding to the active hand. Although we believed that muscle activity related to the dominant hand could have assessed users state – an increased tonic in activity caused by stress or phasic responses triggered by poorly responsive interface – we discarded those recordings as did not find a way to analyze them correctly. The position of the electrodes or the type of interaction may have induced too much noise, or the features were too randomly distributed, in any case there were no conclusive information whatsoever that would gain in being described further. We only kept the description for the sake of a fully described protocol, as participants were all in all heavily equipped.

All physiological measures were recorded within OpenViBE. EEG data was fetched from the acquisition server while the python API of the Bitalino was employed in custom boxes to retrieve data from the others sensors.

6.2.3 Calibrating workload with the N-back task

Each one of us is unique: it is not a vain refrain, we react indeed differently; more particularly, for identical mental states our physiological signals vary. This is why estimating workload levels from EEG, ECG and EDA signals first requires signals labelled with the corresponding user's mental workload.

To obtain a ground truth signal data set to calibrate and validate our workload estimator, we induced 2 different workload levels in our participants. To do so, we had them perform cognitive tasks, the cognitive difficulty of which being manipulated using a protocol known as the N-back task, a well-known task to induce workload by playing on memory load [Owen et al., 2005]. While various sensory modalities can be used to implement a N-back (e.g. sounds), usually visual information, such as letters, are employed. Since the CubTile manipulation involved mainly visual information, we kept those types of item.

In the N-back task, users watch a sequence of letters on screen, the letters being displayed one by one. For each letter the user had to indicate whether the displayed letter was the same one as the letter displayed N letters before or was different – hence users have to remember *n* items at every moment – using a left or right mouse click respectively.

We implemented a version similar to [Grimes et al., 2008], removing vowels to prevent chunk strategies based on phonemes. We used the same time constraint as in [Mühl et al., 2014], i.e. letters appeared for 0.5s, with an inter-stimulus interval of 1.5s. Each user alternated between "easy" blocks with the 0-back task (the user had to identify whether the current letter was a randomly chosen target letter, e.g. 'X') and "difficult" blocks with the 2-back task (the user had to identify whether the current letter was the same letter as the one displayed 2 letters before). For example, in the 2-back task, with the sequence "1: W, 2: Q, 3: R, 4: Q, 5: R, 6: K", users should press on the left buttons at t=4 ("Q" on-screen and "Q" at t=2) and at t=5 ("R" on screen and "R" at t=3) – see Figure 6.4.

Each block contained 60 letters presentations. 4 letters were drawn at the beginning of a block so that the number of target letters accounted for 25% of the trials. Each participant completed 6 blocks, 3 blocks for each workload level (0-back vs 2-back). Therefore, 360 calibration trials (i.e. one trial being one letter presentation) were collected for each user, with 180 trials for each workload level ("low" *vs* "high").

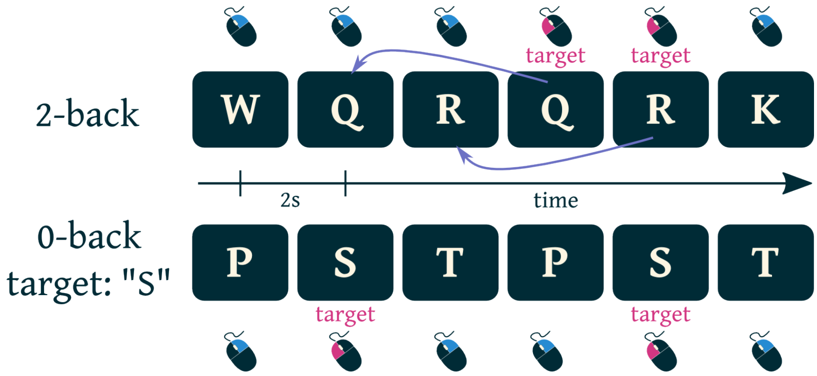

Figure 6.4 – *Top*: 2-back task, the target letter is the one that appeared two steps earlier, users have to select trials 4 and 5. *Bottom*: 0-back task, the target letter “S” is randomly chosen, users have to select trials 2 and 5.

The calibration took approximately 12 minutes. It was possible to configure mouse buttons according to users’ habits, even though that possibility have never been requested. Indeed, some users (e.g. left-handed people) invert right click and left click sequences in their desktop environments so that the default action selection occurs with a right click instead of a left click.

6.2.4 Procedure

The experiment took place in a dedicated experimental room, in a quiet environment. 8 participants (2 females, age from 16 to 29) took part in this study. They were all first-time users of the bridge building application and the CubTile (except for one participant who has used the CubTile before for another application). During the whole duration of the experiment, the participants’ brain and physiological signals were recorded.

The experiment comprised the following stages:

1. Participants entered the room, they were seated before a table. They read and signed the inform consent. They were briefly explained the context and the goal of the study.

2. Participants were equipped with the physiological sensors, starting with the EEG headset. Once the EEG was installed, quick verification were made to control for the quality of the recordings (participant had to blink and to clench their teeth). Then we proceeded to the installation of the different Bitalino sensors. Once again, we verified that the readings were consistent with the sensors, e.g. heart beats visible in the ECG at a non-lethal pace.

3. Participant still seated at the same table, we proceeded to the calibration of workload. This took approximately 15 minutes. Two

other calibration sessions occurred (about 15 minutes each), for two other constructs that were not used nor analyzed for this study – see Conclusion.

4. Once the calibration was complete, the participant was asked to sit on an elevated chair in front of the CubTile which was itself in front of a 65 inches Panasonic TX-P65VT20E screen. The task involving the bridge construction occurred, as described previously. Since the instructions were clearly indicated on-sceen, we had to interact little with participants. On very rare occasions, when we saw that they were in difficulty regarding the exact positioning of bridge pieces, we gave them some advice on how to complete the trial. Mainly these problems were caused by a poor depth perception (offset in the Z-axis between the manipulated element and the target). Apart from these possible interruptions, participants smoothly completed the different tasks at their own pace (no time constraints).

5. After the completions of the 3D docking task, physiological sensors were removed.

Overall, with sensors setup and removal, calibration and docking task, the experiment lasted approximately one hour and a half.

6.2.5 Signal processing

In order to estimate mental workload from brain and physiological sensors, we used a machine learning approach: the measured signals were first represented as a set of descriptive features. These features were then given as input to a machine learning classifier whose objective was to learn whether these features represented a low workload level (induced by the 0-back task) or a high workload level (induced by the 2-back task). Once calibrated, this classifier can be used to estimate workload levels on new data, which we will use to estimate mental effort during 3D object manipulation tasks. This pipeline is analogous to what was used during our study of stereoscopic displays in chapter 5, except for artifacts removal that we disabled in order to use the calibration data against the data collected during the main task (see next section below).

From the signals collected during the N-back tasks described above, we extracted features from each time window of EEG and physiological signals immediately following a letter presentation – 2s time windows for EEG, 10s for ECG and EDA, see below. We used each of these time windows as an example to calibrate our classifier. Note that a classifier was calibrated separately for each participant, based on the examples of brain and physiological signals collected from that participant. Indeed, EEG signals are known to be very variable between participants, hence the need for user-specific classifiers to ensure maximal EEG classification performances [van Erp et al., 2012, Mühl et al., 2014].

6.2.5.1 EEG

We used the EEGLab software [Delorme and Makeig, 2004] to process EEG signals. We filtered the signals in the delta (1-3 Hz), theta (4-6 Hz), alpha (7-13 Hz), beta (14-25 Hz) and gamma (26-40 Hz) bands, as in [Mühl et al., 2014]. For each band, we optimized a set of 6 Common Spatial Patterns (CSP) spatial filters (i.e. linear combinations of the original EEG channels that lead to maximally different features between the two workload levels) [Ramoser et al., 2000, Lotte, 2014]. For each frequency band and spatial filter, we then used as feature the average band power of the filtered EEG signals. This resulted in 30 EEG features (5 bands $\times$ 6 spatial filters per band). Note that high frequency EEG is likely to be contaminated by muscle activity (ElectroMyoGraphy - EMG) from the user's face or neck [Fatourechi et al., 2007, Goncharova et al., 2003]. As such, we explored EEG-based workload estimation based on low frequencies only (delta, theta, alpha) and both low and high frequencies (delta, theta, alpha, beta, gamma).

This signal processing approach is the one we used to discriminate workload levels from EEG signals between 0-back and 2-back tasks, i.e. within the same context on which the workload estimator was calibrated. However, it is known that EEG signals change between different contexts, due, e.g. to the different user's attention and involvement that the context triggers or to different sensory stimulations (e.g. different visual inputs) that change brain responses and thus EEG signals. This means that a workload estimator calibrated in a given context will have poorer performances (i.e. will estimate an erroneous workload level more often) when applied to a different context [Mühl et al., 2014].

In our experiment, the final application context, i.e. 3D objects manipulation, is very different from the calibration context, i.e. the N-back tasks. Indeed, during the N-back tasks the user is moving very little as he/she is only performing mouse clicks, and exposed to very little visual stimulations as the N-back task only involves the display of white letters on a black background. On the contrary, manipulating 3D objects means that the user will be moving more and would be exposed to very rich visual stimulations. As such, a workload estimator simply calibrated on the N-back tasks and applied to the 3D object manipulation tasks is very likely to give very poor results or even to fail. Therefore, we modified the above mentioned signal processing approach to make it robust to EEG signal changes between the two contexts.

Rather than using basic CSP spatial filters, we used regularized CSP spatial filters [Lotte and Guan, 2011] that are robust to changes between calibration and use contexts. To do so, based on [Samek et al., 2013], we estimated the EEG signal covariance matrix from the calibration context (N-back tasks) and from the use context (3D object manipulation tasks), and computed the Principal Components (PC) of the difference between these two matrices. These PC represent the directions along which EEG signals change between calibration and use. These directions are then used to regularize the CSP spatial filters as in [Samek et al., 2013], to ensure that the obtained spatial filters are invariant to these EEG signals changes. These stationary subspace CSP (SSCSP) spatial filters prevented us to use the artifacts removal methods employed in chapter 5. It seemed that signals resulting from ICA components rejections were too dissimilar between training and testing for our implementation of SSCSP to be still effective.

Note that SSCSP filters are only possible here because we perform an offline evaluation, after the 3D manipulation tasks have been performed and the corresponding EEG signals collected. It would not be possible to use the exact same algorithm for a real-time estimation of workload during 3D objects manipulation tasks as the covariance matrix of EEG signals during these tasks is not yet fully known.

6.2.5.2 *ECG*

When heart beats, a clear signature appears in the ECG, a positive peak of large amplitude that is preceded and followed by much smaller peaks and deflections. It corresponds to the blood flowing between the different ventricles. Each peak and deflection is represented by a latter and the overall signature is called the "QRS complex". Feature extractions algorithm can detect those QRS complexes and retrieve at which precise moment they occurred in the ECG recordings. To do so, we used Biosig 2.92 [Vidaurre et al., 2011], a Matlab toolbox that is integrated within EEGLab suite.

Two QRS detection methods are implemented in Biosig: [Afonso et al., 1999] (default) and [Nygårds and Sörnmo, 1983]. We obtained more reliable results using the latter – we compared timings and QRS counts with ECG raw signal on several samples. We did not need to filter the ECG signal, the QRS detection yielded the same results whether we applied a high-pass filter to remove the DC drift or not.

As for newer versions of Biosig, e.g. 2.94, the QRS method we chose back then became the default.

From ECG signals we extracted the Heart Rate (HR) and 2 features from the Heart Rate Variability (HRV), namely the low frequency HRV (0.1 Hz) and the Root Mean Square of Successive Differences, as in [Mehler et al., 2011], using functions from the Biosig toolbox.

6.2.5.3 *EDA*

EDA relates to the sympathetic branch of the autonomic nervous system. Although it is often used as a proxy for arousal, EDA is a multifaceted phenomenon that does not reflect a single psychological process [Figner and Murphy, 2011].

From EDA signals we extracted 3 features: the mean EDA amplitude, skin conductance responses (SCR, here band power between 0.5 Hz and 2 Hz) and the skin conductance level (SCL, 0.1-0.5 Hz) [Figner and Murphy, 2011].

SCL correspond to the tonic response of the skin, a slow shift of conductance from the basal level that characterizes the latent state of the user. The SCR, on the contrary, is linked to events of short duration – e.g. conductance will rise over a few seconds before going back to the previous level when a sudden increase of arousal occur.

6.2.6 Classification

We used a shrinkage Linear Discriminant Analysis (sLDA) classifier [Lotte, 2014] to learn which feature values correspond to a high or low workload level.

Note that since both ECG and EDA analyses rely on low frequencies, we had to extend the time windows from 2s to 10s when we studied those physiological signals (for instance, for HRV at 0.1Hz, 10s are needed to observe a single cycle). As such the number of trials per condition (0-back vs 2-back) in these particular scenarios were reduced from 180 down to 36.

6.3 Results

For each user, we first setup a workload level classifier based on the signals collected during the calibration session (N-back tasks). The next section describes the performances achieved for each participant and each signal type. Then, using the best workload classifier, we could estimate the workload level over time during the 3D docking tasks. This work was done offline, after the experiment.

6.3.1 Accuracy of mental effort detection

First, based on the data collected during the calibration session (N-back tasks), we could estimate how well low workload could be discriminated from high workload based on EEG, ECG and EDA. To do so, we used 2-fold cross-validation (CV) on the calibration data collected. In other words, we split the collected data into two parts of equal size, used one part to calibrate the classifier (CSP filters and sLDA), and tested the resulting classifier on the data from the other part. We then did the opposite (training on the second part and testing on the first part), and averaged the obtained classification accuracies (percentage of signal time windows whose workload level was correctly identified). This CV was performed by using each signal type (i.e. EEG, ECG and EDA) either separately or in combination. Table 6.1 displays the obtained classification accuracies.

Table 6.1 – Cross-validation classification accuracies (%) to discriminate workload levels from EEG (delta, theta and alpha bands), EMG+EMG (delta, theta, alpha, beta and gamma bands), ECG and EDA on the calibration session data. A "*" indicates mean classification accuracies that are significantly better than chance (according to [Müller-Putz et al., 2008]).

Sensor	S1	S2	S3	S4	S5	S6	S7	S8	Mean.
EEG	74.0	76.2	76.5	80.2	84.9	81.9	81.7	75.4	78.9*
EEG+EMG	85.0	93.1	81.7	87.6	94.8	97.3	84.8	84.3	88.6*
ECG	37.3	50.7	45.3	58.7	42.6	55.3	54.9	61.2	50.7
EDA	77.3	52.1	60.0	70.6	74.7	68.4	58.6	54.6	64.5
All	44.0	53.3	44.0	61.5	54.8	52.6	54.6	61.2	53.3

II.6

Classification results highlight that workload levels can be estimated in brain and physiological signals, even though the large inter-participant performance variability suggests that workload levels can be estimated more clearly for some users than for some others.

As can be first observed, it appears that EEG can discriminate workload levels with an accuracy higher than chance level, for all participants. In other words, the classification accuracies obtained are statistically significantly higher than 50% for a 2-class problem, i.e. more accurate than flipping a coin to estimate the workload level. Indeed, according to [Müller-Putz et al., 2008], for 160 trials per class, the chance level for $p < 0.01$ and a 2-class problem is an accuracy of 56.9%. Note that 180 trials per class were available with EEG in our experiment, meaning that the chance level is actually even slightly lower.

Regarding the EDA, it led to a better-than-chance classification accuracy only for some participants, but not for all. Indeed, we had 36 trials per class with EDA (due to the use of longer time windows as mentioned previously), which means a chance level of about 65% for $p < 0.01$ according to [Müller-Putz et al., 2008]. ECG signals could not lead to better-than-chance performances for any participant.

Overall, EEG seems to be the signal type the best able to discriminate workload levels reliably. Moreover, when EEG features include high frequency bands – i.e. when delta, theta and alpha bands are combined with beta and gamma bands – and thus when EEG measures potentially contain EMG activity as well, the performances are the highest, close to 90% on average.

The poor performances of the system when all physiological signals are combined (EEG + EMG + ECG + EDA) may be explained by too important disparities in the features for the classifier to handle them correctly. On a side note, we also tested ECG and EDA on 2s time windows with adapted features – HR for the former, mean value and SCR for the latter. Despite the increased number of trials in training and

Normalizing features with z-score did not help, nor did features selection methods.

testing phases, the results were very similar to those already described in Table 6.1. Altogether, the relatively poor performances obtained with ECG and EDA are likely due to the short time windows (2s or 10s long) used. Much better performances should be expected with larger windows, e.g. with 30s-long or even 2min-long time windows [Mehler et al., 2011], at the cost of a coarser temporal resolution.

Since we already obtained a classification accuracy close to 90% through the sole use of EEG recordings (which possibly include EMG activity as well), we did not push further our investigations about a multimodal (multiple signals) approach to mental effort estimation. Such method would necessitate longer time windows, strong synchronization between signals and extra classification steps, with little benefit to expect considering that a classifier based on EDA hardly reaches 65% of accuracy in our protocol. We then calibrated the workload classifier on EEG signals from both low and high frequency bands (i.e. combining EEG and possibly EMG), and used it to analyze workload variations during the 3D manipulation tasks.

6.3.2 Mental effort during 3D object manipulation

While the participants were performing 3D docking tasks to build the 3D bridge, their brain signals were recorded. By using the workload level classifier obtained offline, such classifier being able to estimate whether the current 2-seconds long time window of signals corresponds to a low or high workload for the user, we could notably continuously estimate the workload levels during the tasks. This gave us unique insights into how much mental effort the participants were devoting to each task, and how this mental effort evolved over time.

Due to the large between-user variability in terms of workload level estimation accuracy, and since these estimations are not 100% accurate, we studied average workload levels to obtain a robust and reliable picture of the mental workload level associated with each task. To do so, we first normalized between -1 and +1 the output that was produced by the classifier for each participant during the virtual bridge construction. As such, a workload index close to +1 during the 3D object manipulation represents the highest mental workload a participant had to endure while performing the 3D docking tasks. It should come close to the 2-back condition of the calibration phase. In a similar manner, a workload index close to -1 denotes the lowest workload (similar to that of the 0-back condition).

Because there was no time constraint regarding task completion – users made as many attempts as needed to complete each one of them – we could not compare directly workload indexes over time. Some participants took more than 13 minutes to complete all the tasks while others finished in less than 5 minutes (mean: 7.7 min, SD: 2.9 min). This is why we averaged the workload index per task. Note that due to technical issues, for some participants the beginning and end of a couple of tasks were not accurately recorded or missing. If it was the case, the workload indexes for this task and participant were not included in the analysis to ensure unbiased results. Altogether, out of the 56 tasks (8 participants × 7 tasks per participant), 13 tasks were not included in the analysis to ensure clean results. More precisely, 1 task was missing for participant S3, 2 tasks were missing for participants S2 and S5, 3 tasks for participant S8, and 5 tasks for participant S4. No tasks were missing for the remaining participants. We followed a rather conservative approach (i.e. we discarded a task in case of doubt), to ensure only clean and meaningful results are presented.

Figure 6.5 displays the workload levels averaged over all participants and over the duration of each docking task. This thus provides us with the average mental workload induced by each 3D object manipulation task.

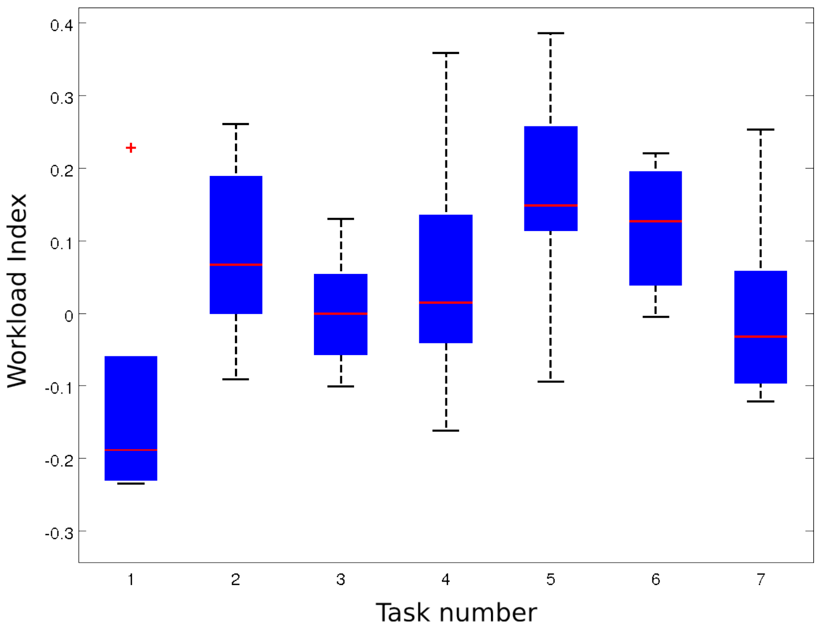

Figure 6.5 – Average workload levels (averaged over participants and task duration) measured for the different 3D docking tasks.

To ensure that the observed workload levels were really due to some information and structure in the data that are detected by the workload classifier, and not just due to chance or to some artifacts that are unrelated to workload levels, we performed a permutation analysis. In particular, we performed the exact same analysis described previously except that we used random classifiers instead of the real workload classifiers trained on the N-back task data. This aimed at estimating the type of workload level indexes we could obtain by chance on our data. To do so, for each participant, we shuffled the labels of the N-back task data, (i.e. the EEG signals were not labelled with the correct workload level anymore), and optimized the spatial filters and classifier based on this shuffled training data. In order words, we built random classifiers that would not be able to detect workload levels. Then using these random classifiers for each participant, we computed the mean normalized workload level indexes for each 3D manipulation task, as described previously. We repeated this process (workload labels shuffling, then random classifier training, and testing of the classifier on the 3D manipulation task data) 1000 times, to obtain the distribution of the mean workload level indexes for each task that can be obtained by chance (see Figure 6.6). More precisely, we estimated the multivariate normal distribution of the vectors of mean workload level per class (i.e. a vector with 7 elements, the i^{th} element being the averaged workload level over participants for task i) obtained for each of the 1000 permutations. This multivariate distribution thus represents the mean workload levels per task that can obtain by chance. We finally compared the actual mean workload levels per task that we obtained using the real workload classifiers (i.e. those optimized on the unshuffled training data, whose output is displayed on Figure 6.5) to this chance multivariate distribution obtained with the random classifiers. This helped us estimate whether the obtained mean workload levels per task were due to chance or not. Results showed that the observed real workload levels are statistically significantly different from that obtained by the chance distribution with $p < 0.001$, i.e. they are not due to chance. This suggests that our workload classifier does find a workload level information during the 3D docking tasks that cannot be found by chance.

II.6

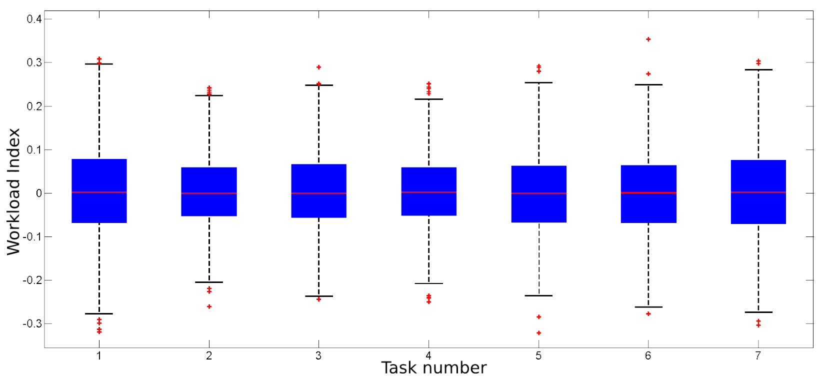

Figure 6.6 – Average workload levels obtained with a permutation test (see text for details), i.e. with random classifiers, for the different 3D docking tasks. The real workload levels we observed (i.e. those displayed in Figure 6.5) significantly differ from those random workload levels, i.e. they are not due to chance.

In order to sense whether or not the workload index fluctuated along tasks completion, we conducted a second analysis. Using the same normalized index, we compared the workload level between the first quarter and the last quarter of every task – average across tasks for each participant (Figure 6.7). A Wilcoxon Signed-rank test showed that there was no significant difference.

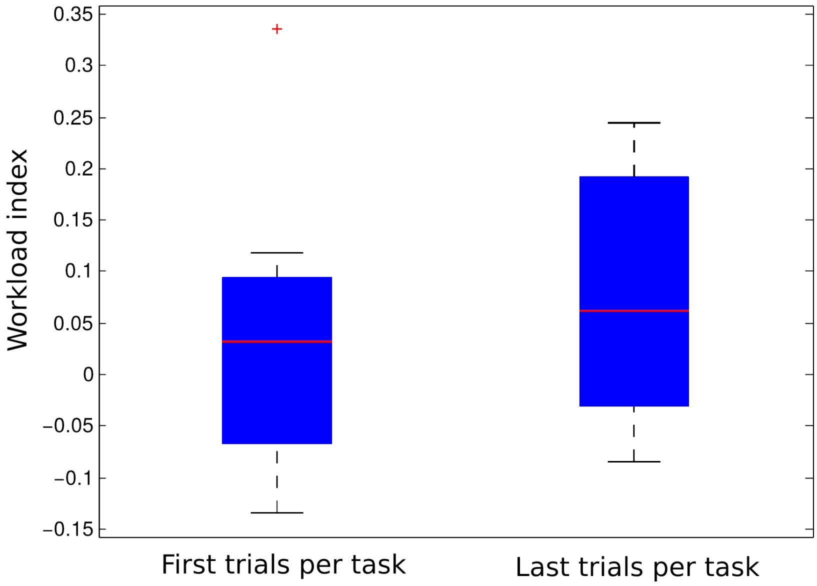

Figure 6.7 – Average per participant of the workload index during the first quarter of every task (left) compared to the last quarter (right).

II.6

6.4 Discussion

First, the fact that the observed workload levels during 3D manipulation tasks are not due to chance on the one hand and that our workload classifiers are calibrated on the N-back task, which is a widely used and validated workload induction protocol [Owen et al., 2005] on the other hand, strongly suggests that our approach may be used to observe how workload levels vary during 3D manipulation tasks. Indeed, our workload classifiers identified a specific EEG+EMG signature of workload levels thanks to the use of the N-back task, which then enabled us to estimate a non-random sequence of workload levels for each task. Naturally, if the variations of another mental state (or artifact) are highly correlated to that of the workload levels, and have a similar EEG+EMG signature as workload so that these variations are picked-up by our classifiers, then the observed variations of workload may be due to variations of another mental state. Therefore, the influence of a confounding mental state or artifacts cannot be completely ruled out without exploring how all possible mental states vary, which is of course impossible. However, the fact that our classifiers are specific to workload variations (since they are calibrated with the N-back task which specifically makes workload levels vary) and that the observed variations are not due to chance makes the influence of a such confounding factor rather unlikely. Based on this interpretation, we can now analyze how the workload level changes during the different 3D manipulation tasks and why.

The observed workload levels suggest that despite the novelty and the complexity of the interaction – handling at the same time rotation, translation and scaling of elements in a 3D environment right from the beginning – the participants did not make an important mental effort to complete the first task. That could be due to the practicality of the CubTile, which may ease 3D interaction thanks to its additional degrees of freedom compared to a traditional input device such as a mouse.

When a constraint appeared concurrently with the second task – pillars were "falling" continuously from the sky and had to be positioned quickly before they touched the ground – the workload index increased substantially. This is consistent with the sudden pressure that was exerted on users. As one could expect, the mental workload lowered and settled in tasks 3 and 4, during which there was no more time pressure – but still more complex manipulations compared to task 1.

We purposely inverted the commands during the fifth task to disorientate participants. As a matter of fact, this is the moment when the workload index was the highest on average among all participants. Then, after this sudden surge of mental stress, once again the measured workload has been reduced in the two subsequent tasks. Interestingly enough, for task 6, in which the control commands were inverted back to normal, the workload indeed decreased as compared to that of task 5, but was still higher than for the other tasks. This probably reflects the fact that users had somehow integrated the counter-intuitive manipulation technique and had to change again the gestures they used to manipulate the 3D object, thus being forced to forget what they had just learned in task 5 which resulted in a high workload. Since the new control scheme was the one they had already used during the previous tasks though, the workload was not as high as in task 5.

Overall, the mental workload that was measured with EEG and EMG along the course of the interaction matches the design of the tasks. Workload increased when a sensitive element of the interaction was deprived – e.g. time or commands– which can be explained by the need to overcome what participants have learned previously and re-learn how to handle the new environment. Afterwards, when going back to the previous scheme, the workload goes back to a low level, as could be expected.

The absence of differences in the workload index between the beginning and the end of the tasks could be due to their duration. We expect to observe a learning effect when the CubTile – or any other input device – is operated during a prolonged period of time in steady conditions; i.e. the workload index would be lower in the end.

These results suggested that continuous mental workload monitoring was possible and could provide us with interesting insights about how cognitively easy-to-use a given 3D interaction technique can be. As compared to previous works, our results show that it is possible to monitor mental workload based on brain and physiological signals, even when the user is actively interacting (and not passively observing as in previous works), moving, and performing more complex and more cognitively demanding 3D manipulation tasks, in a visually rich 3D environment.

Concerning the disappointing performances obtained with heart measures, we realized during our analyses that the ECG signals acquired from the Bitalino contained many artifacts. We quickly understood that they were not due to noise or to movements, but that they originated from problems in data acquisition. We believe that the python scripts we used within OpenViBE to acquire the signal had synchronization issues. When we looked at the raw data we observed that sometimes values were suddenly zeroing, either for several values in a row or for one only. It may be due to a bug in the programming API – for example an absence of interpolation within a data chunk should an insufficient number of samples be fetched from the Bitalino bluetooth feed. Moreover when we used the Bitalino to manufacture wearables later on (see chapter 11) we observed weird behaviors with one of our two Bitalino boards; small values appearing in unused channels next to the one attached to ECG, as if the lower bits of the raw values were not correctly unpacked in floats. As such we also suspect a problem in the Bitalino firmware depending on boards' version.

We did not have time to investigate this issue further, but despite the bad "shape" of raw ECG recordings – i.e. signal drops – the QRS detection may not have been affected so much in this study since statistically few of the QRS complexes were affected (regarding to the size of the data, QRS rarely occurred and during a short time frame). Besides, a simple filtering could overcome in part missing data, and as a matter of fact the method we employed for QRS detection does low-pass filtering [Nygårds and Sörnmo, 1983].

Measuring brain signals with EEG enabled us to perform continuous mental workload monitoring, but only with an offline analysis. Indeed, our algorithm required computing the covariance matrix of EEG signals recorded during the context of use (i.e. here during 3D object manipulation tasks), which would not have been possible if mental workload was to be estimated in real time during these manipulation tasks. The covariance matrix was estimated on all the EEG data collected during the manipulation tasks, and thus could only be estimated once the tasks were completed. In the future, it would be interesting to design a continuous workload estimator that can also be used in real time. To do so, our algorithm could be adapted in two ways: 1) the covariance matrix of the EEG signals recorded during 3D manipulation tasks could be estimated on the first task – or couple of tasks – only, to enable workload estimation in real time on the subsequent tasks; 2) the differences between the calibration context and the use context are likely to be the same across different participants [Samek et al., 2013]. As such, the EEG signals directions that vary between contexts can be estimated on the data from some users, and used to estimate robustly the workload on the data from other users, hence without the need to estimate these variations for a new user, as done in [Samek et al., 2013] for the classification of EEG signals related to imagined hand movements. We will explore these options in the future, which would potentially open the door for robust continuous mental effort estimation during 3D interaction, in real time.

6.5 Conclusion

In this chapter, we have explored a new way to evaluate 3DUI in a more continuous, exocentric and non-interrupting way. In particular we proposed to continuously monitor the mental effort exerted by users of a given 3DUI based on the measure of their brain signals (EEG). We first proposed a method to estimate such level of mental effort from EEG, ECG and EDA signals. EEG outperformed other physiological sensors and we then used the resulting mental effort estimator to study mental workload during a pilot study involving 3D object manipulation tasks with a CubTile. For the first time, we able to transfer the assessment of a construct from a control task (calibration) to a more ecological situation (virtual environment simulating the construction of a bridge). Monitoring workload enabled us to continuously observe when and where the 3DUI and/or interaction technique was easy or difficult to use, unveiling a new path to create better interfaces.

Since the use of EDA and ECG sensors was not likely to improve the overall performance of a system aimed at evaluating HCI – at least in our experimental settings – we did not pursue with these physiological sensors in the next chapter. Although we recorded participants' physiological signals during a calibration task meant to study emotional valence using the IAPS (International Affective Picture System), no sensors or combination of sensors did better than a 57% mean classification accuracy – a score too low for applying the classifier to an ecological task. Therefore, we also discarded the measure of emotions in the studies that followed. Valence calibration – that occurred alongside workload calibration – consisted in the presentation of 140 images, 35 with a negative valence, 35 with a positive valence and 70 with a neutral valence. Since images appeared for only 7s and the sequence were randomly chosen, maybe the time widow was too short for any relevant effect to appear in the physiology – or in the mind.

Another unused calibration occurred during this study with a task that involved error recognition (see part II). We did not analyze this data because there were no markers that we could use to investigate this construct during the 3D manipulation task. This situation highlights one limitation of the current protocol: the software used for the assembly of the bridge was not modified for the purpose of the experiment. As such there was no stimulations (i.e. markers) that was automatically triggered during the task and that could enable an accurate synchronization between physiological recordings and events in the virtual environment – we relied instead on a manual segmentation of the tasks, enough to roughly cluster data but impossible to use for the ERP analyses required by error recognition.

This is why we developed from scratch a virtual environment specifically dedicated at validating the use of EEG for HCI evaluation, that we present in the next chapter. Not only did we establish a protocol that induced a specific amount of workload, but were able to monitor continuously users' mental state. Beside more accurate workload conditions, this was also opportunity to study attention level and error recognition, to measure how intuitive a UI can be and compare various interactions techniques.

7

Validating EEG as an HCI evaluation method

In this chapter, we go beyond the results obtained previously with 3D manipulation tasks. Here the objective is to validate the use of EEG as an evaluation method for HCI in a controlled and carefully crafted environment. We contribute a set of methods to estimate continuously the user's mental workload, attention level and recognition of interaction errors during different interaction tasks.

This work was published in [Frey et al., 2016b].

Special thanks to Maxime Daniel for his hard work all along this project – more in the appendices, section Credits.3.

7.1 Introduction

In HCI, measuring the attention level could help to estimate how much information users perceive. More particularly, in the present work the measure of attention relates to inattentional blindness; i.e. it concerns participants' capacity to process stimuli irrelevant to the task [Cartwright-Finch and Lavie, 2007]. Error recognition relates to the detection by users of an outcome different from what is expected, and *interaction* errors arise when a system reacts in an unexpected way, for example if a touch gesture is badly recognized (see chapter 3). Interaction errors enable to assess how intuitive a user interface is.

Figure 7.1 – We demonstrate how electroencephalography can be used to evaluate human-computer interaction. For example, a keyboard (*left*) can be compared with a touch interface (*middle*) using a continuous measure of cognitive workload (*right*, here participant 4).

We show how these measures can be used to compare different interaction techniques or devices, by comparing here a keyboard and a touch-based interface. Thanks to such framework, EEG becomes a promising method to improve the overall usability of computer systems. Is constitutes a powerful complementary tool for people who develop new interaction techniques, so they could test beforehand new approaches.

In the following sections, we will first describe the virtual environment that we developed, specifically aimed at validating the use of EEG as an evaluation method for HCI. We validated the workload (mental effort) induced by our environment in a first study, with NASA-TLX questionnaires [Hart and Staveland, 1988]. After this pilot study, we will detail how EEG can be put into practice to assess the 3 studied mental states. Finally, during the main study we then employ EEG recordings to measure continuously such workload, altogether with the attention level of participants toward external stimuli and the number of interaction errors they perceived.

To summarize, our main contributions are:

1. To validate the use of EEG as a continuous HCI evaluation method
2. To demonstrate how such tool can assess which interaction technique is better suited for a particular environment
3. To propose a framework that could be easily replicated to improve existing interfaces with little or no modifications

7.2 Virtual 3D maze

This section describes the virtual environment that we purposely developed for the validation of EEG as an HCI evaluation method. Building the interaction from the ground up gave us precise control over the different constructs we wanted to test, i.e. workload, attention and error recognition. The 3D environment uses gamification [Deterding et al., 2011] in order to increase users' engagement and ensure better physiological recordings [Flatla et al., 2011]. Such an environment also

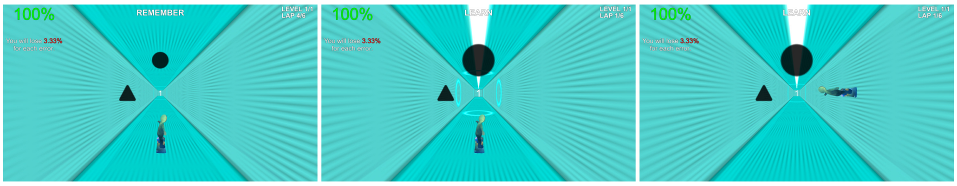

Figure 7.2 – Caption: The virtual environment, where players control a character that moves by itself inside a 3D maze. *Left*: Symbols appear in each tunnel to indicate the possible directions for the next turn; players have to select a particular sequence of symbols/directions. *Middle*: During the "learning" phase, the correct direction is highlighted by a breadcrumb trail and the associated symbol bounces (here the disc on top). *Right*: Controls depend on the position of the character. If the character is on the right side, players have to press *right* in order to go *up*.

enables us to assess these constructs during ecological and realistic interaction tasks. Indeed, such constructs are traditionally evaluated during controlled lab experiments based on protocols from psychology that are vastly different from an actual interaction task, see, e.g., [Grimes et al., 2008].

7.2.1 Overall description

The virtual environment takes the form a maze where players have to learn and reproduce a path by triggering directions at regular intervals (see Figure 7.2). A character displayed with a third person perspective moves by itself at a predefined speed inside orthogonal tunnels (somewhat similar to a *running game*). Soon after the character enters a new tunnel, symbols appear on-screen. Those symbols are basic 2D shapes, such as square, circle, triangle, diamond or star, and their positions (bottom, top, left or right) indicate which directions are "opened". Players must select one of these symbols before the character reaches the end of an intersection, either by pressing a key or touching the screen. If users respond too early – before symbols appeared –, too late, or if they select a direction that does not exist, they loose points and the character "dies" by smashing against a wall, respawning soon after at the beginning of the current tunnel.

The main element of the gameplay consists in selecting the directions in the correct order. Indeed, one level is comprised of two phases. During the "learning" phase a particular sequence of symbols is highlighted; at each symbols' appearance one of them is bouncing to indicate the correct direction. Another cue takes the form of a "breadcrumb trail", a beam of light that precedes the character and points to the right direction (see Figure 7.2, middle). Selecting an available but incorrect direction exist does not result in the character's "death" but still leads to a loss of points. A visual feedback is given to users when they select a direction, the corresponding symbol turns green if the choice is correct and red otherwise. The feedback remains visible until the character reaches the end of the tunnel and turns into the next section. When the sequence is completed and the end of the maze is reached, the character loops over the entire maze so that players have another opportunity to learn the sequence. When the training phase ends the "recall" phase follows, where the symbols are identical but where cues are no more displayed; players have to remember by themselves the right path. Symbols position in each tunnel and symbols sequence are randomly drawn when a new level starts.

Beside learning a sequence, the principal challenge comes from *how* the directions are selected. The third-person view fulfills a purpose: the interface that users are controlling – i.e. keyboard or touch screen – is mapped to the *character position*. Since the character is a futuristic surfer that defies the law of gravity, from time to time it slides by itself from the bottom of the tunnel to one of the walls or to the ceiling. In this latter situation, when the character is upside down, commands are inverted compared to what players are used to, even though symbols remain in the same position. This game mechanism stresses spatial cognition abilities; users have to constantly remain aware of two different frames of reference. For example, if "up" and "left" directions are open in a given tunnel and if the character's position – controlled by the application – is on the right wall, as in Figure 7.2 right, in order to go *up* users have to press *right*. This discrepancy between input and output is a reminder of the problematic often observed with 3D user interfaces, where most users possess a device with 2 degrees of freedom (DOF), such as a mouse, to interact with a 6 DOF environment.

The combination of the game design and game mechanisms herein described offers a wide variety of elements that we put in use so as to investigate users' mental states. We detail below how the study of users' workload, attention and error recognition shaped our choices. In particular, we detail how we tuned the game elements to manipulate the user's workload and attention in controlled ways as well as to trigger interaction errors. Knowing which constructs value (e.g. high or low workload) to expect, we could then validate whether our EEG-based estimates during interaction match these expectations, and thus whether they are reliable.

7.2.2 Manipulating workload

Our virtual environment possesses several characteristics that could be used to induce different levels of mental workload. We can notably adjust 4 parameters:

- *Maze depth*: the number of tunnels players have to cross before reaching the end of the maze, hence the length of the symbols sequence they have to learn. More symbols to learn means more items to be held in the working memory, which increases workload [Grimes et al., 2008, Sternberg, 1966].

- *Number of directions*: at each intersections up to 4 directions are "opened" in the maze; the complexity of the symbols sequence grows as this number increases.
- *Game speed*: the pace of the game can be adjusted to increase temporal pressure. When the speed increases symbols appear sooner and users must respond quicker, thus increasing overall stress [Hart and Staveland, 1988, Maule and Edland, 1997]. In the easier levels the character spends 6s in a tunnel and players must respond within 3s after symbols' appearance; in the hardest levels a tunnel lasts 2s and players have 1s to choose a symbol.
- *Spatial orientation*: in order to keep selecting the correct directions, users have to perform a mental rotation if the character they control jumps from the floor to the walls or to the ceiling. Furthermore, they need to update their frame of reference as often as the character shifts from one side to another. Depending on the spatial ability of users, this mechanism can cause an important cognitive load [Poor et al., 2013].

We used those mechanisms and dimensions to create 4 different difficulty levels for the game: "EASY", "MEDIUM", "HARD" and "ULTRA" (see Table 7.1). These levels affect mostly (symbolic) memory load and time pressure. Indeed, the 3D maze is more about remembering a sequence of symbols or directions rather than spatial navigation per se. Because randomization could create loops in the maze topography and since there were no landmarks, it is unlikely that participants were able to adopt an allocentric strategy.

While the EASY level is designed to be completed with very little effort, the ULTRA level, on the other hand, is designed to sustain a very high level of workload, up to the point that it is barely possible to complete it with no error. While during EASY levels there is no need to perform mental rotations and players have to memorize only 2 symbols that are constrained to either left and right directions, in ULTRA levels the frame of reference changes between each selection and the sequence reaches 5 symbols that could appear in all 4 directions, *and* players have to react thrice as fast. No matter the level, players had 3 "loops" to learn the maze and another set of 3 loops to reproduce the path.

7.2.3 Assessing attention

We relied on stimuli not congruent to the main task in order to probe for inattentional blindness, using the "oddball" paradigm. The oddball paradigm is an experimental design often employed in combination with EEG recordings, as the appearance of rare (i.e. "odd") stimuli among a stream of frequent stimuli (i.e. distractors) triggers a particular

Table 7.1 – Four difficulty levels are created by leveraging on game mechanisms. *Depth*: depth of the 3D maze, hence the number of directions/symbols players have to learn. *Directions*: number of possible directions at each intersection. *Response time*: how much time players have to respond after symbols appearance. *Orientation*: percentage chance that the controlled character changes its orientation before symbols appearance.

Difficulty	Depth	Directions	Resp. time	Orientation
EASY	2	2	3s	0%
MEDIUM	4	3	2.5s	30%
HARD	5	4	2s	60%
ULTRA	5	4	1s	100%

event-related potential within EEG signals [Coull, 1998]. The amplitude of these latter ERP decreases as users are less attentive to stimuli [Fabiani et al., 2007]. Stimuli could be either audio or visual, the advantage of the former being that is does not interfere with the main task in our experimental design.

This mechanism is similar to what was employed in [Burns and Fairclough, 2015] in order to measure how many participants were focused on a video game. While in this latter study audio tones were played externally to the chosen commercial game, we had the opportunity to directly integrate sounds to our virtual environment. As such, while users' character was navigating in the maze, sounds were played at regular intervals, serving as a background "soundtrack" that was consistent with the user experience. 20% of these sounds had a high pitch (odd event) and the remaining 80% had a low pitch (distracting events) – this proportion is on par with the literature [Burns and Fairclough, 2015, Ferrez and Millan, 2008].

Our hypothesis is that the attention level of participants toward sounds – as measured with the oddball paradigm – should decrease as the workload increase, since most of their cognitive resources will be allocated to the main task during the most demanding levels.

7.2.4 Assessing error recognition

EEG could be used to measure interaction errors, i.e. errors originating from an incorrect response of the user interface, that differ from what users were expecting [Ferrez and Millan, 2008]. We have seen in chapter 3, how interaction errors are of particular interest for HCI evaluation since they could account for how intuitive an interface is. In order to test the feasibility of such measure, we decided to implement two different interaction techniques. Both of them use discrete events – i.e. symbols' selection – so that we could more easily synchronize EEG recordings with in-game events later on.

The first technique uses *indirect* interactions by the mean of a keyboard (Figure 7.1, left). In due time, left, right, up or down arrow keys are used to send the character in the tunnel that is situated to *its* left, right, top or bottom. Indeed, we have seen previously that in our virtual environment players have to orientate themselves depending of the position of the character. If the character is moving on the sides, players have to perform a mental rotation of 90°, if it is on the ceiling then the angle is 180°, i.e. commands are inverted.

The second technique uses *direct* interaction by the mean of a touch screen (Figure 7.1, middle). Usually, with touch screen, pointing is co-localized with software events, since users can directly indicate where they want to interact. However, in our case, we decided to mimic exactly the behavior of the keyboard interface. That is to say that with the touch screen as well players have to orientate themselves depending on the position of the character. Hence, if the character is positioned on the *left*, players have to touch the *right* fringe of the screen in order to go *up*. This is mostly counter-intuitive since players have to inhibit the urge to point to the actual direction they want to go; there is a cognitive dissonance.

Since in our experimental design the utilization of the direct (touch-based) interaction is counter-intuitive, we hypothesize that it will lead to an overall higher number of interaction errors compared to the indirect interface (keyboard).

7.2.5 Implementation

We developed our environment using the game engine Unity 5. It ran under Windows 7 64bit, on an Alienware Aurora R4 equipped with an Intel i7-3820 processor, 8GB of RAM and a GeForce GTX 660 Ti graphic card. The screen was a 24-inch multi-touch display (3M M2467PW). During touch-based sessions it was positioned at a comfortable angle for participants.

In order to synchronize accurately in-game events with brain recordings, we used Lab Streaming Layer[1] (LSL) library. LSL is a network protocol dedicated to physiological recordings. It is designed to achieve sub-millisecond accuracy on local networks, however the delay between the data sent over the protocol and actual stimuli (images or sounds) depends on the performance of the software. Even though the refresh rate of the screen was locked at 60Hz, we increased game framerate to 240FPS to prevent any lag. Using OpenVibe 1.0 to acquire EEG signals, we managed to ensure a constant delay of 25ms (SD: 2.5) – ERP detection is particularly sensitive to signals' latency as variations will diminish the amplitude measured in the averaged signals.

Notice to OpenViBE users: set python boxes frequency to 128hz to increase responsiveness.

[1] https://github.com/sccn/labstreaminglayer

7.3 Pilot study: validation of the induced workload level

We designed our virtual environment as a test-bench aimed at inducing several mental states within users. Notably, we implemented several game mechanisms that attempt to modulate mental workload. While we defined a set of parameters using the literature that we adjusted during internal tests, we had to formally validate the mental workload that each one of our 4 different game levels seeks to induce. Each one should be different from the other, and users' workload should range from low to high with the following levels' order: EASY < MEDIUM < HARD < ULTRA.

As such, we conducted a pilot study with no physiological recordings but using the NASA-TLX questionnaire [Hart and Staveland, 1988], a well established questionnaire that accounts for workload.

7.3.1 Protocol

15 participants took part in this study – 4 females, 11 males, mean age 24.53 (SD: 3.00). We used a within-subject experimental design; all participants answered for all 4 difficulty levels. The gaming session occurred with the indirect interaction (keyboard) and started with 2 "training levels", that introduced participants to the game mechanisms. In the first training level, players learned the objective of the game – navigate in the maze and memorize a sequence of items. In the second training level, they discovered how the character could change its orientation by itself. After the completion of this training phase, and once they felt confident enough, participants continued with the main phase of the experiment.

During the main phase of the experiment, participants played once each one of the four main levels (EASY, MEDIUM, HARD or ULTRA). The presentation order was randomly chosen. Immediately after the end of a level, participants were given a NASA-TLX questionnaires to inquire about their mental workload. The questionnaire took the form of a 9-points Likert scale. As in the original questionnaire, it was comprised of 6 items, that assessed mental demand, physical demand, temporal demand, performance, effort and frustration. For example, effort was rated from "low" to "high" with the following question: "How hard did you have to work to accomplish your level of performance?". The experiment lasted approximately 25 minutes and finished once participants played all 4 levels and filled the corresponding NASA-TLX questionnaires.

7.3.2 Results

For each participant and each level of difficulty, we averaged the 6 items of the NASA-TLX questionnaire and normalized the scales from [1;9] to

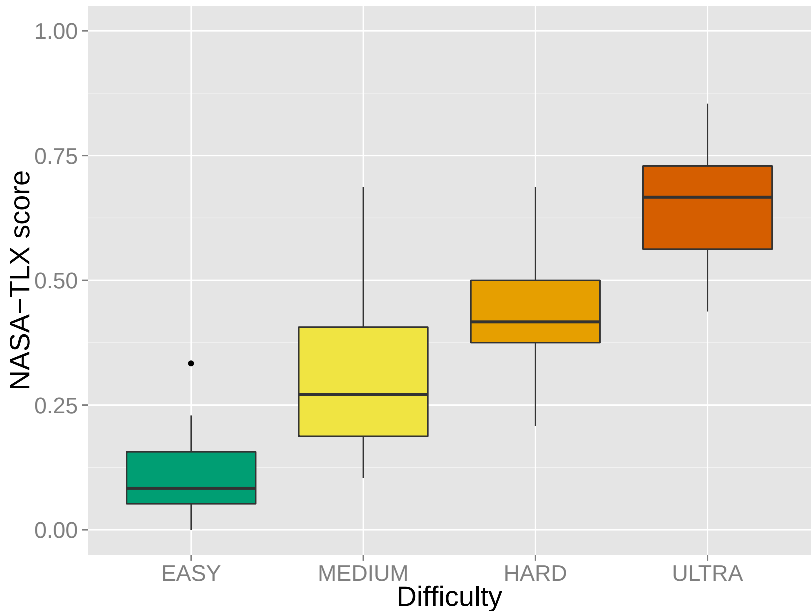

Figure 7.3 – NASA-TLX scores obtained during the pilot study. Each difficulty level differs significantly from the others ($p < 0.01$).

[0;1] – except for the "performance" item, that was normalized from [1;9] to [1;0] because its scale is in reverse order compared to the other items ("1" for "good" and "9" for "poor").

The resulting averaged scores are: EASY: 0.11 (SD: 0.09); MEDIUM: 0.32 (SD: 0.17); HARD: 0.43 (SD: 0.13); ULTRA: 0.65 (SD: 0.13) – see Figure 7.3.

A repeated measures analysis of variance (ANOVA) showed a significant effect of the difficulty factor over the NASA-TLX scores and a post-hoc pairwise Student's t-test with false discovery rate (FDR) [Noble, 2009] correction showed that each levels differed significantly from the others ($p < 0.01$).

7.3.3 Discussion

In this pilot study, we demonstrated through questionnaires that each difficulty level presented in Table 7.1 induces a different workload level. Hence, we can use our virtual environment as a baseline to assess the reliability of analogous EEG measures and put into perspective this new evaluation method.

7.4 EEG IN PRACTICE

In this section, we describe the calibration tasks that we implemented in order to measure workload and error recognition. We chose to use standard tasks, validated by the literature, so that our findings could be easily reproduced. Moreover, as shown later in this chapter, using a single of these tasks to calibrate each construct estimator was enough to obtain reliable estimations of such constructs during different and complex interaction tasks.

Concerning attention, we did not develop a dedicated task *per se* for its calibration. Since the audio probes were already integrated to our virtual environment, we simple used a specific level of our game.

7.4.1 Calibration of workload

We used the N-back task protocol to induce 2 different cognitive loads and calibrate our workload estimator. We used the exact same protocol and implementation as the one described in the previous chapter, when we studied 3D manipulations tasks. All details are there – there were also 360 trials, divided in 6 blocks of 60 letters that alternated between 0-back and 2-back conditions.

7.4.2 Calibration of error recognition

We replicated the protocol described in [Ferrez and Millan, 2008] to calibrate the system regarding error recognition, since it could be considered as a standard approach to evaluate interaction errors. The task simulates a scenario in which users control the movements of a robot. The robot appears on screen and has to reach a target. At each turn users order the robot to go right or left in order to reach the target as fast as possible (with the least steps). Except that the robot may understand badly the given command. This is simulated by some trials during which the command is (on purpose) erroneously interpreted; hence an interaction error happens.

The calibration task is a simplified version of this scenario: the robot is pictured by blue rectangle on screen that users control with the arrow key, the target is represented by a blue outline. The robot is constrained to the X axis and along this axis there are only 7 different positions both for the "robot" and the target (see Figure 7.4).

We choose a ratio for the occurrence of interaction errors that is consistent with the literature. 80% of the movements matched the actual key pressed and for the other 20% the "robot" moved in the opposite direction. It was necessary not to balance both events since the kind of EEG features that interaction errors are triggering relates to the oddball paradigm – it is called an "error related potential", ErrP [Ferrez and Millan, 2008]. A timer was set to prevent the appearance of artifacts, such as muscle movements, within EEG recordings. The rectangle moves 1s after a key was pressed, and after movement completion users have to wait another 1s before they could press a key again. Rectangle turned yellow as long as users could not control the rectangle.

Do not confuse an ErrP for an ERP!

A trial is completed once the robot reaches the target. A trial fails if after 10 attempts the robot is not yet on target. Whenever the trial is a success of a failure, the screen is reinitialized and a new trial begins, with a new position for the robot and the target. The last trial occurred

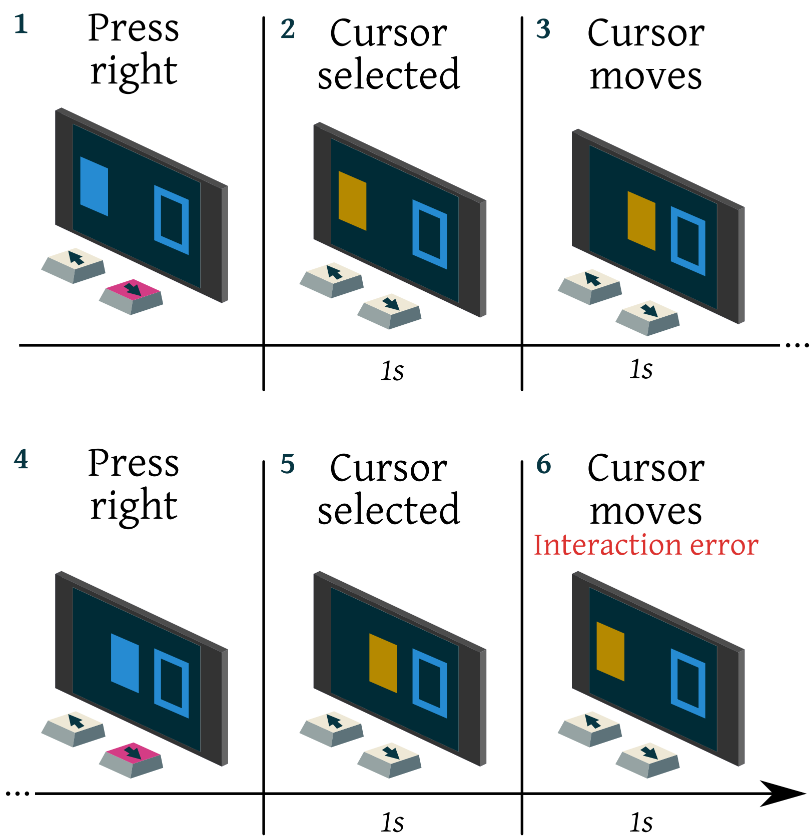

Figure 7.4 – Error recognition calibration task. Users control a blue filled rectangle. They have to move it to an outlined target by pressing left or right arrow key. The rectangle moves 1s after key press and users have to wait another 1s to move it again. 20% of the time, the rectangle goes in the opposite direction, thus causing an interaction error.

II.7

after 350 interactions were performed. On average this calibration phase lasted 15 minutes.

7.4.3 Calibration of attention

The calibration of attention occurred within a simplified version of the virtual environment. Users did not have to control the character during this special level, it was moving by itself through the maze. They were asked to watch the character and count in their head how many times they heard the "odd" sound – the sound of a high pitched bell lasting 200ms. The distractor was a low pitched beat of 70ms – we did not use pure tones to improve users experience. The pace of the game was adjusted so that a sound (target or distractor) was played every second. Once again, since the probes for attention relies to the oddball paradigm, we chose a 20% of appearance of the target event. The calibration lasted for approximately 7 minutes, after 350 sounds were played. Note that participants were instructed to count the "odd" events *only* during the calibration phase, and *not* during the completion of the 3D maze.

7.5 Main study: EEG as an evaluation method

The main study consisted in the evaluation of the game environment with two different types of interface using EEG recordings. As such we created a 4 (difficulty: EASY, MEDIUM, HARD, ULTRA) $\times$ 2 (interaction: KEYBOARD *vs* TOUCH) within-subject experimental plan. Our hypotheses are:

1. The workload index measured by EEG is higher in TOUCH and increases with the difficulty, reflecting NASA-TLX scores obtained during the pilot study.
2. The attention level of participants decreases as the difficulty increases.
3. The TOUCH condition induces a higher number of interaction errors compared to the KEYBOARD condition.

The gaming phase was split in two, one for each interaction technique. In order to avoid a too tedious experiment, participants alternated between those game sessions and the 3 calibration tasks (workload, attention and error recognition). Since the analysis were performed offline, there was no need to cluster all the calibrations at the beginning of the experiment. The resulting variety of the exercises helped participants to remain engaged during the various tasks, which is of importance since from the reliability of the signals during the calibration depends the accuracy of the final measures.

The order of the gaming sessions and calibration phases was counterbalanced between participants following a latin square (12 combinations possible, see Figure 7.5). After the experiment, the signals gathered from the calibration tasks were processed in order to evaluate both the virtual environment (difficulty levels) and the chosen interaction techniques.

7.5.1 Apparatus

Besides the material related to the virtual environment, EEG signals were acquired at a 512Hz sampling rate with 2 g.tec g.USBamp amplifiers. This medical-grade equipment handles 32 electrodes. In the international 10-20 system, electrodes were placed at AF3, AFz, AF4, F7, F3, Fz, F4, F8, FC3, FCz, FC4, C5, C3, C1, Cz, C2, C4, C6, CP3, CPz, CP4, P7, P3, Pz, P4, P8, PO7, POz, PO8, O1, Oz and O2 sites.

12 participants took part in this study – 3 females, 9 males, mean age 26.25 (SD: 3.70). All of them reported a daily use of tactile interfaces.

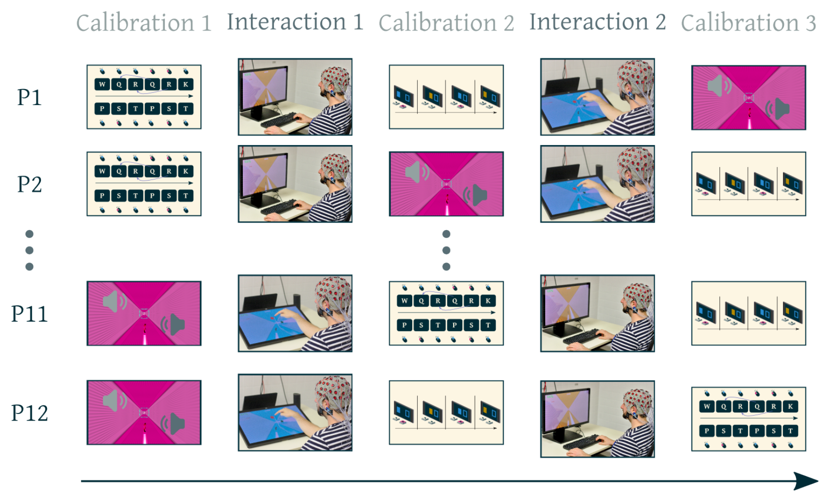

Figure 7.5 – The order of the 3 calibration tasks and 2 interactions techniques was counter-balanced between the 12 participants to improve engagement.

The experiment occurred in a quiet environment, isolated from the outside. There were two experimenters in the room and the procedure comprised the following steps:

1. Participants entered the room, read and signed an informed consent form and filled a demographic questionnaire.
2. While one of the experimenter installed an EEG cap onto participants' heads, the other experimenter introduced participants to the virtual environment. They played 2 training levels and the 4 main levels in an increasing order of difficulty. They could redo some levels if they did no feel confident enough.
3. One of the 3 calibration tasks occurred (workload, attention or error recognition).
4. Participants played to the game using one of the 2 interaction techniques (KEYBOARD or TOUCH). The four levels of difficulty (EASY, MEDIUM, HARD, ULTRA) appeared twice during the session, in a random order. In the case of TOUCH, a dedicated training session occurred beforehand so that participant could get used to this interaction technique.
5. Another calibration task occurred, different from step 3.
6. Participants tested the second interaction technique. As in step 4, in case of TOUCH it was preceded by a training session, that lasted until participants felt confident enough to proceed to the main task.
7. Participants performed the last remaining calibration task.

A game session (steps 4 and 6) took approximately 20 minutes and the whole experiment lasted 2h.

7.5.2 EEG Analyses

The calibration tasks were used to train a classifier specific to each one of the studied mental states. Classifiers were calibrated separately for each participant, user-specific classifiers ensuring maximal EEG classification performances. We used EEGLAB 13.4.4b and Matlab R2014a to process EEG signals offline. While the descriptive EEG features associated to workload relate to the frequency domain, the features associated to attention and error recognition relate to temporal information. We detail below those pipelines.

7.5.2.1 *Processing workload*

The signal processing concerning workload was exactly the same as what we successfully applied in our previous study – we employed the EEG pipeline that was the most effective.

Once again, we used 2s time windows extracted during the N-back tasks to train the classifier. We filtered EEG signals in the delta (1-3 Hz), theta (4-6 Hz), alpha (7-13 Hz), beta (14-25 Hz) and gamma (26-40 Hz) bands. To reduce features' dimension, we used for each band a set of Common Spatial Patterns (CSP) spatial filters and reduced the 32 original channels down to 6 "virtual" channels. Since the calibration (N-back task) and use contexts (virtual environment) differs also substantially in this study, we used a regularized version of these filters called stationary subspace CSP (SSCSP) – refer to Chapter 6.2.1 for details.

Once calibrated, this classifier can be used to estimate workload levels on new data, which we will use to estimate users' mental effort whilst interacting with the virtual environment.

7.5.2.2 *Processing attention and error recognition*

Since in our experimental design both attention and error recognition relied on the oddball paradigm, they share the same signal processing.

We selected time windows of 1s, starting at the event of interest (i.e. sounds for attention, rectangle's movements for error recognition). In order to utilize temporal information, feature extraction relied on regularized Eigen Fisher spatial filters (REFSF) method [Hoffmann et al., 2006]. Thanks to this spatial filter, specifically designed for ERPs classification, the 32 EEG channels were reduced to a set of 5 vectors.

This time the pipeline is similar to the study measuring visual comfort with stereoscopic displays.

To reduce furthermore the number of features, we decimated the signal by a factor 32. The "decimate" function of Matlab was used, which applies a low-pass filter after decimation to prevent aliasing. As a result, there was 80 features by epoch (5 channels $\times$ 512Hz $\times$ 1s / 32).

7.5.2.3 Classification

We used shrinkage LDA (linear discriminant analysis) as a classifier; which is more efficient compared to regular LDA when it comes to a high number of features [Ledoit and Wolf, 2004].

For each construct there was two steps: first we used the data collected during the calibration tasks to estimate the performance of the classifiers. Second, we studied the output of the different classifiers to evaluate the virtual environment and test our hypotheses.

To assess the classifiers' performance on the calibration data, we used 4-fold cross-validation (CV). I.e. we split the collected data into 4 parts of equal size, selecting trials randomly, used 3 parts to calibrate the classifiers and tested the resulting classifiers on the unseen data from the remaining part. This process occurred 3 more times so that in the end each part was used once as test data. Finally, we averaged the obtained classification accuracies. The accuracy was measured using the area under the receiver-operating characteristic curve (AUROCC). The AUROCC is a metric that is robust against unbalanced classes, as it is the case with attention and error recognition (20% of targets, 80% of distractors). A score of "1" means a perfect classification, a score of "0.5" is chance.

Compared to previous studies, here we switched to AUROCC metric to account for unbalanced classes.

Once the classifiers were trained thanks to the calibrations tasks, we could use them on the EEG signals acquired while participants were interacting with the virtual environment, to estimate the different constructs values.

For workload, we used 2s long sliding time windows that were overlapping by 1s, to extract signals and feed the classifier. From the outputs that was produced by the LDA classifier for each participant (i.e., the distance to the separating hyperplane), we first removed outliers by iteratively removing one outlier at a time using a Grubb's test with $p = 0.05$, until no more outlier was detected [Grubbs, 1969]. We then normalized the outlier-free scores between -1 and +1. As such, for all participants a workload index close to +1 represents the highest mental workload they had to endure while they were playing. It should come close to the 2-back condition of the calibration phase. On the opposite, a workload index close to -1 denotes the lowest workload, similar to the 0-back condition.

Outliers removal was added to the pipeline to account for the greater heterogeneity of the LDA outputs between subjects with the other constructs, but it does not concern more than few data points.

The process was similar for attention, but we only extracted epochs that corresponded to the target stimuli onset, i.e. when the high pitch sound was played. Note that contrary to [Burns and Fairclough, 2015], that studied the amplitudes of ERPs and did not use the data gathered during the calibration phase, here we kept the machine learning approach. As such, the resulting scores can be seen as a confidence index of the LDA classifier about whether or not participants noticed odd events while they were playing.

As for the classifier dedicated to error recognition, the processing of its output differs. Indeed, we could not assume which interaction yielded or not an interaction error, i.e. if and when participants perceived a discrepancy between what they intended to do and what occurred. Consequently, we simply counted over an entire game session the number of times the classifier labelled an interaction as being erroneous in the eye of the participants.

7.5.3 Results

Unless otherwise noted, we tested for significance using repeated measures ANOVA. For significant main effects, we used post-hoc pairwise Student's t-test with FDR correction.

Table 7.2 – Classification accuracy during the calibration tasks for the 3 measured mental states (AUROCC scores).

	P1	P2	P3	P4	P5	P6	P7
Workload	0.85	0.93	0.98	0.95	0.97	0.97	0.79
Attention	0.83	0.82	0.96	0.81	0.85	0.90	0.82
Err. recog.	0.88	0.57	0.90	0.90	0.86	0.90	0.78
	P8	P9	P10	P11	P12		*Avg*
Workload	0.87	0.87	0.98	0.95	0.94		*0.92*
Attention	0.82	0.86	0.92	0.88	0.83		*0.86*
Err. recog.	0.80	0.88	0.78	0.85	0.74		*0.82*

7.5.3.1 *Workload*

On average, the classifier's AUROCC score during the training task was 0.92 (SD: 0.06) – see Table 7.2. Over the test set there were on average 2171 data points per participant across all condition (time windows).

The statistical analysis of the classifier output during the game session showed a significant effect of the difficulty factor ($p < 0.01$); the workload index increasing along the difficulty of the levels (Figure 7.6, top). The post-hoc analysis showed that all difficulty levels significantly differs one from the other with $p < 0.01$; except for the MEDIUM level, which differs from EASY with $p < 0.05$ and with HARD only by a margin ($p = 0.11$).

There was as well as a significant effect of the interaction factor ($p < 0.01$), the workload being higher on average during the TOUCH condition. There was no interaction between difficulty and interaction factors.

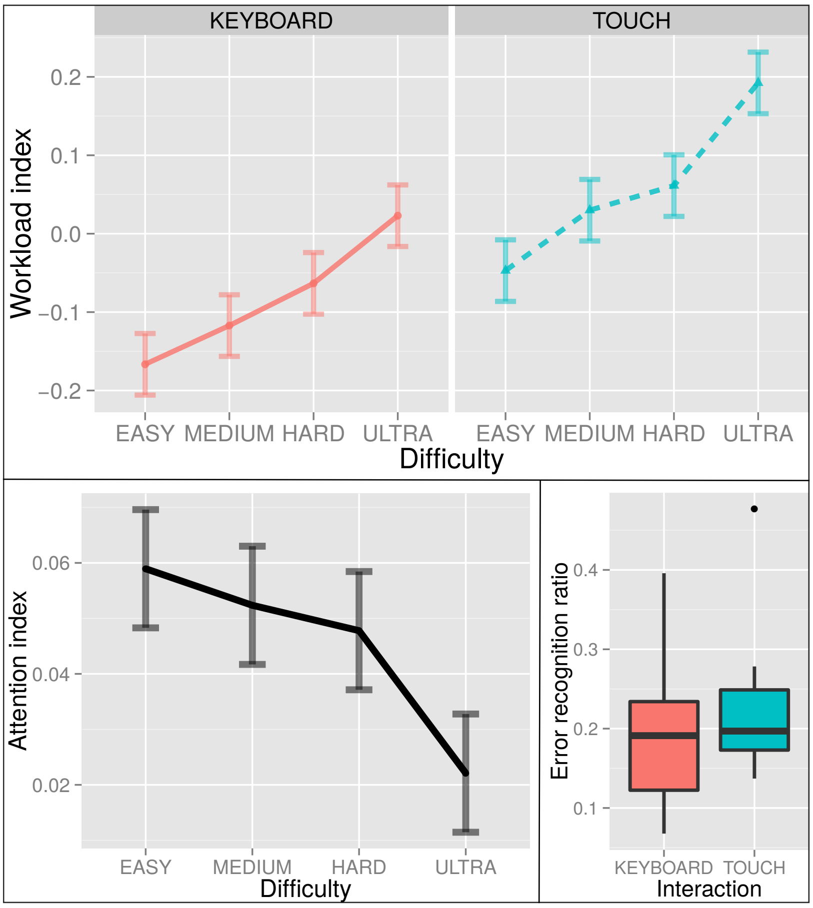

Figure 7.6 – EEG measures. *Top*: The workload index significantly differs across difficulties and between interaction techniques. *Bottom left*: The attention index significantly differs across difficulties. *Bottom right*: The number of interaction errors differs by a tendency between KEYBOARD and TOUCH.

7.5.3.2 *Attention*

On average, the classifier's AUROCC score during the training task was 0.86 (SD: 0.05) – see Table 7.2. Over the test set there were on average 497 data points per participant across all conditions (odd events).

The statistical analysis of the classifier output during the game session showed a significant effect of the difficulty factor ($p < 0.01$) but not of the interaction. The attention index decreases as the difficulty decreases (Figure 7.6, bottom left). The post-hoc analysis showed that the ULTRA level significantly differs from the others ($p < 0.05$).

7.5.3.3 *Error recognition*

On average, the classifier's AUROCC score during the training task was 0.82 (SD: 0.10) – see Table 7.2. Over the test set there were on average 388 data points per participant across all conditions (interactions).

Due to the nature of the data (numbers of interaction errors across entire game sessions), we used a one-tailed Wilcoxon Signed Rank Test to stress our hypothesis. The number of interaction errors differs by a tendency (p = 0.08) between the KEYBOARD and the TOUCH conditions. 19% of the interactions (SD: 9%) was labelled as interaction errors by the classifier for KEYBOARD *vs* 22% (SD: 9%) for TOUCH (Figure 7.6, bottom right).

7.5.4 Behavioral measures

Besides EEG metrics, we had the opportunity to study participants' reaction time and performance so as to get a clearer picture of their user experience.

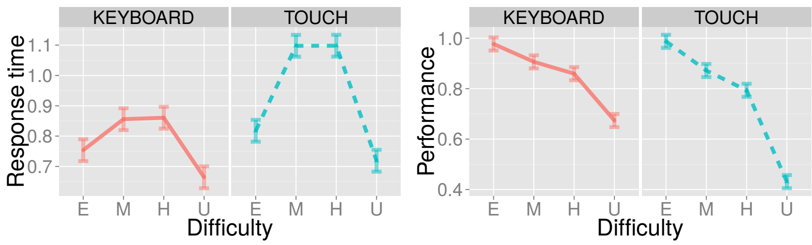

Figure 7.7 – Behavioral measures: reaction time in seconds (left) and performance (proportion of correctly selected directions – right) significantly differs between difficulty levels and interactions. E: EASY, M: MEDIUM, H: HARD, U: ULTRA.

7.5.4.1 *Reaction time*

There was a significant effect of both the difficulty and interaction factor, as well as an interaction effect between them (p < 0.01). Post-hoc tests showed that all difficulty levels differ one from the other (p < 0.01), except for MEDIUM and HARD, which do not differ significantly (p = 0.91).

The mean reaction times (SD) were respectively for EASY, MEDIUM, HARD and ULTRA: 0.78s (0.14), 0.97s (0.18), 0.98s (0.15) and 0.69s (0.06). Mean reaction time for KEYBOARD: 0.78 (SD: 0.12); for TOUCH: 0.93 (SD: 0.13). See Figure 7.7, left. Note that users had less time to respond during higher difficulty levels.

7.5.4.2 *Performance*

The performance was computed as the ratio between the number of correct selections and the total number of interactions. There was a significant effect of both the difficulty and interaction factor, as well as an interaction effect between them (p < 0.01). Post-hoc tests showed that all difficulty levels differ one from the other (p < 0.01).

II.7

The mean performance (SD) was respectively for EASY, MEDIUM, HARD and ULTRA: 98% (3), 89% (12), 83% (17) and 55% (21). Mean performance for KEYBOARD: 85% (SD: 13); for TOUCH: 77% (SD: 13). See Figure 7.7, right.

7.5.5 Discussion

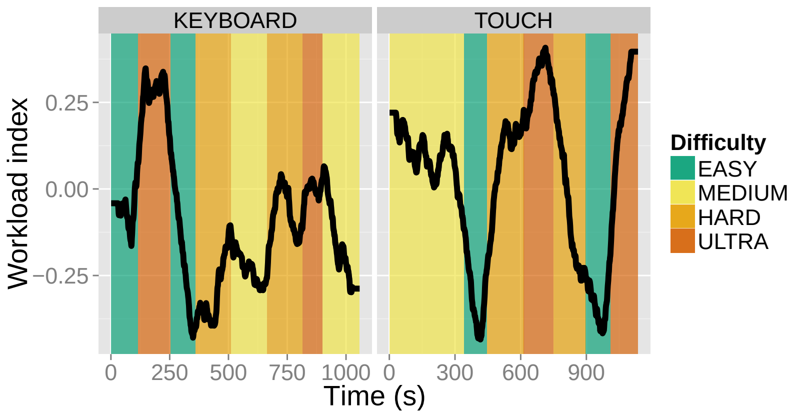

Figure 7.8 – Workload index over time for participant 3 – 60s smoothing window. *Left*: KEYBOARD condition, *right*: TOUCH condition. Background color represents the corresponding difficulty level.

Most of the main hypotheses are verified. The workload index as computed with EEG showed significant differences that match the intended design of the difficulty levels. It was also shown that in the highest difficulty level the attention of participants toward external stimuli was significantly lower – i.e. inattentional blindness increased. Concerning the interaction techniques, the number of interaction errors as measured by EEG was higher with the TOUCH condition, but this is a tendency and not a significant effect. The workload index, on the other hand, was significantly higher in the TOUCH condition compared to the KEYBOARD condition.

Thanks to the ground truth obtained during the pilot study with the NASA-TLX questionnaire, these results validate the use of a workload index measured by EEG for HCI evaluation and set the path for two other constructs: attention and error recognition. Beside the evaluation of the content (i.e. difficulty levels) we were able to compare two interaction techniques. These are promising results for those who seek to assess how intuitive a UI is with exocentric measures.

In this study, we chose to use the particularity of the touch screen to make the task *more* difficult. Indeed, while we used a touch screen for its possibility of direct manipulation, we kept the character as a frame of reference, resulting in input commands that were (patently) not co-localized with output directions. Besides results denoting the differences between the conditions, participants also spontaneously reported how non intuitive this condition was. We wanted to investigate our evaluation method on a salient difference at first. Then our framework could well be employed to go further; for example seeking physiological differences between direct and indirect manipulation interfaces in more traditional tasks.

It is interesting to note how those EEG measures could be combined with existing methods to broaden the overall comprehension of the user experience. For instance, while we did show significant differences across difficulty levels and between interaction techniques with behavioral measures (reaction time and performance index), EEG measures could help to understand the underlying mechanisms. Because we have a more direct access to brain activity, we can make assumptions about the cause of observed behaviors. For example participants' worse performance with TOUCH than with KEYBOARD could be due to the fact that they anticipate less the outcomes of their actions (more interaction errors); the higher reaction time may not only be caused by the interface *per se*, but by a higher workload. And while participants manage to cope with the fast pace of the ULTRA level (the smallest reaction time), the increase in perceptual load lower their awareness to task-irrelevant stimuli

Additionally, while the performances obtained at the easy, medium and hard levels are very similar with the keyboard and the touch screen – see Figure 7.7, right –, the analysis of the workload levels from EEG reveals that the workload was significantly higher in the TOUCH condition, meaning that users had to perform significantly more mental efforts to reach the same performance. This further highlights that EEG-based measures do bring additional information that can complement traditional evaluations such a behavioral measure.

Measuring users' cognitive processes such as workload and attention may prove particularly useful to assess 3D user interfaces (3DUI), since they are known to be more cognitively demanding. They require users to perform 3D mental rotation tasks to successfully manipulate objects or to orientate themselves in the 3D environment. Moreover, the usual need for a mapping between the user inputs with limited degrees-of-freedom (DOF) and the corresponding actions on 3D objects makes 3DUI usually difficult to assess and design. Indeed, 3D environments possess typically 6 DOF while mouses possess only 3 DOF, and multiple DOF input controls demand complex coordination [Zhai and Milgram, 1998]. We reproduced part of this problematic with our game environment and obtained coherent results from EEG measures.

Above all, an evaluation method based on EEG enables a continuous monitoring of users. The intended use case of our framework is to enroll dedicated testers that would wear the EEG equipment and perform well during the calibration tasks. As a matter of fact, the best performer during workload calibration (participant 3 in Table 7.2) shows patterns that clearly meet the expectations concerning both difficulty levels and interactions, as pictured in Figure 7.8.

7.6 Limitations and future challenges

Although using EEG measures as an evaluation method for HCI was proven conclusive regarding workload – we obtained a continuous index on par with a ground truth based on traditional questionnaires – the two other constructs we studied could benefit from further improvements.

Despite the direct interaction (TOUCH) being more disorienting for users than the indirect one (KEYBOARD), the recognition of interaction errors differed only by a tendency. This could be explained by the fact that the calibration task was too dissimilar to the virtual environment. Notably, while there was few and slow paced events during the calibration, users were confronted to many stimuli while they were playing, hence overlapping ERPs must have appeared within EEG, which that may have disturbed the classifier. A calibration task closer to real-life scenarios than the one described in [Ferrez and Millan, 2008] should be envisioned. Such task should remain generic in order to facilitate the dissemination of EEG as an evaluation method for HCI.

II.7

Another mean to facilitate the transfer of the classification between a standard task and the evaluated HCI lies in signal processing. Indeed, if our results demonstrate that the EEG classification of workload could be transferred from the N-back tasks to a dissimilar virtual environment and user interface, we benefited from spatial filters that specifically take into account the variance between calibration contexts and use contexts – the stationary subspace CSP that we used both in this chapter and in chapter 6. Since ERPs may also slightly differ in amplitudes and delays between calibration and use contexts, in the future, it would be worth designing similar approaches to optimize temporal or spatial filters for ERPs as well.

Concerning the nature of the measures we made, our protocol for assessing participants' attention level is inspired by [Burns and Fairclough, 2015]. In this latter study, however, authors sought *immersion*. Their assumption was that the less players were attentive to external stimuli, the more they were immersed in the game. Because in the present study the stimuli not congruent to the tasks were embedded directly in the virtual environment, we assumed that the less players were attentive to sounds the lower was their overall attention level. This is an open issue, and one may try in the future to look more closely if it is either immersion on distraction that change such attention level, and in which cases being less aware of external stimuli is a concern.

Influence of physical characteristics and the need for involvement is not limited to EEG, this is also the case for fNIRS, among others.

The reliability of mental states' measures is strongly correlated to the quality of EEG signals. The hardware is not the sole factor, though, there is also a lot of variability between individuals. Brain patterns differs from one to another. One's brain activity is neither "wrong" or "good", but depending on the considered signal processing the system *will* pickup more easily features with some. Even anecdotal body differences, such as scalp's thickness, influence EEG recordings.

Participants' mindset during the recordings is another the factor influencing EEG signals. Their awareness and involvement toward the tasks improve system's accuracy. The form of the calibration tasks could be enhanced to engage more users, for example through gamification [Flatla et al., 2011] – and our virtual environment proved to be suitable to do so. Whereas our participants were volunteers enrolled among students, in the end the outcome of an evaluation method based on EEG should be strengthened by recruiting dedicated testers, using as selection criteria how reliably the different constructs could be estimated from their EEG signals during calibration tasks.

Finally, one should acknowledge that when it comes to recordings as sensitive as EEG, artifacts such as the ones induced by muscular activity are of major concern. The way we prevented the appearance of such bias in the present study is threefold. 1) The hardware we used – active electrodes with Ag/AgCl coating – is robust to cable movements, see e.g., [Wilson et al., 2012b]. 2) The classifiers were trained on features not related to motion artifacts or motor cortex activation. 3) The position of the screen during the "touch" condition minimized participants' motion, and gestures occurred mostly before the time window used for detecting interaction errors. These precautions are important for the technology to be correctly apprehended.

To further control for any bias in our protocol, we ran a batch of simulations where the labels of the calibration tasks had been randomly shuffled, similarly to the verification process described in chapter 6. Should artifacts bias our classifiers, differences would have appeared between the KEYBOARD and TOUCH conditions even with such random training. Among the 20 simulations that ran for each of the 3 constructs (workload, attention, error recognition), none yielded significant differences.

7.7 Conclusion

In this chapter, we demonstrated how brain signals – by the mean of electroencephalography – could be put into practice in order to obtain a continuous evaluation of different interaction techniques, their ergonomic pros and cons. In particular, we validated an EEG-based workload estimator that does not necessitate to modify the existing software. Furthermore, we showed how users' attention level could be evaluated using background stimuli, such as sounds. Finally, we investigated how the recognition of interaction errors could help to determine the best user interface. Shaping a set of limited mental tasks – e.g. a sole workload calibration task to evaluate a broad range of interfaces and situations – will considerably ease the application of an evaluation framework based on EEG.

Being able to estimate these three constructs – workload, attention and error recognition – continuously during realistic and visually complex interaction tasks opened new possibilities. Notably, it enabled us to obtain additional and more exocentric metrics of user experience, based on the users' cognitive processes. It also provided us with additional insights that traditional measures (e.g. behavioral measures) could not reveal. To sum up, this suggests that combined with existing evaluation methods, EEG-based evaluations tools as the ones proposed here can help to understand better the overall user experience. Future work should apply this framework to other contexts and may refine the distinction between, from the one hand, the evaluation of the interface and, from the other hand, the evaluation of the interaction technique.

This study is a step forward from the assumptions made in the first part of this thesis. However, for such approach to disseminate, physiological sensors in general and EEG devices in particular must become widely available. The need for sensors practical to use is a quest that has to be completed before we could effectively benefit the HCI field. I have attempted to tackle some of these issues during my thesis, and this more technical aspect of my thesis is presented in the appendices.

PART III

FOSTERING SOCIAL INTERACTIONS

This part starts the shift toward practical applications of physiological computing among the general public. One seeks how a pervasive feedback based on users' own heart rate can improve the social presence of embodied agents; another how that same physiological activity can foster a new kind of interaction between several board game players.

These two works own a debt to the young PhyCS conference (International Conference on Physiological Computing System). The idea of the first study came to my mind during the first instalment of PhyCS in 2014 – . . . where I presented my first paper ever – and was presented the year after at PhyCS '15 [Frey, 2015].

PhyCS '14 was also the time when I discovered remote PPG, which seemed just the right technology to avoid the cumbersomeness that too often goes with physiological sensors. While we had been discussing for some months after that within the team of the possibility of an application that could combine a "bluff game" and heart rate sensing, it's when I came back from PhyCS '15, fuelled by a replenished motivation, that I found the will to finally put that idea into practice.

A modest attempt at exploring which "killer app" could help to disseminate even more physiological computing.

8

Physiological similarity-attraction

Physiological sensors are gaining the attention of manufacturers and users, as denoted by devices such as smartwatches or by the popularity of smartphone apps that track heart rate during fitness activities. Soon, physiological monitoring could become widely accessible and transparent to users, especially since remote sensing is within reach (see appendix D).

We demonstrate here how one could take advantage of this situation to increase users' engagement and enhance user experience in human-agent interaction. We created an experimental protocol involving embodied agents – "virtual avatars". Those agents were displayed alongside a beating heart. We compared a condition in which this feedback was simply duplicating the heart rates of users to another condition in which it was set to an average heart rate. Results suggest a superior social presence of agents when they display feedback similar to users' internal state. This physiological "similarity-attraction" effect may lead, with little effort, to a better acceptance of agents and robots by the general public.

8.1 Introduction

Covert sensing of users' physiological state is likely to open new communication channels between human and computers. When anthropomorphic characteristics are involved – as with embodied agents – mirroring such physiological cues could guide users' preferences in a cheap yet effective manner.

One aspect of human-computer interaction (HCI), albeit difficult to account for, lies in users' engagement. Engagement may be seen as a way to increase performance, as in the definition given by [Matthews et al., 2002] for task engagement: an "effortful striving towards task goals". In a broader acceptation, the notion of engagement is also related to fun and accounts for the overall user experience [Mandryk et al., 2006] (see chapter 3). Several HCI components can be tuned to improve engagement. For example, content and challenge need to be adapted and renewed to avoid boredom and maintain users in a state of flow [Berta et al., 2013]. It is also possible to study interfaces: [Karlesky and Isbister, 2014] use tangible interactions in surrounding space to spur engagement and creativity. When the interaction encompasses embodied agents – either physically (i.e. robots) or not (on-screen avatars) – then anthropomorphic characteristics can be involved to seek better human-agent connections.

Following the affective computing outbreak [Picard, 1995], studies using agents that possess human features in order to respond to users with the appropriate emotions and behaviors began to emerge. [Prendinger et al., 2004] created an "empathic" agent that serves as a companion during a job interview. While playing on empathy to engage users more deeply into the simulation was conclusive, the difficulty lies in the accurate recognition of emotions. Even using physiological sensors, as did the authors with electrodermal activity and electromyography, no signal processing could yet reach an accuracy of 100%, even on a reduced set of emotions – see [Lisetti and Nasoz, 2004] for a review.

Humans are difficult to comprehend for computers and, still, humans are more attracted to others – human or *machine* – that match their personalities [Lee and Nass, 2003]. This finding is called "similarity-attraction" in [Lee and Nass, 2003] and was tested by the authors by matching the parameters of a synthesized speech (e.g. paralinguistic cues) to users, whenever they were introverted or extroverted. An analogous effect on social presence and engagement in HCI has been described as well in [Reidsma et al., 2010], this time under the name of "synchrony" and focusing on nonverbal cues (e.g. gestures, choice of vocabulary, timing, ...). Unfortunately, being somewhat linked to a theory of mind, such improvements lean against tedious measures, for instance psychological tests or recordings of users' behaviors. What if the similarity-attraction could be effective with cues that are much simpler and easier to set up?

Indeed, at a lower level of information, [Slovák et al., 2012] studied how the display of heart rate (HR) could impact social presence during human-human interaction. They showed that, without any further processing than the computation of an average heartbeat, users did report in various contexts being closer or more connected to the person with whom they shared their HR. We wondered if a similar effect could be obtained between a human and a machine. Moreover, we anticipated the rise of devices that could covertly measure physiological signals, such as the Kinect 2, which can use its cameras (color and infrared) to compute users' HRs – the use of video feeds to perform volumetric measurements of organs is dubbed as "photoplethysmography" [Kranjec et al., 2014].

Consequently, we extended on the theory and **we hypothesized that users would feel more connected toward an embodied agent if it displays a heart rate similar to theirs, *even if users do not realize that their own heart rates are being monitored*.**

By relying on a simple mirroring of users' physiology, we elude the need to test users' personality [Lee and Nass, 2003] or to process – and eventually fail to recognize – their internal state [Prendinger et al., 2004]. Creating agents too much alike humans may provoke rejection and deter engagement due to the uncanny valley effect [MacDorman, 2005]. Since we do not emphasize the link between users' physiological cues and the feedback given by agents, we hope to prevent such negative effect. The similary-attraction applied to physiological data should work at an almost subconscious level. Furthermore, implicit feedback makes it easier to improve an existing HCI. As a matter of fact, only the feedback associated with the agent has to be added to the application; feedback that can then take a less anthropocentric form – e.g. see [Harrison et al., 2012] for the multiple meanings a blinking light can convey and [Huppi et al., 2003] for a use case with breathing-like features. Ultimately, our hypothesis proved robust, it could benefit to virtually any human-agent interaction, augmenting agent's social presence, engaging users.

The following sections describe an experimental setup involving embodied agents that compares two within-subject conditions: one condition during which agents display heartbeats replicating the HR of the users, and a second condition during which the displayed heartbeats are not linked to users. Our main contribution is to show first evidence that displaying identical heart rates makes users more engaged toward agents.

8.2 Experiment

The main task of our HCI consisted in listening to embodied agents while they were speaking aloud sentences extracted from a text corpus, as inspired by [Lee and Nass, 2003]. When an agent was on-screen, a beating heart was displayed below it and an audio recording of a heart pulse was played along each (fake) beat. This feedback constituted our first within-subject factor: either the displayed HR was identical to the one of the participant ("human" condition), either it was set at an average HR ("medium" condition). The HR in the "medium" condition was ranging from 66 to 74 BPM (beats per minute), which is the grand average for our studied population [Agelink et al., 2001].

Agents possessed some random parameters: their gender (male or female), their appearance (6 faces of different ethnic groups for each gender), their voice (2 voices for each gender) and the voice pitch. Those various parameters aimed at concealing the true independent variable. Had we chosen a unique appearance for all the agents, participants could have sought what was differentiating them. By individualizing agents we prevented participants to discover that ultimately we manipulated the HR feedback. To make agents look more alive, their eyes were sporadically blinking and their mouths were animated while the text-to-speech system was playing.

In order to elicit bodily reactions, we chose sentences for which a particular valence has been associated with, and, as such, that could span a wide range of emotions. Valence relates to the hedonic tone and varies from negative (e.g. sad) to positive (e.g. happy) emotions [Picard, 1995]. HR has a tendency to increase when one is experiencing extreme pleasantness, and to decrease when experiencing unpleasantness [Winton et al., 1984].

Our experiment was split in two parts (second within-subject factor). During the first session, called "disruptive" session (see Figure 8.1), participants had to rate each sentence they heard on a 7-point Likert scale according to valence they perceived (very unpleasant to very pleasant). Sentences came from newspapers. A valence (negative, neutral or positive) was randomly chosen every 2 sentences. Every 4 sentences, participants had to rate the social presence of the agent. Then a new randomly generated agent appeared, for a total of 20 agents, 10 for each "human"/"medium" condition.

As opposed to the first part, during the second part of the experiment, called "involving" session, sentences order was sequential (see Figure 8.2). Agents were in turns narrating a fairy tale. Participants did not have to rate each sentence's valence, instead they only rated the social presence of the agents. To match the length of the story, agents were shuffled every 6 sentences and there were 23 agents in total, 12 for the "human" condition, 11 for the "medium" condition.

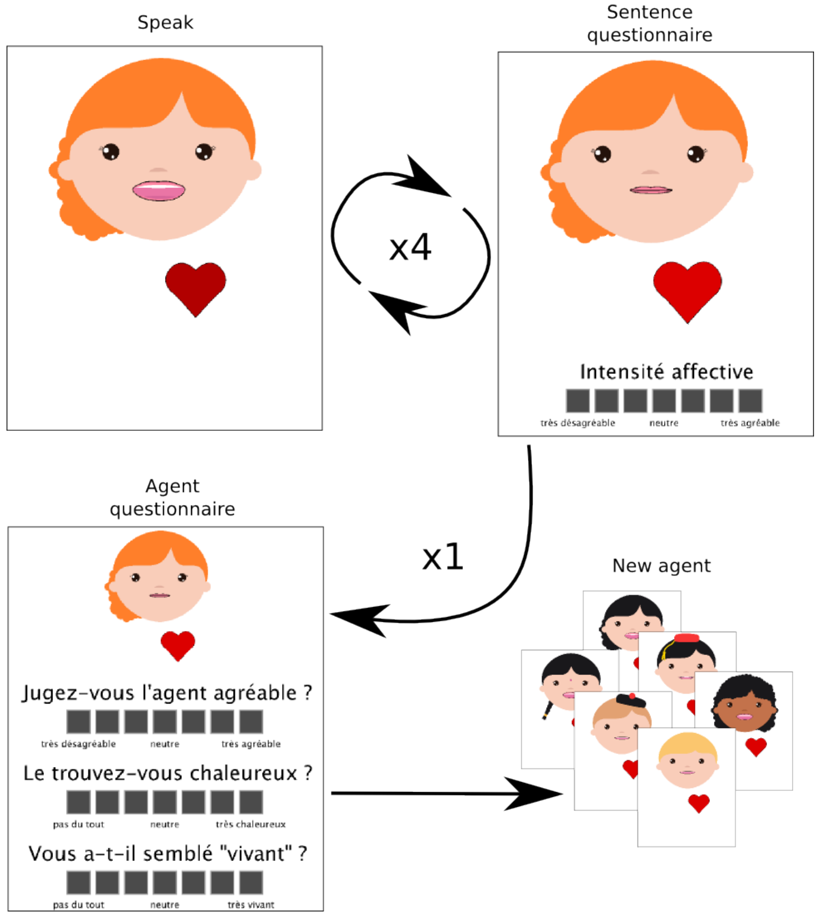

Figure 8.1 – Procedure during the "disruptive" session: participants rate the valence of each one of the sentences spoken by an agent. After 4 sentences, they rate agent's social presence (3 items). Then a new agent appears. 20 agents, average time per agent ≈ 62.2s.

Because of its distracting task and the nature of its sentences, the first part was more likely to disrupt human-agent connection; while the second part was more likely to involve participants. This let us test the influence of the relation between users and agents on the perception of HR feedback. We chose not to randomize sessions order because we estimated that putting the "disruptive" session last would have made the overall experiment too fatiguing for participants. A higher level of vigilance was necessary to sustain its distracting task and series of unrelated sentences. Participants' cognitive resources were probably higher at the beginning of the experiment.

We created a 2 (HR feedback: "human" *vs* "medium" condition) × 2 (nature of the task: "disruptive" *vs* "involving" session) within-subject experimental plan. Hence, our two hypothesis. **H1**: Hear rate feedback replicating users' physiology increases the social presence of agents. **H2**: This effect is more pronounced during an interaction involving more deeply agents.

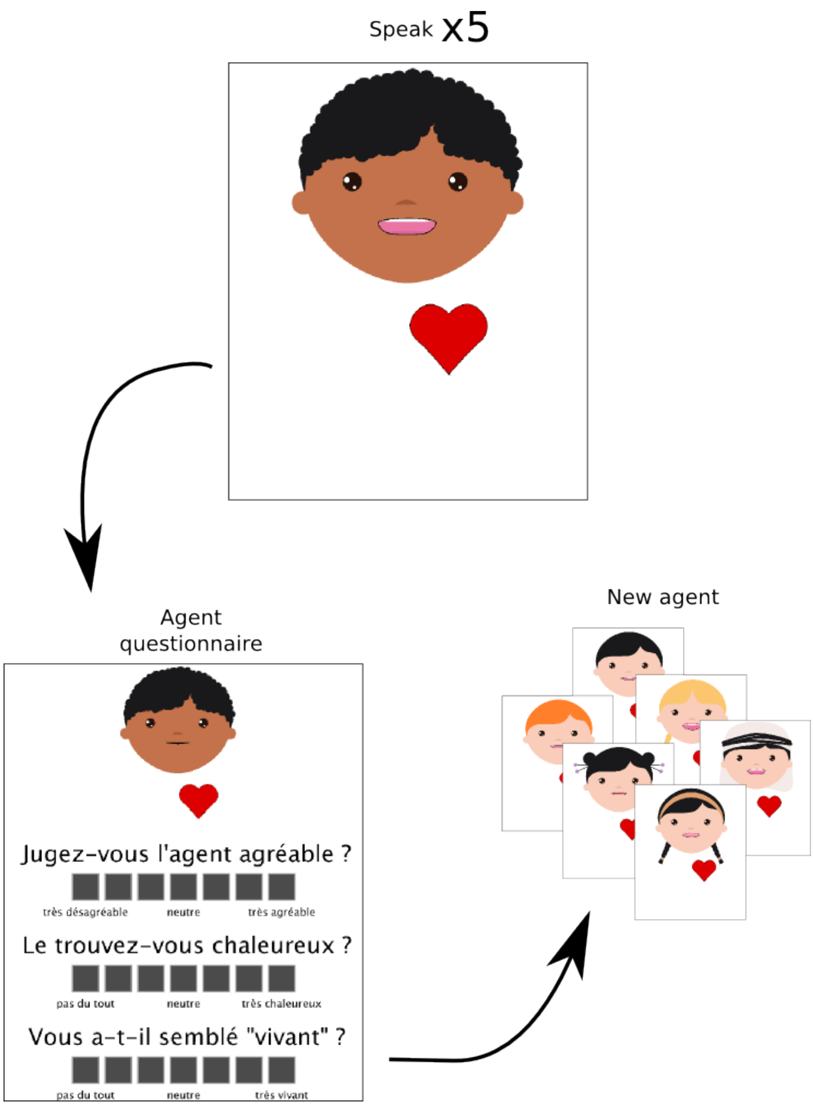

Figure 8.2 – Procedure during the "involving" session: participants rate agent's social presence after it recited all its sentences. Then a new agent appears, continuing the tale. 23 agents, average time per agent $\approx$ 46.6s.

8.2.1 Technical description

Most of the elements we describe in this section, hardware or software, come from open source movements, for which we are grateful. I would also like to thank the artist who made freely available the graphics on which agents are based[1]. All code and materials related to the study are freely available at https://github.com/jfrey-phd/2015_phycs_HR_code/.

[1] http://harridan.deviantart.com/

8.2.1.1 *Hardware*

We chose to use a BVP (blood volume pulse) sensor to measure HR, employing the open hardware Pulse Sensor[2] (see Figure 8.3 for a closeup). It assesses blood flow variations by emitting a light onto the skin and measuring back how fluctuates the intensity of the reflected light thanks to an ambient light photo sensor. Each heartbeat produces a characteristic signal. This technology is cheap and easy to implement. While it is less accurate than electrocardiography (ECG) recordings, we found the HR measures to be reliable enough for our purpose. Compared to ECG, BVP sensors are less intrusive and quicker to install – i.e. one sensor around a finger or on an earlobe instead of 2 or 3 electrodes on the chest. In addition, as far as general knowledge is concerned, BVP sensors are less likely to point out the exact nature of their measures. This "fuzziness" is important for our experimental protocol, as we want to be as close as possible to the real-life scenarios we foresee with devices such as the Kinect 2, where HR recordings will be transparent to users.

One of the creator of the Pulse Sensor is behind OpenBCI. Note that BVP and PPG are synonyms; we try here to avoid confusion with the remote sensing we saw.

The BVP sensor was connected to an Arduino Due[3] (see Figure 8.3). Arduino boards have become a well-established platform for electrical engineering. The Due model comes forward due to its 12 bits resolution for operating analog sensors. The program uploaded into the Arduino Due was feeding the serial port with BVP values every 2ms, thus achieving a 500Hz sampling rate.

Fortunately we deal with not expensive components; since it was a pet project at the time both the Arduino and the sensor were on me.

III.8

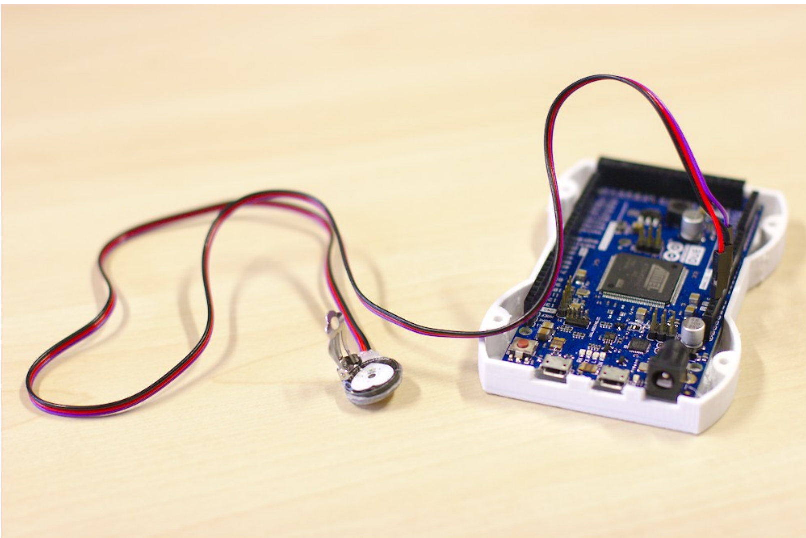

Figure 8.3 – BVP (blood volume pulse) sensor measuring heartbeats, connected to an Arduino Due.

Two computers were used. One, a 14 inches screen laptop (Alienware M14x), was dedicated to the participant and ran the human-agent

The second (and way older) computer was my personal, as well as the Fooloose (the remake) goody used as headphones. What would'nt I give for Science...

[2] http://pulsesensor.myshopify.com

[3] http://arduino.cc/

interaction. This computer was also plugged to the Arduino board to accommodate sensor's cable length. A second laptop (Lenovo ThinkPad T61p) was used by the experimenter to monitor the experiment and to detect heartbeats. Computers were connected through an ethernet cable (network latency was inferior to 1ms).

8.2.1.2 *Software and signal processing*

Computers were running Kubuntu 13.10 operating system. The software on the client side was programmed with Processing framework[4], version 2.2.1. Data acquired from the BVP sensor was streamed to the local network with ser2sock[5]. This serial port-to-TCP bridge software allowed us to reliably process and record data on our second computer. OpenViBE [Renard et al., 2010] version 0.18 was running on the experimenter's computer to process BVP, with custom python scripts to retrieve signals and send stimulations.

I did not implement LSL back then, synchronization would have been easier.

Within OpenViBE the BVP values were interpolated from 500 to 512Hz to ease computations. The script which received values from TCP was downsampling or oversampling packets' content to ensure synchronization and decrease the risk of distorted signals due to network or computing latency. A 3Hz low-pass filter was applied to the acquired data in order to eliminate artifacts. Then a derivative was computed. Since a heartbeat provokes a sudden variation of blood flow, a pulsation was detected when the signal exceeded a certain threshold. This threshold was set during installation: values too low could produce false positives due to remaining noise, and values too high could skip heartbeats. Eventually a message was sent. See figure 8.4 for an overview of the signal processing.

Using raw TCP communication and Ethernet cable, delays, for signal or stimulation where negligible (inferior to 1ms). Even the bottleneck induced by a 60 FPS of the main program couldn't impact our processing.

Once the main program received a pulse message, it computed the HR from the delay between two beats. As a failsafe measure against poor beat detection – e.g. noise due to head motion – the HR value computed within Processing could not vary by more than 10% and a min/max threshold was set (30/200). This value was passed over the engine handling the HR feedback during the "human" condition. We purposely created an indirection here – using BPM values in separate handlers instead of triggering a feedback pulse as soon as a heartbeat was detected – in order to suit our experimental protocol to devices that could only average HR over a longer time window (e.g. fitness HR monitor belts or remote PPG). It should be easier to replicate our results

[4] http://www.processing.org/
[5] https://github.com/nutechsoftware/ser2sock

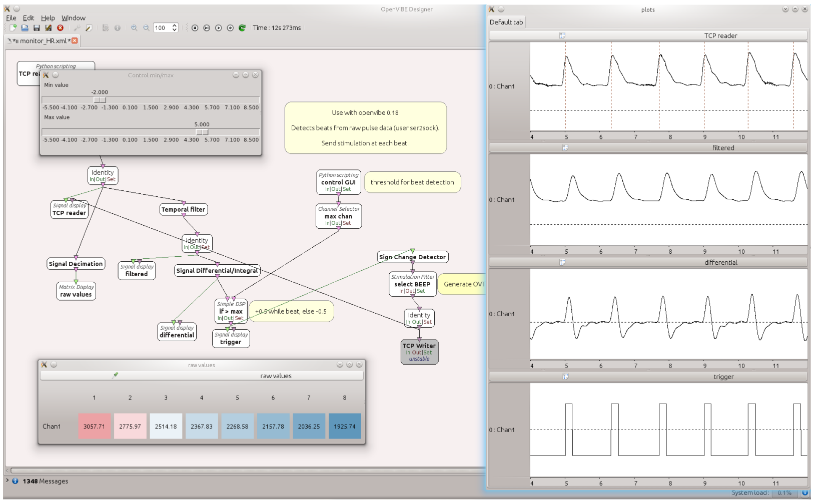

Figure 8.4 – Signal processing of the BVP sensor with OpenViBE. A low-pass filtered and a first-derivative are used to detect heartbeats.

without the need to synchronize precisely feedback pulses with actual heartbeats.

The TTS (text-to-speech) system comprised two applications. eSpeak[6] 1.47.11 was used to transform textual sentences into phonemes and MBROLA[7] 3.01h to synthesize phonemes and produce an actual voice. The TTS speed was controlled by eSpeak (120 word per minutes), as well as the pitch (between 65 and 85, values higher than the baseline of 50 to match the teenage appearance of the agents). The four voices (2 male and 2 female, "fr1" to "fr4") were provided by the MBROLA project. Sentences' valence did not influence speech synthesis.

8.2.2 Text corpuses

During the first part of the experiment (i.e. the "disruptive" session) sentences were gathered from archives of a french-speaking newspaper. These data were collated by [Bestgen et al., 2004]. Sentences were anonymized, e.g. names of personalities were replaced by generic first names. A panel of 10 judges evaluated their emotional valence on a 7-point Likert scale. The final scores were produced by averaging those 10 ratings. We split the sentences in three categories: unpleasant (scores between $[-3; -1[$, e.g. a suspect was arrested for murder), neutral (between $[-1; 1]$) and pleasant (between $]1; 3]$, e.g. the national sport team won a match) – see section 8.2.

The sentences of the second part (i.e. the "involving" session) come from the TestAccord Emotion database [Le Tallec et al., 2011]. This database originates from a fairy tale for children – see [Wright and McCarthy, 2008] for an example of storytelling as an incentive for

Never had time to run some machine learning against those valence scores to see how HR perform with narrative content. Data are here for those interested...

[6] http://espeak.sourceforge.net/
[7] http://tcts.fpms.ac.be/synthesis/mbrola.html

empathy. We did not utilize *per se* the associated valences (average of a 5-point Likert scale across 27 judges for each sentence), but as an indicator it did help us to ensure the wide variety of the carried emotions. For instance, deaths or bonding moments are described during the course of the tale.

It is worth noting that when the valence of these corpuses has been established, sentences were presented in their textual form, not through a TTS system.

8.2.3 Procedure

The overall experiment took approximately 50 minutes per participant. 10 French speaking participants took part in the experiment; 5 males, 5 females, mean age 30.3 (SD=8.2). The whole procedure comprised the following steps:

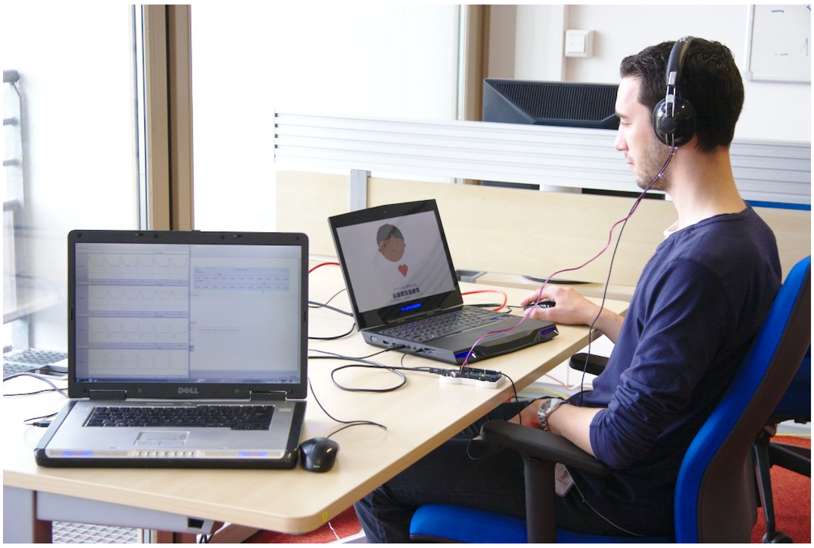

Figure 8.5 – Our experimental setup. A BVP sensor connects participant's earlobe to the first laptop, where the human-agent interaction takes place. Participant is wearing a headset to listen to the speech synthesis. A second laptop is used by the experimenter to monitor heartbeats detection.

1. Participants were given an informed consent and a demographic questionnaire. While they filled the forms, the equipment was set up. Then we explained to them the procedure of the experiment. We emphasized the importance of the distraction task (i.e. to rate sentences' valence) and explained to the participants that we were monitoring their physiological state, without further detail about the exact measures. ≈ 5 min.

2. The BVP sensor was placed on the earlobe opposite to the dominant hand, so as not to impede mouse movements. Right after,

the headset was positioned. We ensured that participants felt comfortable, in particular we checked that the headset wasn't putting pressure on the sensor. We started to acquire BVP data and adjusted the heartbeat detection. $\approx$ 2 min.

3. A training session took place. We started our program with an alternate scenario, adjusting the audio volume to participants' taste. Both parts of the experiment occurred, but with only two agents and with a dedicated set of sentences. This way participants were familiarized with the task and with the agents – i.e. with their general appearance and with the TTS system. During this overview, so as not to bias the experiment, "human" and "medium" conditions were replaced with a "slow" HR feedback (30 BPM) and a "fast" HR feedback (120 BPM). Once participants reported that they understood the procedure and were ready, we proceeded to the experiment. $\approx$ 5 min.

4. We ran the experiment, as previously described. First the "disruptive" session (80 sentences, 20 agents, $\approx$ 22 min), then the "involving" session (138 sentences, 23 agents, $\approx$ 17 min). We were monitoring the data acquired from the BPV sensor and silently adjusted the heartbeat detection through OpenViBE if needed – rarely, a big head movement could slightly move the sensor and modify signal amplitude. Figure 8.5 illustrates our setup. $\approx$ 40 min.

The newspapers' sentences being longer than the ones forming the fairy tale, agents on-screen time varied between both parts. Agents mean display time during the first part was 62.2s, during the second part it was 46.6s.

8.2.4 Measures

We computed a score of social presence for each agent, averaged from the 7-point Likert scales questionnaires presented to the participants before a new agent were generated. This methodology was validated with spoken dialogue systems by [Möller et al., 2007]. This score was composed of 3 items, consistent with ITU guidelines [ITU, 2003]. Translated to English, the items were: "Do you consider that the agent is pleasant?" ("very unpleasant" to "very pleasant"); "Do you think it is friendly?" ("not at all" to "very friendly"); "Did it seem 'alive'?" ("not at all" to "much alive").

8.2.5 Results

We compared agents' social presence scores between the "human" and the "medium" conditions for each part. Statistical analyses were performed with R 3.0.1. The different scores were comprised between 0 (negative) and 6 (positive), 3 corresponding to neutral.

A Wilcoxon Signed-rank test showed a significant difference ($p < 0.05$) during the "disruptive" session (means 3.29 *vs* 2.91) but no significant difference ($p = 0.77$) during the "involving" session (means: 3.30 *vs* 3.34). H1 is verified while H2 cannot be verified. Besides, when we analyzed further the data, we found no significant effect ($p = 0.27$) of the "human"/"medium" factor on the valence scores attributed to the sentences during the "disruptive" session (means: 3.06 *vs* 2.91).

H1: p<0.01 with "greater" hypothesis.

Participants' HRs were a little higher than expected during the experiment: mean $\approx$ 74.73 BPM (SD = 5.59); to be compared with the average 70 BPM set in the "medium" condition. We used Spearman's rank correlation test to check whenever this factor could have influenced the results obtained in the "disruptive" session. To do so, we compared participants' average HRs with the differences in social presence scores between "human" and "medium" conditions. There was not significant correlation ($p = 0.25$).

8.3 Discussion

In the course of the "disruptive" session our main hypothesis has been confirmed: users' engagement toward our HCI increased when agents provided feedback mirroring their physiological state. This result could not be explained by a preference for a certain pace of the HR feedback. For instance, even though their HRs were higher than average, participants did not prefer agents of the "human" condition because of faster heartbeats. Some of them did possess HRs lower than 70 BPM. The only other explanation lies in the difference of HR synchronization between "human" and "medium" conditions.

III.8

Beside agents' social presence, similarity-attraction effect may influence the general mood of participants, as they had a slight tendency to overrate sentences valence during "human" condition. It is interesting to note that while the increase in social presence scores is not huge (+13%), it shifts the items from slightly unpleasant to slightly pleasant.

Maybe the effect would have been greater in a less artificial situation. Indeed, despite our experimental protocol, participants reported afterwards that the TTS system was sometimes hard to comprehend, which bothered them on some occasions. It may have resulted in a task not involving enough for the participants to really "feel" the emotions carried by the sentences.

Several reasons could explain why the effect appeared only during our "disruptive" session. During the first session agents were displayed on a longer duration (+33%) because of the longer sentences used in the newspapers. The attraction toward a mirrored feedback could take time to occur. In addition, because the task was less disruptive in the second session, participants were more likely to focus their attention on the content (i.e. the narrative) instead of the interface (i.e. the feedback). This could explain why they were less sensible to ambient cues. Participants were less solicited during the "involving" session; we observed that between agents questionnaires they often removed their hands from the mouse, leaning back on the chair. Lastly, the "involving" session systematically occurred in second position. Maybe the occurrence of the similarity-attraction effect is correlated to the degree of users' vigilance.

As for participants' awareness of the real goal of the study, during informal discussions after the experiments, most of them confirmed that they had no knowledge about the kind of physiological trait the sensor was recording, and none of them realized that at some point they were exposed to their own HR. This increases the resemblance of our installation with a setup where HR sensing occurs covertly.

8.4 Conclusion

We demonstrated how displaying physiological signals close to users could impact positively social presence of embodied agents. This approach of "ambient" feedback is easier to set up and less prone to errors than feedback as explicit as facial expressions. It does not require prior knowledge about users nor complex computations. For practical reasons we limited our study to a virtual agent. We believe the similarity-attraction effect could be even more dramatic with *physically* embodied agents, namely robots. That said, other piece of hardware or components of an HCI could benefit from such approach. While its appearance is not anthropomorphic, the robotic lamp presented by [Gerlinghaus et al., 2012] behaves like a sentient being. Augmenting it with physiological feedback, moreover when correlated to users, is likely to increase its presence.

Further research is of course mandatory to confirm and analyze how the similarity-attraction applies to human-agent interaction and to physiological computing. The kind of feedback given to users need to be studied. Are both audio and visual cues necessary? Does the look of the measured physiological signal need to be obvious or could a heart pulse take the form of a blinking light? In human-human interaction such questions are more and more debated [Slovák et al., 2012, Walmink et al., 2014]. Obviously, one should check that a physiological feedback does not *diminish* user experience. [Lee et al., 2014] suggest it is not the case, but the comparison should be made again with human-agent interaction.

Various parameters in human-agent interaction need to be examined to shape the limits of the similarity-attraction effect: exposure time to agents, nature of the task, involvement of users, and so on. Especially, we suspect the relation between human and agent to be an important factor. Gaming settings are good opportunities to try collaboration or antagonism. Concerning users, some will perceive differently the physiological feedback. As a matter of fact, interoception – the awareness of internal body states – varies from person to person and affects how we feel toward others [Fukushima et al., 2011]. It will be beneficial to record finely users reactions, maybe by using the very same physiological sensors [Becker et al., 2005].

Finally, our findings should be replicated with other hardware. We used lightweight equipment to monitor HR, yet webcams could enable remote sensing in the near future (see appendix D). But with the spread of devices that sense users' physiological states, it is essential not to forgo ethics.

Measuring physiological signals such as HR enters the realm of privacy. Notably, physiological sensors can make accessible to others data unknown to self [Fairclough, 2014]. Even though among a certain population there is a trend toward the exposition of private data, if no agreement is provided it is difficult to avoid a violation of intimacy. Users may feel the urge to publish online the performances associated to their last run – including HR, as more and more products that monitor it for fitness' sake are sold – but experimenters and developers have to remain cautious.

Physiological sensors are becoming cheaper and smaller, and hardware manufacturers are increasingly interested in embedding them in their products. With sensors acceptance, smartwatches may tomorrow provide a wide range of continuous physiological data, along with remote sensing through cameras. If users' rights and privacy are protected, this could provide a wide range of areas for investigating and putting into practice the similarity-attraction effect. Heart rate, electrodermal activity, breathing, eye blinks: we "classify" events coming from the outside world and it influences our physiology. An agent that seamlessly reacts like us, based on the outputs we produce ourselves, could drive users' engagement.

If we do not formally test it afterwards, the similarity-attraction effect is the paradigm that supports most of our work in the following chapters. It is applied to human-human interaction in chapter 9 and to human-proxy[8] interaction in chapters 10 and 11. All those 3 works also take great care at giving explicitly a feedback corresponding to the recorded measures in order to give the upper hand to users regarding what data they share.

III.8

[8]Teegi and Tobe are not *exactly* robots. Not yet.

9

Using heart rate to foster social interactions

III.9

This chapter describes an application of physiological computing that uses heart rate monitoring as an incentive for social interactions. A traditional board game has been “augmented” through remote physiological sensing, using webcams. Projection helped to conceal the technological aspects from end users. We detail how players reacted – stressful situations could emerge when users are deprived from their own signals – and we give directions for game designers to integrate physiological sensors.

A shorter version of this work was published in [Frey, 2016b].

This projected originated in passionate discussions with Renaud Gervais and Jérémy Laviole; the experimental setup as well as the few game mechanics found in the conclusion are a mere ersatz of the idea we exchanged back then.

Besides ideas, both of them contributed the work presented below. I thank Renaud for having introduced me to Complot – what a funny and handy card game indeed – and Jérémy for the work he has done on the code that displays the heart rate feedback.

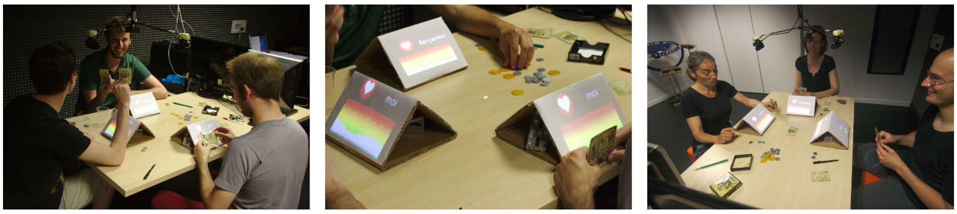

Figure 9.1 – A board game session that we augmented with remote physiological monitoring and projection.

9.1 Introduction

Through the rise of wearables – such as smartwatches – physiological sensors are among the technologies that are gaining increased attention. Heart rate belts, for example, are sold in many sports shops across the globe, heart monitoring being used by sportsmen to increase their performance [Tholander and Nylander, 2015] – heart rate activity is probably the measure that is the most commonly made outside of medical facilities.

However, maybe because of such "priming" effect, up until now the range of applications of physiological sensing among the general public has been limited. Despite the fact that already two decades ago physiological sensors were envisioned as a mean to shape new interactions through the so called "affective computing" [Picard, 1995], they are more often than not experienced within medical or sports settings only.

Now that physiological sensors are increasingly present around us, they are still missing an application that could spread their *use* outside people with special needs, in everyday life

Fortunately this situation can change. Lately, sensors have been investigated as a supplementary communication channel. They have been used to enhance existing interactions; implicitly with video games that adapt to users' state [Nijholt et al., 2009]; explicitly when the physiological activity is visualized for telepresence [Lee et al., 2014] or to enhance human-agent interaction, as we have seen in chapter 8.

While those latter examples mainly deal with human-computer interactions, physiological computing can also intercede between people that interact directly with each other. New usages emerged for social interplay [Walmink et al., 2014] or for mediating affect [Williams et al., 2015] – see [Chanel and Mühl, 2015] for a more thorough review regarding social interaction. Among those works, a study showed that heartbeats was a meaningful source of information for players and helped people to "connect" between each others [Slovák et al., 2012]. Authors mentioned how displaying heartbeat during a real life poker game suddenly arose players' interest.

"Human-human" interaction, such as card games, could well constitute the entry point for physiological computing to come out of its shell.

In return, board games constitute a good context to study complex social interactions in a close environment. Other studies used poker to test different aspects of physiological monitoring, for example if players could use the information to control their signals in [Yamabe et al., 2010] or how a “nervousness” indicator could affect gameplay in [Dang and André, 2010]. This is the kind of findings that could foster the demand for physiological sensors by the general public.

In this chapter, our main interest is to use a game as a dedicated use case of physiological monitoring’s influence over social interactions. Indeed, such information could help to create deeper interactions [Janssen et al., 2011], enriching social presence.

Even though previous works combined physiological monitoring and board games, they did not focus on how users reacted to this new feedback – nor did they consider the benefits for board games in general. For instance, [Slovák et al., 2012] investigated how people comprehend heart rate feedback in various situations, and the appearance of a gaming application among users was incidental. The biofeedback was studied for training in [Yamabe et al., 2010] as a way to help poker players gain control over themselves. Another combination of poker and physiological signals is sketched in [Dang and André, 2010], but heart activity only stands as an additional feature of a new human-computer interaction technique.

We seek to use a game as a dedicated use case of physiological monitoring’s influence over social interactions. We also want to explore how we could maximize user experience by integrating seamlessly the technology behind. As a matter of fact, the tabletop setup proposed in [Dang and André, 2010] requires additional gestures from users to perform actions as simple as hiding cards and in [Slovák et al., 2012] each player needed a laptop. Yamabe and al. [Yamabe et al., 2010] used a projector to display the heartbeat directly on the gaming table, but, as in each other previous work, users still had to wear sensors. It is possible for technology to be even less intrusive, and we solution both kinds of artifacts.

In the present work we used projection to display information, because it is less likely to disrupt the gaming experience than relying on screens to display heart rates; spatial augmented reality (SAR, introduced in [Raskar et al., 2001]) brings digital content to the physical world and facilitates the merge between computer science and existing board games (Figure 9.1). Furthermore, as opposed to sensors that are attached to the body – i.e. to the skin – we relied on a system that uses non-contact sensors to record heart rate by the mean of photoplethysmographiy (PPG). Detailed in appendix D, this system makes it possible to record heart rate by processing the subtle variations in skin colors while blood is flowing.

Through games and afferent social interactions, physiological sensors may finally reach end users in a casual way. Using SAR and PPG,

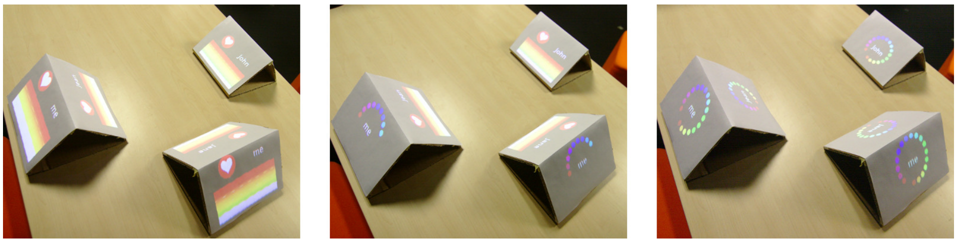

Figure 9.2 – Our experimental conditions regarding heart rate (HR) visualizations. *Left*: HR visible by all players. *Center*: HR visible by the others but not by self. *Right*: HR not visible.

the technology behind disappears in the eyes of the players, keeping the genuine "feel" of traditional board games. But then we need to ensure that such addition does not hinder user experience and, more importantly, that sharing an information that usually belongs to the realm of the self does not deter how users feel. Some may not like that others see "through" them, especially when heartbeats may relate to intimacy [Janssen et al., 2011]. In particular, this negative effect may be more likely to arise if the situation between players is not perceived as being "fair", e.g. if the biofeedback is seen by others and *not by the one being measured*.

Our first hypothesis is that the presence of a biofeedback equally shared between players – i.e. a heart rate visible by all – will improve game experience and social presence. Our second hypothesis is that an *asymmetrical* biofeedback – i.e. players see opponents' heart rate but not their own – will on the contrary cause more stressful situations and *deter* game experience.

In order to test these hypotheses, we used the versatility offered by SAR to create three different biofeedback conditions: heart rates visible by all, heart rates visible by others but not by self, no heart rates displayed.

The main highlights of this chapter are:

1. To describe an application that could facilitate the dissemination of physiological computing to the general public
2. To investigate how biofeedack influences user experience
3. To demonstrate how physiological monitoring could be integrated seamlessly to an environment that usually does not involve computers

9.2 Technical description of the system

One of the idea behind this study was to propose a "sit and play" setup for 3 persons, where players would not have to endure any supplementary equipment before experimenting physiological sensing.

Instead of choosing electrocardiography – that requires electrodes on the torso or on the wrists, or a pulse oximeter – that would clip to the finger, we turned to video analysis to record heart activity. The various modules that we developed in order to go from the webcam acquisition down to heart rate measures are described in appendix D. Our pipeline ends with heart rate feedback, which is projected in the gaming environment.

We used spatial augmented reality for mainly two reasons. It enables a seamless integration of the digital world into the physical world, the "disappearance" of computers resulting in a physiological monitoring that becomes one of the game mechanism of a traditional board game. SAR was also a way to multiply the displays without the need of adding physical screens – tablets, laptops, ... – to the players' surroundings. For instance, since we wanted to compare whether or not the visual feedback of oneself heart rate would change the social interaction, we just had to craft display stands with two sides, onto which we projected either a heart rate or an idling animation (see Figure 9.2), instead of using 6 separate screens.

The video projector, a Viewsonic PLED-W800 with a 1280 by 800 pixels native resolution, was positioned in a top-down orientation 1.5m above the table. The display surface was 1.2 by 0.75m. The projector was calibrated using ProCamCalib [Audet and Okutomi, 2009]. The positions of the stands were tracked beforehand using SURF algorithm [He et al., 2009] and then fixed for the rest of the experiment in order to save processing power for the heart rate measures. It was also an incentive for the participant not to move those display surfaces so much that they could see both faces of their stand and bias our experimental conditions. III.9

The visual feedback of the heart rate had two modalities. An icon shaped as a heart that was beating at the pace recorded by PPG, and beneath was a histogram plotting the BPM of the previous 20 seconds. We did not give a more explicit feedback – e.g. BPM values – so as not to distract too much the attention of the players from the main interaction, that is to say from the board game. Finally, the names of the players were displayed on the stands' sides facing others – "me" on the side facing them, helping to raise both their presence and their social awareness.

To obtain the desired visualization we used a framework developed in Processing[1] that could be easily grasped by game makers or artists [Laviole and Hachet, 2012].

All computations, for all 3 players, were done on a single computer, an Alienware Aurora R4 equipped with an Intel i7-3820 processor, 8GB of RAM and a GeForce GTX 660 Ti graphic card. The computer was running Kubuntu 14.04 operating system.

[1] https://processing.org/

9.3 User study

In this study we compared how users felt – regarding themselves and regarding the others – in three different conditions of heart rate feedback: heart rate visible by all ("HR all"), heart rate visible by the others but not by self ("HR others"), heart rate visible by none (control condition, "HR none") – see Figure 9.2. We used a within-subject experimental design. The conditions were set for all 3 players of a group at the same time, and each condition occurred once. The order of the conditions was counter balanced between groups following a latin square – hence we recruited 6 groups.

Beside the experiment that we describe below, this user study was also an opportunity to demonstrate how physiological sensors could be integrated to a traditional interaction, and to check that the technological choices we made – SAR and remote PPG – did not impact negatively the player experience.

9.3.1 Board game

As opposed to other studies, we did not choose poker as a card game to test the impact of physiological monitoring. Poker has too much of a background and, for some people, it may be associated to negative feelings due to the competitive spirit that surrounds it.

We chose instead a board game less known to people, more friendly and casual, "Coup" – edited by *Indie Boards and Cards*[2] – in its French version, "Complots". *Coup* possesses bluffing as one of the core elements of its gameplay. This is an incentive for players to use the physiological signals, since for the general public heart rate is strongly related to emotions. In *Coup* the main goal of the game is to "kill" the two characters of the other players, with various occasions to interact – block or counter attacks, steal money, and so on. Unless someone is "challenged" by an opponent, players never have to actually show their cards when they use the power associated to a character. These situations are most engaging for the players, whether they knew each others beforehand or not.

Except for two players, none of the participants were familiar with the game. *Coup* being quick to learn, every player was on par once the study started. *Coup* is also fast to play, with one game lasting about 5 minutes on average during our study, so we could have several games for each of our conditions.

9.3.2 Participants

18 participants took part in this study – 6 groups of 3 players, 5 females, 13 males, mean age 23.3 (SD: 6.9). Within each group, most of the participants knew each others beforehand. Half of them reported a

[2] http://www.indieboardsandcards.com/

III.9

previous use of physiological sensors, each time associated to sport – indoor bike or running – or to medical activities.

9.3.3 Protocol

The user study took place in a dedicated experimental room with homogeneous artificial light coming from the ceiling. The participants came by 3 to play the card game. Overall, a session lasted approximately 90 minutes and comprised the following steps:

1. The participants entered the room and sat around one end of a table, the other end being occupied by the experimenter. They read and signed a consent form and filled a demographic questionnaire. $\approx$ 5 minutes.
2. The rules of the card game were explained by the experimenter and a "warm up" game took place – two games if necessary. $\approx$ 15 minutes.
3. Once the players were confident they knew the rules, the SAR system was switched-on with fake signals so that the experimenter could explain how to read the feedback.
4. Then PPG signals were fed to the system and one of the 3 condition occurred. Participants played on their own. After about 10 minutes and a game ended, the experimenter interrupted the players.
5. While the SAR system was momentously switched-off, the participants filled 2 questionnaires related to emotion and social presence (see next section). $\approx$ 5 minutes.
6. Step 3 and 4 were repeated two more times, one for each other conditions of the user study.
7. After the completion of the 3 conditions, participants filled one last questionnaire to sense their overall feeling about the setup. $\approx$ 5 minutes.

Due to limited space the opposite end of the table was occupied by the experimenter, but the two control screens and the disposition of the room still gave some privacy to the participants while they were playing. Apart when – very occasionally – they asked about some specific rule, they were in fact so engaged in the game that they tended to forgot the presence of the experimenter. Generally there were two games over the course of one condition, sometimes 1 or 3 depending on the speed of the players.

9.3.4 Measures

Our main metric is composed by the two questionnaires given after each condition occurred – step 5 of previous section. Besides those measures, aimed at comparing our experimental conditions, we also wrote down participants' reactions while they were playing in order to gather more insights about what they experienced.

The first questionnaire is the self-assessment manikin (SAM) [Bradley and Lang, 1994]. Using various pictures of a cartoony manikin, respondents could indicate their emotions. There are three types of manikins – hence, 3 axis. One related to valence, one to arousal, and one to dominance. *Valence* relates to the hedonic tone and varies from negative to positive emotions (e.g. frustration *vs* pleasantness); *arousal* relates to bodily and mental activation and varies from "calm" to "excited" (e.g. satisfaction *vs* happiness) – see chapter 3. *Dominance* relates to the degree of control. Choosing between pictures instead of using words helps people to express feelings that could be difficult to externalize. We used a 9-points Likert scale version of the SAM and computed 3 scores for each condition.

The second questionnaire is the Social Presence in Gaming Questionnaire (SPGQ) [de Kort, 2007], that we translated to French. Originally developed for video games, its aim is to qualify social presence between players on three different axis: "empathy", "negative feelings" and "behavioral engagement" – "empathy" and "negative feelings" are both linked to psychological involvement. The SPGQ is rated on 5-points Likert scales and contains 21 items in total. A score related to "Empathy" is computed by averaging 7 of them (e.g. "When the others were happy, I was happy"); "negative feelings" is computed over 6 items (e.g. "I felt revengeful"); "behavioral engagement" over 8 items (e.g. "The others paid close attention to me"). As for SAM, 3 scores were computed for each HR condition out of the SPGQ.

III.9

9.3.5 Results

For each questionnaire and each axis, we used a Friedman test and post-hoc pairwise Wilcoxon signed-rank tests to compare our 3 heart rate feedback conditions ("HR none", "HR others", "HR all"). The results were adjusted for multiple comparisons with false discovery rate [Noble, 2009].

Regarding the SAM, there was no significant differences between the scores reported for valence, arousal and dominance between the 3 conditions.

There was no significant differences either in the SPGQ, although we found a tendency for the "negative feelings" score ($p \approx 0.1$) between "HR others" and "HR all" conditions. There was slightly more negative feelings reported in "HR others" compared to "HR all" condition – 1.19 *vs* 1.06 (SD: 0.87 *vs* 0.64), see Figure 9.3.

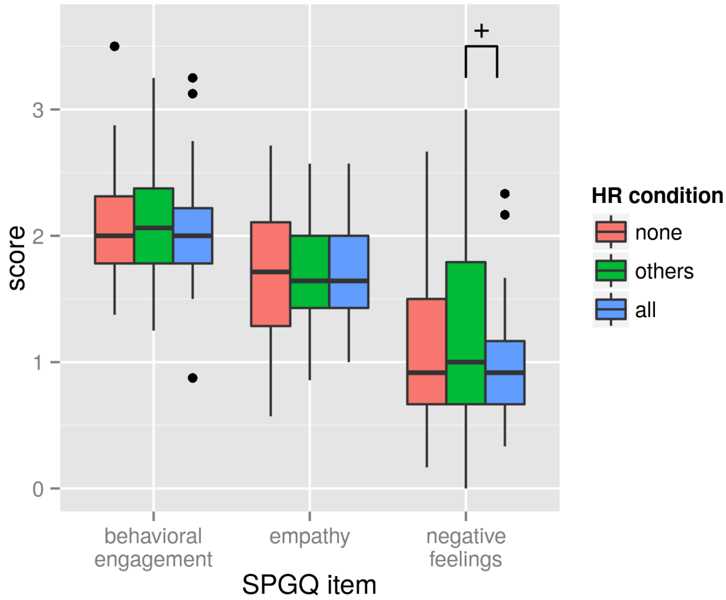

Figure 9.3 – Results of the SPGQ questionnaire – tendency is marked with a “+” sign.

Besides those two sets of questionnaires, at the end of the user study we also asked informally our participants their opinions about each experimental condition and about the technical aspects of the setup, using 5-points Likert scales ranging from “I did not like at all” (score: 0) to “I liked a lot” (score: 4).

The “HR all” condition was slightly favored over the “HR others” condition – 2.83 *vs* 2.78 (SD: 0.79 *vs* 1.06), and the condition with no HR feedback was ranked last – mean: 2.44 (SD: 0.86). On the technical aspects of the setup, participants praised the SAR display – 3.28 score, SD: 0.89, and were satisfied with the remote PPG heart rate measure – 2.89 score, SD: 0.96.

Concerning players’ comments to each other during the game, here is a selection of what participants said when they were referring out loud to the heart rate feedback:

- “Your rate is really high now, it’s because you’re upset!”
- “It’s stressful because I don’t see my heart!” (*HR others* condition)
- “Look at how his heart’s beating, he’s going to make a mistake I think…”
- “Damn, I got a huge spike, it’s because I won, I killed someone!”
- “You’re a bit fast, you look stressed!”
- “I see your plot and I see you bluffing.”

- "We're seeing your plot, don't go crazy!"
- "I don't own the game anymore, I cannot bluff…"
- "You saw how it went up suddenly?" / "Yes, it's because I was happy."
- "He said he liked her, and his heart increased…"
- "You're totally busted, I saw it, it increased!"

The proportion of sentences referring to emotions or decision processes, to events in-game or out-game, is representative of what we annotated during the different games.

As for *how* the physiological feedback was utilized, we observed two different kinds of players, roughly in equal proportions. First, the players that did not use explicitly the heart rate display during the game. Even when they liked to see it, they did not use this information while interacting with other players. During informal inquiries, those players reported that the provided heart rate display was hard to interpret. The second profile is among players that did use the feedback; participants that knew beforehand the game and the rules were immediately attracted by the heart rate feedback and remained the most enthusiast throughout the game.

9.4 Discussion

Between the two HR conditions, players had a tendency to report more negative feelings in the SPGQ questionnaire when others could see their heart rate but when themselves could not. From direct observation, it seems that players used this asymmetry to "tease" themselves, giving false or exaggerated feedback to the one that could not see by herself or himself the real heart rate. Sharing the information evenly should prevent social stress – unless of course a game designer *wishes* to create a very competitive gameplay. These findings only partially go along our hypotheses since players still preferred an asymmetrical feedback rather than no HR feedback at all when they were asked to rank explicitly the experimental conditions at the end of the study.

Concerning the utilization of the heart rate display, we did not give absolute values to avoid a too intrusive feedback, and we used the same scale for all players in the "history" plot to adapt to all metabolisms. These choices may have prevented some players to comprehend well the information. Hence, other visualizations should be investigated as well as other feedback modalities, such as sounds.

Many players also reported being too much focused on the actual game and on their adversaries' strategies to gather yet another clue. In this case the novelty of the game may explain why players' attention was centered elsewhere. On the contrary, players that had already played the game and knew the rules before the user study were immediately attracted by the heart rate feedback; they were the most enthusiast thorough the game. Physiological signals, for instance, could be used to add another layer to an existing game, especially for experts.

Among the players that did use the heart rate, most of them associated the heart rate to emotions, even though in reality this is *one* among many different traits that could be inferred from heart rate [Kivikangas et al., 2010]. This is easily explained by the common knowledge surrounding heart activity and by our cultural references. Still, other participants did not hesitate to interpret various events with regard to the heart rate feedback – players who spoke first during the exchanges were mostly looking at others' signals in order to bother playfully opponents.

Maybe the loss and gains would have been the same without, but physiological sensors altered the gaming experience, improving the richness of the relationships.

9.5 Conclusion

We presented a framework that combined remote heart rate monitoring and projection in order to bring anew an existing board game, without modifying the latter. We showed how the scope of physiological signals can be extended to reach a new population, beyond medical and sport contexts. Finally, we reported how it enriched social interactions.

During our study we sensed how a discrepancy between what is recorded by the system and what is showed to the user could lead to stress, when only the other players could see one heart rate. Overall our observations seem to indicate that the presence of a physiological feedback improved the richness of the relationships between the players, even though more thorough examinations are needed before we could draw solid conclusions about how the gaming experience is altered.

Game designers may use PPG or any other technology to integrate heart rate measurements to the gameplay. In the end, they have the possibility to develop a new game system. For example, in a card game a special picture could trigger the masking of one's heart rate for a fixed duration. In this case it would be up to the opponents to decide if the player wants to hide something… or if it is a strategy to make it believe so. Another card, or a gesture, could switch feedback between players. A common shared space, thanks to SAR, could also favor the collaboration or competition over physiological states – e.g. synchronize signals. A "game master" could use biofeedback as an input in a role-playing game. The possibilities are limitless, and may rapidly be explored by the players themselves. As matter of fact, not all the modifications observed by *our* players were related to actual physiological changes, especially when it was sudden. Some players reported rightfully that when the webcam was "seeing the hair", the values changed. Attempting to deceive opponents is one way to play with physiological signals.

The visual feedback we provided through projection was relying on simple 3D models. While this kind of projection was already sufficient to obtain a game room where anybody, at anytime, could take a seat and start to play, others may wish to venture further into spatial augmented reality. Projection could be mapped to detailed objects, small avatars that could be tracked in real time for example. Tangible interaction could be a powerful incentive for novices to grasp physiological sensors [Frey et al., 2014a].

In the last part of this thesis, we purposefully venture into tangible avatars that act as proxies for inner physiological activities and mental states. We study how this form factor help to relieve the fears and misunderstandings that surround a technology such as EEG, prompting users to engage in self-investigations (Teegi, chapter 10). Then we investigate how such augmented "puppets" help to better know ourselves – cognitive processes that go beyond emotions – and how they favour the emergence of new interactions between users.

PART IV

PHYSIOLOGY TO MEDIATE ONESELF

Over the course of various projects we saw how physiological signals could be used to asses mental states and how they could improve the quality of existing interactions – either between human and computers or between peers.

These different pieces wandered and are now about to be brought together.

This part describes two applications were physiological signals go back to a body. A body, not the one signals were originating from at start, but a physical persona that will act as a proxy between users and their signals. Two examples of embodiment that are a tangible representation of users' inner states.

Teegi lets novices discover about their brain activity; Tobe is the personification of a platform that enables users to shape and share their physiology.

Such tangible artifacts help to investigate the concepts underneath physiological signals, making literally easier to grasp one's state. Teegi and Tobe, two physical personas, two siblings soon to be reunited to properly mediate oneself using physiological computing, acting like a social prosthesis and facilitating introspection.

Those tangible avatars would not be if it was not for Renaud Gervais. Renaud brought his vision, knowledge and skills as much as I did. Both our theses share the pages that follow, however we approached these works with two different angles. Only half of the story lies in the present text. You may not wish to read two manuscripts in a row, but you should at least check out the backstory by jumping to the appendices, section Credits.6. And if you wonder how tangible augmented objects could leak information between the screen and your surroundings, shattering forever the boundaries between digital and physical worlds, well, it's never too late to switch.

10

Teegi: Tangible EEG Interface

Teegi is a Tangible ElectroEncephaloGraphy (EEG) Interface that enables novice users to get to know more about something as complex as brain signals, in an easy, engaging and informative way. To this end, we have designed a new system based on a unique combination of spatial augmented reality, tangible interaction and real-time neurotechnologies. With Teegi, a user can visualize and analyze his or her own brain activity in real time, on a tangible character that can be easily manipulated, and with which it is possible to interact. An exploratory study has shown that interacting with Teegi seems to be easy, motivating, reliable and informative. Overall, this suggests that Teegi is a promising and relevant training and mediation tool for the general public.

This chapter is a (slightly) extended version of the work published in [Frey et al., 2014a]. Beside the authors credited in this paper, I would like to thank Jérémy Laviole – who helped to implement the "on-the-go" version of Teegi that we brought outside the lab – and Maxime Duluc, who's late work concerned an instrumented version of Teegi (codename "Teegi disco", coming soon in a store near you!).

10.1 Introduction

Electroencephalography (EEG) measures the brain activity of participants under the form of electrical currents, through use of a set of electrodes connected to amplifiers and placed on the scalp [Niedermeyer

Figure 10.1 – Teegi (Tangible EEG Interface) is a friendly interactive character that users can manipulate to observe and analyze their own brain activity in real time.

and da Silva, 2005] – appendix C describes how such device could be made.

Although the scope of this thesis is limited to mental states assessment and Brain-Computer Interfaces (BCI, see part II), EEG is first and foremost widely used in medicine for diagnostic purposes, e.g. for the diagnosis of sleep disorders or epilepsy [Niedermeyer and da Silva, 2005]. While these emerging technologies are becoming increasingly more popular, they feed into fears and dreams in the general public, where many fantasies are linked to a misunderstanding of their strengths and weaknesses. *No, it is not possible to read thoughts!* But what can be done exactly? Our motivation is to provide a tool that allows one to better learn how EEG works, and to better understand the kinds of brain activity that can be detected in EEG signals. Beyond the knowledge of the brain that a user can acquire, we believe that a dedicated tool may help demystify BCI, and consequently, it may favor the development of such a promising field.

We followed a multidisciplinary approach, combining Human-Computer Interaction (Spatial Augmented Reality, Tangible User Interfaces), Neurotechnologies (EEG, brain signal processing) and Psychology/Human sciences (Human Learning and Representations, Scientific Mediation) to design an interactive multimedia system that enables novice users to get to know more about something as complex as EEG signals and the brain, in an easy, engaging and informative way. Our final goal is to enhance learning efficiency and knowledge acquisition by letting users actively and individually manipulate and investigate the concept to be learned [Vosniadou et al., 2001], i.e. EEG signals.

This gave birth to Teegi (Tangible EEG Interface), a physical character that users can manipulate in a natural way to observe and analyze their own brain activity projected in real time on the character's head (see Figure 10.1. Beyond the technical description of Teegi, this paper depicts an exploratory study we conducted, which provides an experimental basis for discussions and future works. Our major contribution for this paper is the design of the first system to make EEG signals and brain activity easily accessible, interactive and understandable. This work is based on theoretical foundations, technical developments, and preliminary investigations.

10.2 Neuroimaging and EEG

EEG signals are small electrical currents (in the μV range) that can be measured on the surface of the scalp [Niedermeyer and da Silva, 2005]. They reflect the synchronous activity of millions of neurons from the brain cortex (i.e. the outer layer of the brain).

The currently available tools used to visualize and analyze such signals are tailored for experts with a deep understanding of the brain, EEG principles and EEG signal processing [Niedermeyer and da Silva, 2005]. Figure 10.2] (left and center) shows some typical visualizations of EEG signals used by experts, i.e. EEG signal traces and a 2D topographic map. More complex visualizations have been proposed, such as 3D topographic maps (Figure 10.2, right), but they require many mouse inputs to be observed from all angles, which make them inconvenient to use in practice.

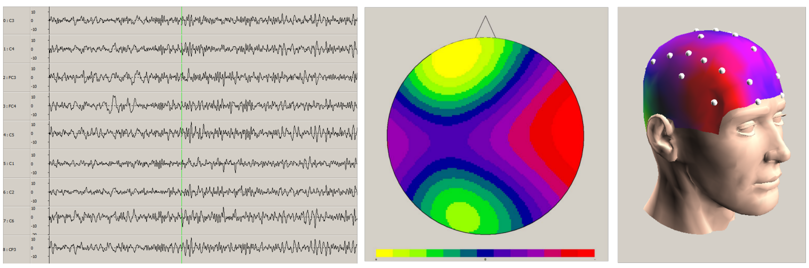

Figure 10.2 – (Left) A trace of EEG signals collected from multiple sensors. 2D (center) and 3D (right) topographic maps. (Screenshots from OpenViBE [Renard et al., 2010]). The first two views are traditionally used by experts.

Although EEG visualizations are intended for experts only, the general public is often compelled by how the brain works and how its activity is measured. Anyone wondering about brain injuries, epilepsy, sleep or learning disorders, aging, etc. may want to seek further knowledge about how the brain works. Currently, the public is increasingly exposed to neurotechnologies due to the availability of consumer grade EEG devices, such as the Emotiv EPOC or the Neurosky MindWave. Consequently, it has become necessary to design tools and user interfaces which will allow the general public to visualize, understand and interact with EEG signals. For instance, Mullen et al. proposed a software solution to process EEG signals collected using wearable EEG devices and visualize them in 3D [Mullen et al., 2013]. This software enables the user to estimate brain activity sources and connectivity but is still mainly designed for brain signal and neuroscience experts, and not so public-friendly. Another recent work, more suited to lay persons, is the "Portable Brain Scanner" [Stopczynski et al., 2014]. This system makes use of a consumer-grade EEG device (the Emotiv EPOC) and a smartphone to provide a cheap and portable solution enabling anyone to visualize the sources of their brain activity on their smartphone in 3D. Another more attractive work, which is the most closely related to our Teegi system, is the "Mind-Mirror" system [Mercier-ganady et al., 2014]. This system combines Augmented Reality (AR), 3D Visualization, and EEG to enable users to visualize their own brain activity in real time superimposed to their own head, thanks to a semi-transparent mirror-based AR setup.

This short review of the existing literature about making EEG accessible to the general public revealed that this is still a vastly unexplored area. Moreover, these solutions do not take into account any representation that the general public may have regarding the brain and EEG signals – many lay people do not even know what EEG signals are – in order to provide suitable visualizations and interaction devices to better understand these concepts. Some rare studies have indicated that misconceptions about brain functions prevail in general public [Dekker et al., 2012, Herculano-Houzel, 2002, Simons and Chabris, 2011]. These works stress the importance of popular scientific communication and indicate that communication efforts should be focused on increasing public awareness. It is important to note that the existing works mentioned above are mostly centered on visualization, with little or no interaction possibilities to manipulate and understand the EEG signals in real time and in a friendly way. This further deters the general public from understanding brain activity [Vosniadou, 1992]. Therefore, with the aim to enhance general public awareness, our work associates technical innovation and user-centered design.

10.3 Introducing Teegi

10.3.1 Founding principles

Design choices were made according to pedagogical principles. It has long been recognized that learner-centered education is much more effective than transmission-based education, even in informal situations [Wellington, 1990]. According to the constructivist paradigm, people create unique personal meanings by reflecting on interactive learning experiences. Therefore, people/learners should investigate and manipulate in order to become conscious of complex phenomena, change their misconceptions and construct scientific knowledge [Vosniadou et al., 2001]. In association, meaningful models play an important role in this type of learning processes [Fleck and Simon, 2013]. This motivated the design of an anthropomorphic interface that can be freely manipulated.

Our user-centered interactive media uses Spatial Augmented Reality (SAR) and tangible interaction. SAR, introduced by Raskar et al. [Raskar et al., 2001], adds dynamic graphics to real-world surfaces by the means of projected light. Many systems were designed using projectors to add "painted" surface [Raskar et al., 2001] or to give the illusion of virtual elements actually being there [Wilson et al., 2012a, Benko et al., 2012]. A related approach is Tangible User Interface (TUI). TUI is concerned with providing tangible (i.e. physical) representations to digital information and *controls* [Ishii and Ullmer, 1997, Shaer, 2009]. One of the strengths of tangibles is their situatedness: the interaction takes place in a real-world environment that often hides most of the technological aspects to expose physical interaction components only. They are particularly well suited for mediation purposes as they tend to be more inviting compared to mouse-screen based interfaces [Horn et al., 2009].

IV.10

SAR and TUIs are often found together [Underkoffler and Ishii, 1999, Piper et al., 2002]. They are very complementary in that they both take place in the real world, in a common canvas. The tangible interaction allows for a hands-on approach by offering different input affordances (as well as physical constraints) to the user while the SAR technology allows for a flexible and situated way to give feedback. SAR can also be used as an affordable way to embed dynamic graphics on a physical surface that would otherwise require curved displays [Brockmeyer et al., 2013] or rear projection [Benko et al., 2008] – SAR helped for example to integrate seamlessly 6 different viewports to the surrounding space of game board players in chapter 9.

There are examples of systems that use either tangible or AR principles to interact or review physiological data. Hinckley et al. [Hinckley et al., 1994] designed a system which used tangible props to do neurosurgical planning. A small tangible head was used in conjunction with a plastic plane to select the cutting planes to be visualized on a screen. Also mentioned above, the “Mind-Mirror” [Mercier-ganady et al., 2014] is the work closest to Teegi. However, with Teegi, the data is not co-localized with the data source. It provides flexibility and easier visualization as the users can change viewpoints by tangible interactions instead of rotating their head while keeping their eyes on the mirror. This “out-of-body” visualization also enables collaboration where multiple users can explore the data.

10.3.2 General description

Teegi is a tangible interface that enables users to visualize and analyze a representation of their own brain activity recorded via an EEG system in real time and displayed on a physical character. After some processing of the raw signals, a dedicated visualization is projected directly on top of the character. This character is tracked, which allows us to co-locate the projection with the character’s head, at any time. Hence, the user can easily visualize a realistic modeling of the EEG signals in any part of the scalp by manipulating the character, while maintaining a good spatial topology of the observed data. Teegi was purposely given a child-like appearance, as well as animated eyes (also projected) that blink at the same time as the users do, in order to breathe life to the character and enhance attractiveness. Indeed blinking can be easily detected in electrodes neighboring the eyes.

Three different filters can be applied to the raw data (see the technical section for details) enabling users to investigate influences of motor motions, visual activities or meditation, on their brain activity in real time. To remain consistent with the tangible philosophy of this project, we decided to control the filters by way of small tangible characters (mini-Teegis) that can be moved on a “filter area”, which is highlighted on the table by a projected halo (see Figure 10.4). For example, if a user wants to apply a filter that will allow her to better see what happens when moving her hand, she just needs to take the dedicated mini-Teegi, i.e. the one with the colored hands, and to move it to the filter area. Then, by moving her right hand, she should see changes in EEG amplitude on the left hemisphere of Teegi’s head, as illustrated in Figure 10.3. The manipulation of Teegi requires a motor activity. Therefore, when the motor filter is on, manipulating Teegi will obviously lead to observable changes in brain activity.

Later on during this chapter, we present an exploratory study we conducted to obtain feedback about the main features of Teegi. However, Teegi is not limited to these first features. In the next section, we describe additional interaction metaphors we have explored, and that may benefit more advanced users. These advanced features were not evaluated during the study.

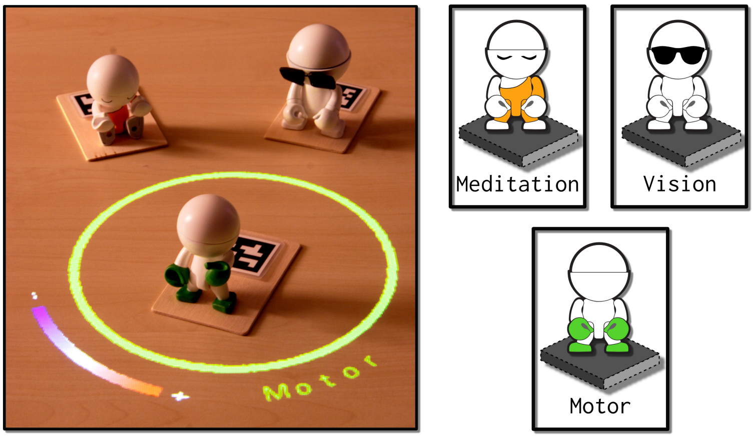

Figure 10.3 – Three mini-Teegis can be used to apply high-level EEG filters to highlight brain processes associated to *Motor*, *Vision* and *Meditation* activities. To do so, the user simply needs to move the desired mini-Teegi into a specific zone projected on the table (green circle).

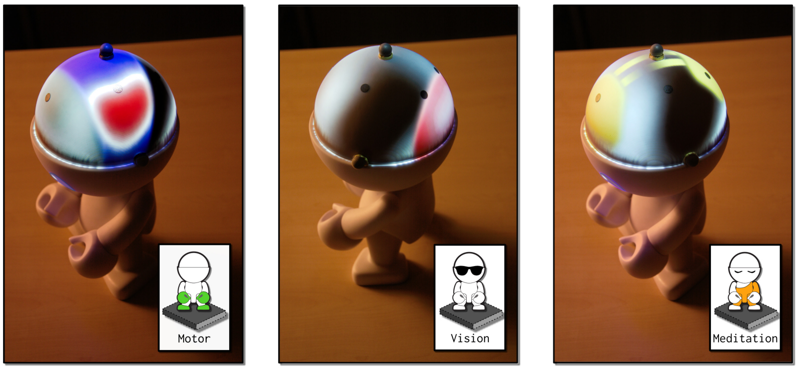

Figure 10.4 – Examples of the displayed visualizations on Teegi for each of the provided filters. Once a filter is active, the brain area corresponding to the selected and processed activity is highlighted in colors while the remaining EEG signals are displayed in grayscale.

IV.10

10.3.3 Advanced features

Visualizing the raw signal recorded on each electrode of the EEG is not very informative for the general public. However, this can be instructive for students who are learning EEG signal processing and analysis. In our approach, we can display on the table these raw data, as shown in Figure 10.5 (left). This creates a visual link between what is recorded with the EEG system, and the visualization that is provided on Teegi's head. This is possible because we know the rough position of the user, and the exact position of Teegi. When applying a filter, as described in the previous section, the user can see the effect of his or her action on the

VRPN was used with Teegi to tranmit signals from OpenViBE to vvvv; it was a time before we had LSL implementations running..

signal (see Figure 10.5, right). Compared to a standard approach where everything takes place on a screen, we believe that such a spatial and tangible approach might ease the understanding of the filters' effect.

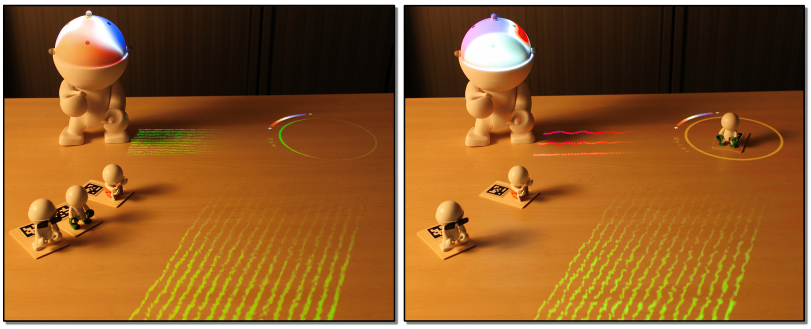

Figure 10.5 – *Left*: the raw EEG readings are displayed going from the user to the filter area and then rerouted towards Teegi. *Right*: When a mini-Teegi (i.e. a filter) is active, the corresponding filtered signals are displayed between the filter area and Teegi instead.

Another dimension we explored is the use of tangible actions to control some parameters of the EEG signal processing. As an example, we have implemented a technique where the user can control the amplitude of the visualization color map by moving a tangible object on the table (figure 10.6). This could be useful to reveal tiny fluctuations of EEG signals. With such interaction techniques, the whole analysis could be conducted without the use of a screen or a mouse, which remains consistent with the tangible philosophy of the project.

Eventually, the idea would be to have the whole visual programming environment of OpenViBE on the table with tangible tokens.

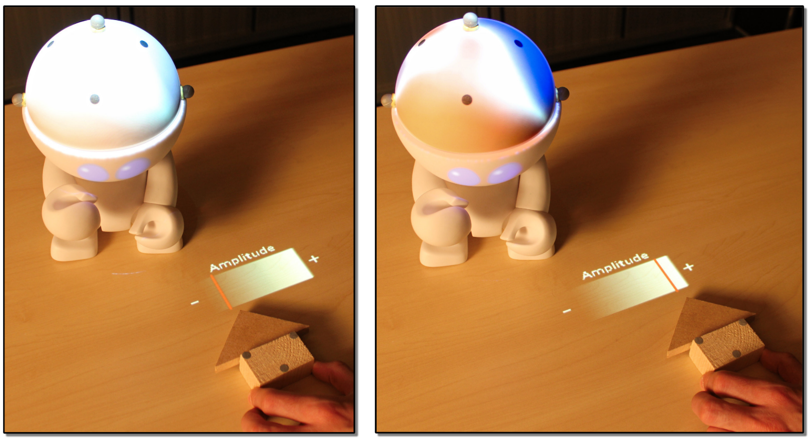

Figure 10.6 – A moving tangible cursor is controlling the amplitude of the visualization color map.

Finally, we developed a solution that highlights the relationship between EEG signals and localized cortical sources, that is where the signals come from *inside* the brain. Using sLORETA inverse modeling [Pascual-Marqui, 2002] and Brainstorm to compute the kernel matrix [Tadel et al., 2011], we obtained a model of the cortex containing 2002 voxels linked to the 32 EEG electrodes we used. We can then project in real time the activity which arises from the outer regions of the cortex on an object representing the brain, alongside with Teegi (Figure 10.7). Since both Teegi and the brain proxy are tracked, it becomes possible to manipulate two synchronized representations of the same brain activity (the source at the surface of the brain and the measures on the scalp). This opens way to mediation activities that are more advanced all the while keeping the simplicity and ease of use brought forth by using SAR and tangible interaction.

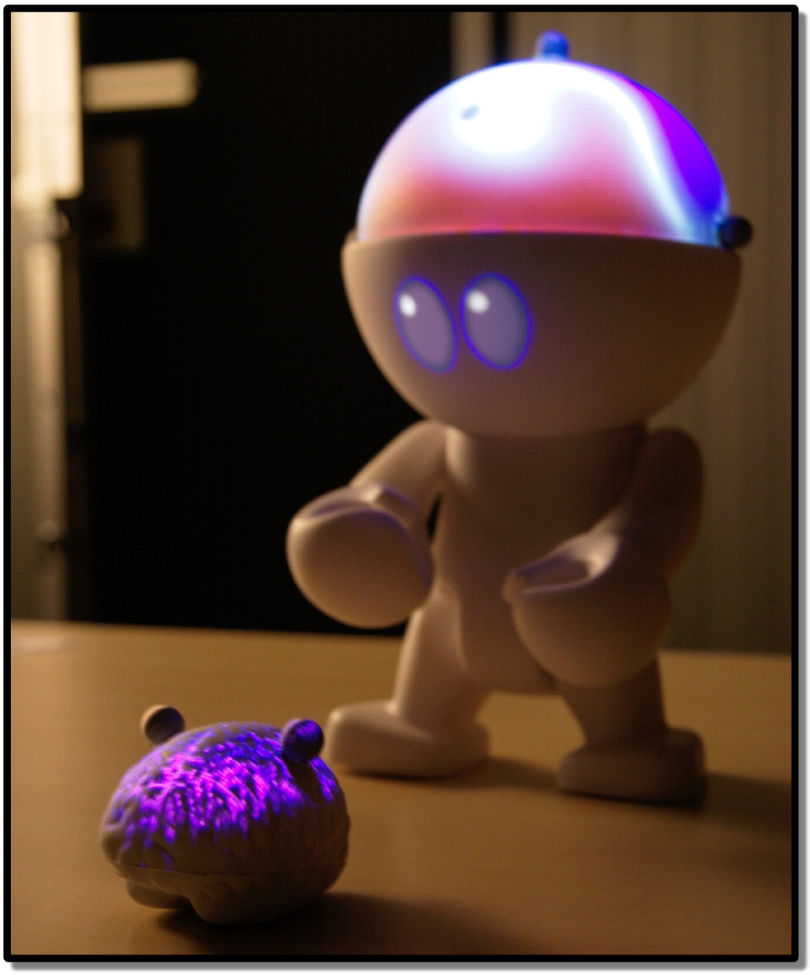

Figure 10.7 – Using an inverse model, the cortical activity and EEG measures are presented together to users.

10.4 Technical description

10.4.1 EEG

We designed different EEG signal processing pipelines that each create a specific visualization tailored to identify specific elements in the signal. The details of these pipelines are transparent to the user. Each pipeline corresponds to a mini-Teegi filter. In particular, we set up the following EEG signal processing pipelines:

1. Wide-band EEG activity: EEG signals were band-pass filtered between 3Hz and 26Hz, in order to filter DC drift and part of the artifacts (e.g. facial muscle activity [Fatourechi et al., 2007]) that may pollute them. Their power is then computed before being displayed. This corresponds to unspecific brain signals, hence they were labelled as "raw" signals.

2. Sensorimotor activity: EEG signals were first band-pass filtered in the β band (16-24Hz), a brain rhythm highly involved in sensorimotor tasks [Pfurtscheller and Lopes da Silva, 1999]. Then, they were spatially filtered, i.e. the signals from several neighboring EEG sensors were combined in order to enhance the signal of interest. In particular, we used and displayed Laplacian spatial filters around electrodes C3, C4 and Cz. This enabled the users to visualize EEG activity changes due to movements of the left hand, right hand and feet. Indeed, it is known that the power of EEG signals in the β rhythm decreases in electrodes C3/Cz/C4 during right hand/feet/left hand movements respectively, and increases just after the end of this movement [Pfurtscheller and Lopes da Silva, 1999].

3. Visual activity: EEG signals were band-pass filtered in the α band (8-12Hz), then only electrodes P3, Pz, P4, PO3, POz, PO4, O1, Oz and O2 (located on the back of the head, above the neck) were selected and displayed. These electrodes are indeed located over the visual cortex of the brain, i.e. the brain area in charge of visual information processing. The amplitude of the α rhythm is actually known to increase while the user is closing his/her eyes and is thus not processing any visual information [Niedermeyer and da Silva, 2005]. To ensure that the user could perceive this increase after he/she reopened his/her eyes, the visualization was delayed by 0.5s.

4. Meditation: on a more exploratory note, we used the synchronization between the signals from the anterior and posterior cortex (AFz/Pz), which was measured in a 7-28Hz band with instantaneous phase locking value [Lachaux et al., 2000]. There are different outcomes (increase/decrease in synchronization)

depending on meditation type. Mindfulness and body focus practices decrease the synchronization while transcendental practice increases it [Lehmann et al., 2012].

EEG signals were acquired with a 32-channels EEG device (made of two g.tec g.USBAmp EEG amplifiers). This professional-grade system ensured that our prototype had a good signal-to-noise ratio and accurate electrode location, avoiding unneeded uncertainties. Signals were processed in real time using OpenViBE [Renard et al., 2010]. For pipelines 1 to 3, the displayed colors correspond to signal power strength; for pipeline 4 they correspond to the degree of synchronization.

10.4.2 Spatial Augmented Reality

In order to create an augmented character, we have designed a tabletop augmentation setup (see Figure 10.8). Teegi itself is a 25cm high Trexi DIY toy. The mini-Teegis are also 10cm high Trexis. The main program handling the whole installation was created with vvvv[1].

The primary projected content (Teegi augmentation and GUI display) is handled with a single wide lens projector ProjectionDesign F20SX of resolution 1024x768 located over the table in a top-down orientation. The tracking of Teegi is achieved with an OptiTrack V120:Trio. It runs at 120 FPS with an overall latency of 8.3ms and a precision of 0.8mm. The OptiTrack is located in the same configuration as the main projector and both devices are calibrated together manually. The tracking data is sent to vvvv using OptiTrack's NatNet protocol. Teegi's eyes are projected using a second projector (Vivitek Qumi Q2) that is located on the side of the table.

The filter selection is done using a Sony PSEye web camera pointed at the position of the program selection GUI. Each mini-teegi representing a filter has a fiducial marker attached to it. The library ARToolkitPlus [Wagner and Schmalstieg, 2007] is used to detect which marker is currently selected.

The OpenViBE software that processes EEG also generates a grayscale texture of the scalp signals. This texture is then exported to a local shared samba folder which is then fetched and remapped to an appropriate color scale in vvvv before being mapped to Teegi's head. In addition, the raw EEG signals are sent to vvvv over VRPN for display purposes (see Figure 10.5).

10.5 Explorative study

10.5.1 Protocol

We conducted an exploratory study where participants had to manipulate Teegi following a given scenario. The objectives of this study were

[1] http://vvvv.org

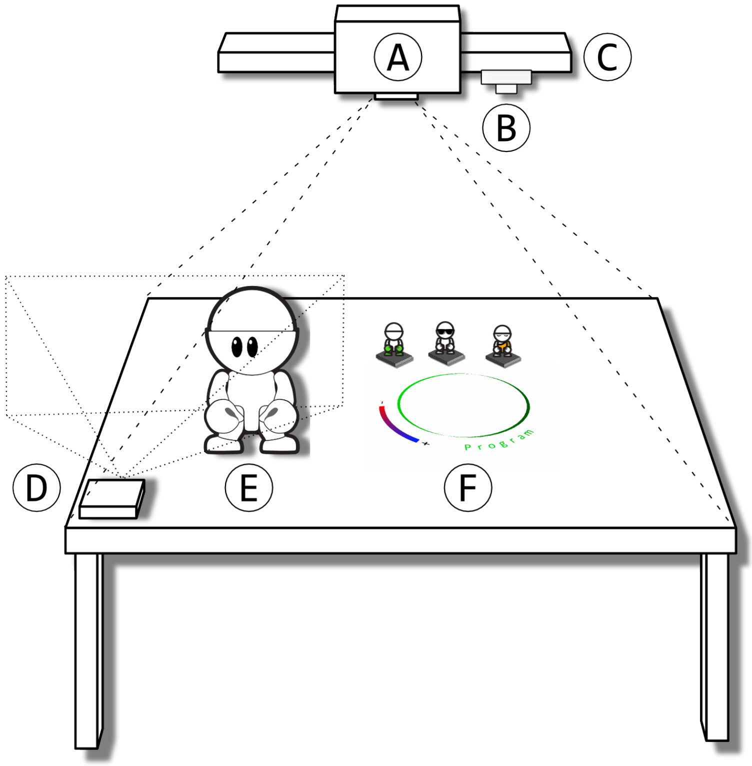

Figure 10.8 – Diagram of the installation. (A) ProjectionDesign F20SX projector (B) Sony PSEye web camera (C) OptiTrack V120:Trio (D) Vivitek Qumi Q2 projector (E) Teegi (F) Program selection zone and mini-Teegis.

to 1) evaluate the general usability of the interface and 2) obtain initial feedback about the relevance of the approach to help users understand EEG signals and the brain. Ten participants – 6 males, 4 females, mean age 28.6 (SD=9.7) – took part in this study. Pre-tests confirmed they were rather naive on the subject. They manipulated the version of Teegi described in the General Description section (no advanced features). The general procedure was as follows:

1. Pre-tests: The participant answered a first questionnaire assessing his or her representation of the brain. The participant then filled in different forms to measure his or her previous knowledge; one form per studied brain process (motor, vision and meditation, see Figure 10.9, top).

2. Setting-up: The experimenter positioned the EEG cap on the participant's head. In parallel, the participant, guided by the experimenter, was made aware of the four didactic "cards" explaining the different filters i.e. *Motor*, *Vision*, *Meditation* and *Raw* (Figure 10.9, middle). Each card was comprised of an image of the mini-Teegi associated with the filter along with basic instructions

to follow (e.g. the *Motor* card indicated to the participants to move their hands or feet while staying relaxed). There were also two cards describing the two types of visualization participants could face, *signal strength* and *synchronization* (Figure 10.9, bottom). Once the participant was equipped, a quick calibration phase occurred. While Teegi was still inactive, participants were asked to close their eyes for a few seconds, and to move their hands and feet in order to identify the baseline activity for visualization.

3. Personal Investigation: The participant was asked to freely manipulate Teegi as well as the filters to be able to answer the following questions:

    - What happens when you move your hands or feet?
    - What happens when you close your eyes?
    - What happens when you meditate?

    During the whole study the participant sat comfortably in a chair. To avoid the occurrence of muscle artifacts that may pollute the signals, the user was instructed to stay relaxed and to refrain from making strong head movements.

4. Post-tests: The participant answered the questions above on dedicated forms, the same that were given at the beginning of step 1. Finally, he or she filled in a user survey questionnaire based on a 7-point Likert scale.

The whole session lasted approximately 1.5 hours per participant, with 15 to 20 minutes of hands-on time with Teegi. Each session was video-recorded. Video segments were separately visualized and labelled with the corresponding behavior (i.e. tangible and visual interactions, emotional expressions, and investigation strategies) using The Observer XT 11.5 (Noldus, Info Tech, Wageninen, The Netherlands). After the session, the experimenter had an informal talk with the participant. He corrected the answers, making sure the participant was not leaving with false knowledge, and explained in more detail some aspects of the system (e.g. relationship between visual filter and attentional states, the various effects of meditation, ...). This phase lasted from 30 min to 1 hour depending on the participant's curiosity.

10.5.2 Results and discussions

To better understand the inherent strengths of Teegi towards learning, we assessed three main aspects of Teegi: its technical reliability, its relevance to ease understanding for non-experts, and the User eXperience (UX) it provides. This evaluation is based on 1) the results of the questionnaire that are summarized in Figure 10.10, 2) the analysis of the video recordings, and 3) the analysis of the forms the participants filled in to assess their pre and post-knowledge of the brain and EEG.

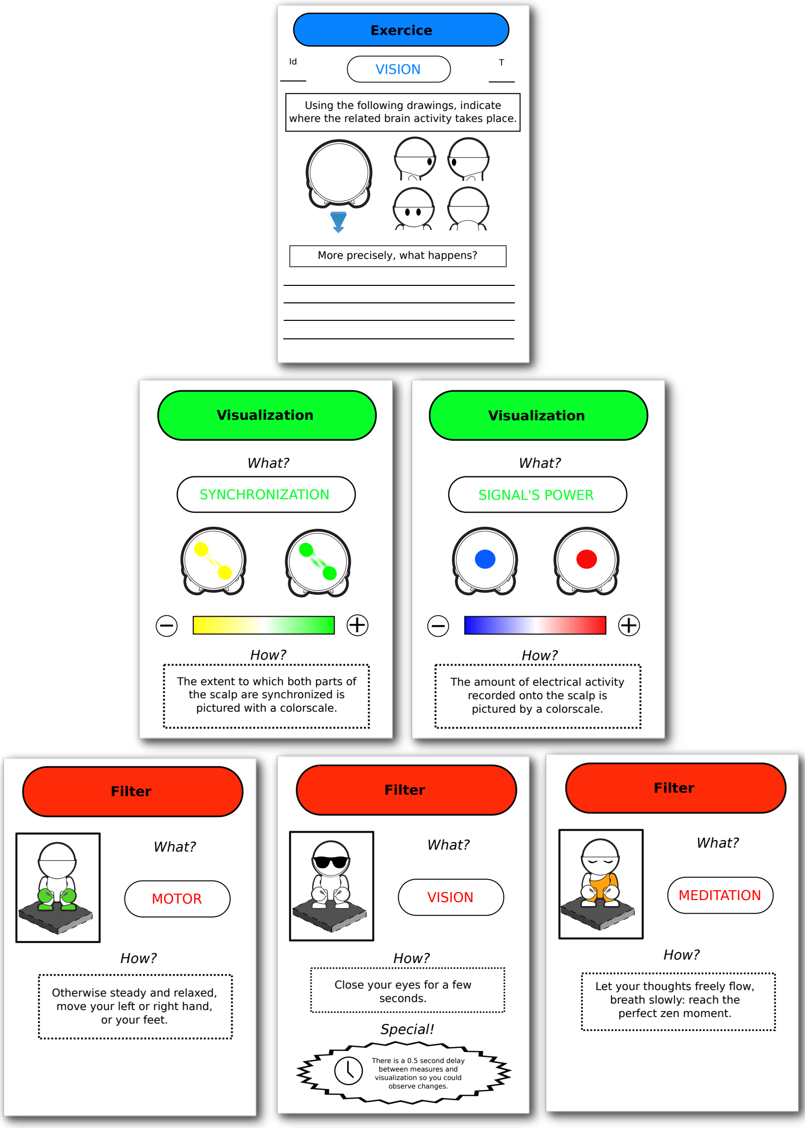

Figure 10.9 – *Top*: An example of question card given during pre and post test. *Middle*: Cards decribing the meaning of the visual feedback. *Bottom*: Filters that could be investigated with Teegi.

10.5.2.1 *Technical reliability*

Participants unanimously reported that the whole system worked properly. The quality of the SAR display is valued by the participants. In particular, they reported that the resolution was appropriate, and they did not report problems of offset between the display and the physical character. Participants declared that they were not disturbed by occlusion problems. The mild temporal delay between their action and their consequences seems not to be an issue.

Manipulations of Teegi were numerous and frequent. Teegi was touched or moved on average 25% of the session's duration, twice

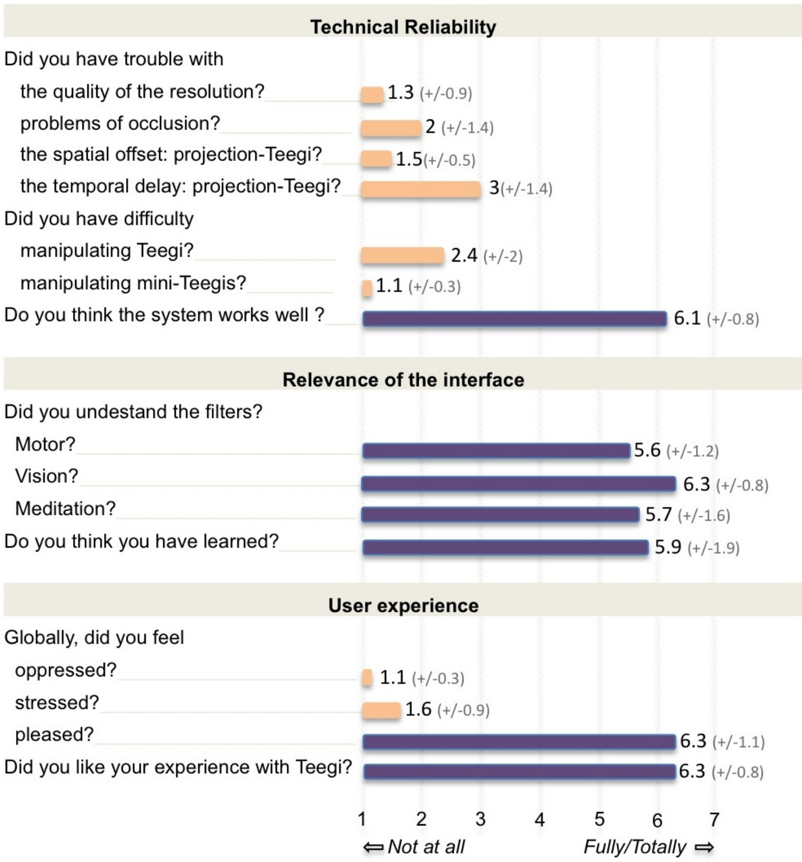

Figure 10.10 – Results of the questionnaire (selected questions). Note that purple (resp. orange) bars indicate questions measuring Teegi's qualities (resp. limitations).

IV.10

per minute. These manipulations consisted mostly of rotations, and to a lesser extent of lifting Teegi to enhance visual perception. Two participants reported difficulties in grasping Teegi while the remaining 8 were comfortable with the form of the character. Video analyses did not show difficulties for the manipulation of Teegi. Similarly, applying filters by manipulating the mini-Teegis seemed easy for the participants.

10.5.2.2 *Relevance of the interface to ease understanding*

The participants reported that they understood the visualization associated with the filters. Video analyses indicated that they systematically used all filters several times (3 times per session on average) for a similar duration – Raw filter: 30.4% (SD 13.3) of session duration; Motor filter: 26.0% (8.3); visual filter: 16.9% (5.5); meditation filter: 26.6% (8.7). Interestingly, the visual activity filter seemed slightly easier to understand than the other filters. Moreover, video analyses indicated that the participants did not have difficulty observing the signals on Teegi's head, as soon as they found the right location to observe. Overall, participants reported that they were able to use Teegi without any difficulties.

All participants completed the required tasks. They used instruction cards 5 times per session on average. They reported that they could focus on the tasks rather than on the mechanisms used to achieve them. This suggests that Teegi is a rather transparent interface. Regarding learning of brain processes and EEG, participants reported that they believed they had learned while doing the study. This was confirmed by the results of the pre- and post-test assessments (see Figure 10.11). These assessments focused on the recognition and the understanding of brain activation during Motor activities, Visual activities and Meditation. Understanding was marked as acquired if 1) the activated areas were correctly localized and 2) the explanations of the brain process were correct. It was marked as under way if only 1) or 2) was satisfied but not both; and as not acquired if neither 1) nor 2) were satisfied. The marks obtained by the participants improved after using Teegi. Overall, this suggests that Teegi offers many interesting features to ease learning and mediation.

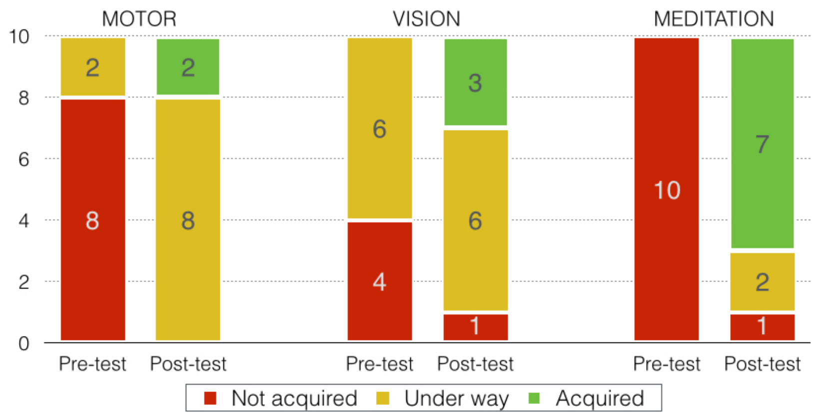

Figure 10.11 – Marks obtained by the participants during the pre- and post-test assessments. See text for details.

All our results indicate that Teegi clearly promotes real-time tangible interactions, which contributes to enhancing awareness. Constructivism and inquiry-based science education principles indicate that, to become conscious of complex phenomena and construct scientific knowledge, people/learners have to experiment by interacting with and physically manipulating the content [Vosniadou, 1992]. This is particularly true for brain activity that is difficult to understand because it cannot be sensed [Damasio, 1994], contrary to other activities (e.g. respiratory) that are perceived through sensory-motor mechanisms. Hence, brain activities need to be conceptualized, and the success of learning processes strongly depends on the interface. Teegi, which has been largely promoted by the participants, seems to fulfill this function.

IV.10

10.5.2.3 User eXperience

The general experience with Teegi was rated as pleasant, attractive and stimulating, and participants did not feel stressed or oppressed. Overall, participants reported that they liked interacting with Teegi. The emotion expression analyses confirmed those statements. They showed that on average participants expressed curiosity and questioning about Teegi feedback during almost 20% (20.1% SD=9.1) of the manipulation duration. Other emotion expressions observed for all participants were joy and pleasure (e.g. smile, laugh, joyful verbal expression…). They occurred during almost 10% (9.8% SD= 6.7) of the interaction duration with Teegi. Surprise emotions were observed but less frequently. Interestingly, boredom, weariness expressions rarely occurred (only for 2 users) and only at the end of the manipulation time. We did not observe any occurrence of exasperation or irritation. These results suggest a high level of acceptance for Teegi. This is a fundamental requirement for a tool aiming at improving access to knowledge.

Behavior observations indicated that the majority of participants spoke with Teegi and used morphological zones specific to human interactions while manipulating it. For example, they held its hands and held it up by the waist as one would do with a child. Some users spoke in the first person when they observed changes on the character's scalp for example "so, when I move my hands, I light up on the sides"; many said aloud that Teegi was their own image, for example "so, Teegi is me!". This identification suggests that an activation of associations between the perceived character's personality and self-perception may have occurred [Paiva et al., 2005]. It is known that identification can be associated with increasing loss of self-awareness, and its temporary replacement with elements of the perceived character's personality [Cohen, 2001]. Therefore, a human shaped, child-like character, made lifelike by animated projected eyes, could enhance both empathy and implicit self-perception of one's own brain activity, as provided by our interactive media. The anthropomorphic appearance of Teegi could explain the motivation and positive UX reported by the users. All these hypotheses would be the aim of a more extensive UX study.

Such engagement toward an anthropomorphic character encouraged us to keep this form factor with Tobe in next chapter.

Regarding visual attention, the participants were apparently paying attention to Teegi most of the time (83.3%, SD 7.6). This supports the fact that Teegi mobilized user attention. It also indicates a cognitive user engagement. Personal investigations were permanent (only 1.9% of inactivity was measured during the session duration; SD=1.7). Behavior analyses indicate that participants made predictions, hypotheses and tested them by conducting experiments. Numerous trial and error strategies were frequently used. This clearly indicates personal active control of the task and inquiry processes. Overall, Teegi stimulates investigations and encourages persistence in task completion.

10.6 Teegi for scientific outreach

We built Teegi in part to bridge the gap between, from the one hand, BCI researchers and, from the other hand, the general public and the media. The BCI community agrees that scientists should prevent inaccurate statements and correct representations in the media [Nijboer et al., 2011]; a tool such as Teegi could facilitate the dialogue and ease the transition from the laboratory to the outside world.

Figure 10.12 – Teegi has been used for scientific outreach during the science and technology festival *IIT Techfest*.

IV.10

Since we first published our work in [Frey et al., 2014a], we developed a portable version, based on the Emotiv Epoc headset, that we demonstrated on several occasions, both in local and in international manifestations. Notably, we were invited to present Teegi during the *IIT Techfest* in Bombay, the largest science and technology festival in Asia (≈ 200 000 attendees, see Figure 10.12). We could appreciate in the field how Teegi raised people's interest. Visitors were driven by their curiosity and we used this opportunity to introduce them to BCI technologies and discuss with the public issues related to the field.

Techfest version is powered by Jérémy Laviole's Papart framework.

Our various interventions echoed in online mass media. For example, not only did a major hub dedicated to new technologies relayed our work[2] or a journal picked Teegi as one of the "best projects" of the fair we've been demonstrated in[3], but a general information website saw the potential of our approach for teaching about brain activity[4]. Teegi

Sadly, Yahoo also demonstrates by few off sentences that there is still a long way to go before a true comprehension of BCI.

[2] http://gizmodo.com/7-experimental-interfaces-that-show-the-future-of-ui-de-1642890943

[3] http://www.dnaindia.com/scitech/slideshow-the-best-projects-at-the-iit-bombay-tech-fest-2015-2048986

[4] https://www.yahoo.com/tech/using-brain-wave-technology-to-create-art-108761329324.html

proved to be an effective medium to raise awareness, demonstrating the viability of the project for scientific outreach.

10.7 Conclusion

In this chapter, we presented Teegi, a tangible interface that makes EEG understandable to non-expert users. Our main contribution is the interface itself, which is built from both theoretical foundations, notably from human learning and scientific mediation and technical developments, including spatial augmented reality, tangible interaction and real-time neurotechnologies. We demonstrated that this interface was well accepted by a first pool of users. We also showed that it appealed to novices' interest in public exhibitions.

In the future, we plan to make a more in-depth investigation into how well users are able to learn about EEG and brain activity with Teegi. To this end, we will conduct dedicated experiments with students and/or visitors in scientific museums. We would also like to precisely evaluate how Teegi benefits learning compared to standard approaches. For more advanced users, ad-hoc tangible filter creation could prove to be of great interest, adding flexibility to the overall system. Finally, it is known that BCI requires the user to learn to control his/her own brain activity to input computer commands [Wolpaw and Wolpaw, 2012], which is a long and tedious task. We expect Teegi to be a motivating and informative way to support this training.

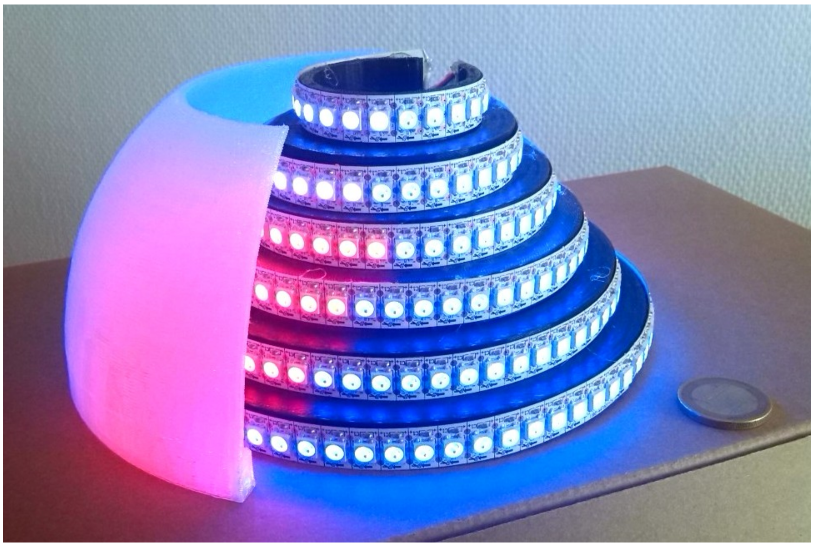

Figure 10.13 – Prototype of a spherical display composed of LEDs (Adafruit NeoPixels) that could replace projection to display EEG signals with spatial augmented reality.

In order to ease the deployment of Teegi, we are currently finishing an instrumented version of the puppet, that is using a spherical display made of LEDs (Figure 10.13) and that embeds signal processing thanks to a Raspberry Pi inside the body – RFID chips within the mini-Teegis would be used to to select the EEG filters. No external hardware will be required anymore beside the EEG device. This will significantly cut the costs and will favour the spread of such tool.

The next chapter addresses a project that springs from Teegi. Although "Tobe" is as well a tangible avatar that deals with EEG, it encompasses other physiological sensors – ECG, EDA and breathing – and is more about high level mental states – such as attention level or emotions – rather than preprocessed physiological signals. The framework underneath has also been improved so it is now scalable, with multiple avatars that could exist at the same time. Thanks to those new possibilities, with Tobe the focus shifts from scientific outreach to social interactions.

11

Tobe: Tangible-Out-of Body Experience

We present a toolkit for creating Tangible Out-of-Body Experiences: exposing the inner states of users using physiological signals such as heart rate or brain activity. Tobe can take the form of a tangible avatar displaying live physiological readings to reflect on ourselves and others. As a toolkit, it can help the general public familiarize itself with Science Technologies Engineering and Mathematics (STEM) disciplines and cognitive science. Through a co-design approach, we investigated how everyday people picture their physiology and we validated the acceptability of Tobe in a scientific museum. Finally, we describe a "design space" that frames how Tobe could be put into practice – in a medical context or not, whether there is one or several users or Tobes involved. We give a practical example of a scenario where 2 users have to relax together, with insights on how Tobe helped them to synchronize their signals and share a moment.

Pinacle of this thesis, the Tobe platform aims at eventually holding it all together. Mental states (I) measured with passive BCI (II) using practical sensors (Appendices) could facilitate social interactions (III) once made tangible (IV). A social prosthesis to foster social presence and empathy among peers.

The work described in this chapter was presented at TEI '16 [Gervais et al., 2016].

Many different persons participated in one way or another in this work. Besides my co-authors, I would like to thank Didier Laval from Cap Sciences[1], Pierre-Alain Joseph, Éric Sorita, Matthew S. Goodwin and Christelle Godin – more about their respective role in the appendices, section Credits.7.

11.1 Introduction

Wearable computational devices are more accessible and more popular than ever. These devices are personal and could be embedded with physiological sensors similar to the ones we have seen thorough this manuscript – i.e. sensors that can monitor signals such as heart beats or electrodermal activity. Nowadays even brain activity is within reach of consumers thanks to cheap alternatives to medical equipment, such as the Emotiv Epoc[2] or, closer to the Do-It-Yourself community, the OpenBCI board[3] (see appendix C). Physiological computing is becoming mainstream, however for the general public the use of such sensors seems mostly centered around performance. Despite an era of personal development, well-being and communication, how many smart watches and heart rate belts advertise themselves as sportspersons' best buddies, while they can account for so much more than *physical* health? Indeed, besides the social applications retold in part III, physiological computing is mature enough to asses mental states [Fairclough, 2009, Picard, 1995, Zander and Kothe, 2011, Frey et al., 2014b] (part I). Therefore, it could be used as a mean to better know our own self and others.

On the one hand, physiological technologies are not exploited to their full potential, on the other hand, we have end users that ignore what technology has to offer for their well-being. Some companies are pioneers, as for example Empatica and its Embrace smart watch[4], but such companies focus on health applications – e.g. monitor with epilepsy to predict seizures – and, consequently, the targeted consumers are still a niche. Both a process that will raise public awareness and a collection of meaningful use cases are missing. Finally, when bodily activity and mental states are at stake – which are difficult to conceptualize and often difficult to perceive – the feedback given to users matters for them to comprehend at first sight what is being measured. How to represent the arousal state of someone? How would *you* represent cognitive workload? We found little examples besides pies and charts, which are not always obvious informants in data visualizations – e.g. [McCandless, 2010].

IV.11

[1] http://www.cap-sciences.net/
[2] https://emotiv.com/
[3] http://www.openbci.com/
[4] https://www.empatica.com/

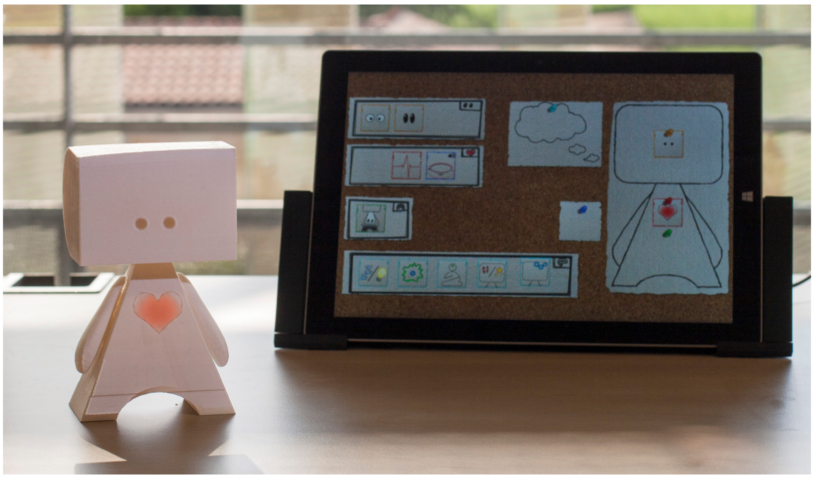

Figure 11.1 – Tobe, the tangible avatar displaying real-time physiological readings along with the interface to control the different visualizations.

To address these issues, we first conducted surveys and interviews to gain insight about physiological feedback. We then created Tobe (to be pronounced “`tobi`”), a Tangible Out-of-Body Experience shaped as a tangible avatar (Figure 11.1). This avatar lets users freely explore and represent their physiological signals, displayed on the avatar itself using spatial augmented reality. The overarching goal is to help one reflect on his physiological and mental states in *his* own way. The main activity would be for users to actively *build* from the ground up their own self-representation and then visualize physiological signals through it. As such, we designed a modular toolkit around Tobe that can be used to customize any part of the system. Tobe has been tested on two different occasions in a scientific museum to collect user feedback. A specialized version of the system was also built to give biofeedback to multiple users in a relaxation task. Beside these two implementations, we frame potential uses of the system, such as a biofeedback device for stroke rehabilitation or replaying inner states synchronized along with videos of cherished memories. The latter example could help create more cherishable versions of personal digital data [Golsteijn et al., 2012].

Previous works do not embrace such system as a whole and are limited either to low-level signals or to emotions. Wearables were used in [Williams et al., 2015] to mediate affect using multimodal stimuli (sounds, heat, vibration, …). However, as with the “Social Skin” project [Uğur, 2013] – that also embodies emotions into actuated wearables –, the information given to those around was rather implicit. When a more comprehensible feedback was studied, as in [Norooz et al., 2015], it was limited to anatomical models, for instance to teach children how the body works. Tobe, on the other hand, gives both access to meaningful visualizations and to additional cognitive states. Tangible proxies and material representations were already studied in [Khot et al., 2014], although the feedback was not dynamic and, once again, constrained to bodily activity. With Teegi (chapter 10) a tangible puppet was already used as a proxy for brain activity, but the settings concerned scientific outreach and the feedback focused only on preprocessed brain signals and not on higher level mental states (which Tobe does). Our toolkit pushes further the boundaries of the applications. By giving access to physiological signals, high-level mental states, dynamic and customizable feedback, a tool that helps to communicate and that facilitates social interactions have emerged.

Our contribution for this chapter are:

1. A toolkit enabling users to create an animated tangible representation of their inner states, encompassing the whole workflow including the physical avatar creation, sensors, signal processing, feedback and augmentation.
2. Two use cases of Tobe which were tested in public settings.
3. Users' feedback about the Tobe system and how they perceive physiological signals and mental states.

11.2 Representing Physiological Signals

Exposing physiological signals in a way that makes sense for the user is not trivial. Some types of signals might be more obvious to represent than others. For example, heart activity could be understood using a symbolic heart shape due to largely accepted cultural references. This question is, however, harder when talking about more abstract mental states such as workload. Nevertheless, even the dynamic representation of low-level physiological signals is still an open question at the moment [Chanel and Mühl, 2015]. We conducted two surveys to gain more insight about the knowledge and the representation people had of different types of signals and high-level mental states.

In the first survey, conducted online, we asked 36 persons about their knowledge of physiological signals in general. We inquired about the self-awareness of inner states on a 7-points Likert scale (1: no awareness, 7: perfectly aware). About "internal physiological activities", the average score was 3.5 (SD=1.4) and for "mental states", the average score was 4.9 (SD=1.3). The latter score indicates that the participants thought they *knew* their inner states – even though a whole literature demonstrates how difficult this is [Nisbett and Wilson, 1977]. Interestingly, we also observed that most of the participants reduced mental state and physiology to emotions only. Mentions of any cognitive processes such as vigilance and workload were very rare (7 out of 36). This lack of knowledge about the inner self and the different cognitive processes is an opportunity to raise awareness of the general public about the complexity of the mind. When inquired about possible uses of a Tobe system, very few respondents (6 out of 36) gave examples other than sports or health. This emphasis the fact that the general public is unaware of possibilities of technology for well-being.

With motion we're entering in the robots' realm

The second survey specifically investigated how users would shape the feedback. We focused on visual cues because it was easier to express on paper, but note that other modalities could be explored. Sound is among them – e.g. [Janssen et al., 2013],Mealla2011 and chapter 8 – but motions may play also an important role, especially when affect is involved [Cooney et al., 2014].

We asked 15 participants to express with drawings and text how they would represent various metrics (Figure 11.2). There was little resemblance between participants for a given high-level metric and even low-level ones – breathing and heart activity – sprang different views. For example, some people drew a physiologically accurate heart instead of a simple sketch. Overall, there was a wide variety of sketches and people were very creative. This highlighted the absence of consensus on how we conceive and view our inner states. Therefore, people could benefit from being able to tailor a meaningful and personal feedback.

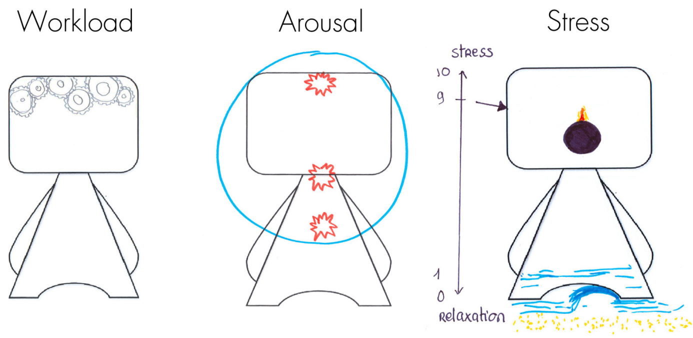

Figure 11.2 – Sample of the drawings made by participants to represent various high-level metrics.

11.3 Toolkit

We created a tangible anthropomorphic avatar, named Tobe as a host for displaying real-time feedback. We chose this form factor because we found evidence in the literature that this combination of anthropomorphism and tangibility can foster social presence and likability [Schmitz, 2010, Hornecker, 2011, Horn et al., 2009]. This also reminds users and observers that the feedback is linked to an actual being; it helps to recognize Tobe as a persona and to bond with it, hence it facilitates engagement. Finally, this was a logical step forward after we had developed Teegi. During the work described in the previous chapter, we witnessed firsthand the benefits of such type of proxy.

Our implementation uses open or low-cost hardware and we are in the process of releasing as open-source software the entire pipeline, thus facilitating reproduction and dissemination.

11.3.1 General Approach

We conceived a toolkit to assist the creation of representations of inner activities – our body at large and the hidden processes of our mind in particular, making it visible to oneself and to others. The different components are highlighted in Figure 11.3. The first step consists in choosing a metric, e.g. the arousal level. For this given metric there are different ways to measure it, that include a combination of one or multiple sensor(s) and signal processing algorithm(s). One chooses a support to express this metric (e.g. tangible avatar, screen, speaker for sound) and creates a shape associated to it (e.g. a circle with a changing color, a rhythmic tone). The conjunction of both the shape and the support produces the feedback. It is an iterative process because when one acknowledges the feedback, it changes one's self-representation. Moreover, it creates a feedback loop which affects one's biosignals.

In order to help users mold the system to their likening, we identified three different degrees of freedom:

- The measured physiological signal or mental state (Metric)
- The form factor (Support)
- The display of the signals (Shape)

Beside the opportunity to answer to a specific need, the process of *building* something matters. Between 2 equivalent things, we prefer the one that took us efforts to achieve / acquire [Norton et al., 2012], hence the bond we seek between users and Tobes should be enhanced at the same more degrees of freedom are given.

11.3.1.1 *Sensors and Signal Processing*

Sensors are the hardware used to capture the raw signals of the body. These encompass heart measures such as electrocardiography (ECG) and photoplethysmography (PPG), brain activity measured by electroencephalography (EEG), electrodermal activity (EDA, i.e. perspiration), etc. Once the raw data is acquired, it needs to be processed in order to produce any relevant metric. As an example, heart rate variability can be inferred from the combination of a set of electrodes attached to the chest and a QRS wave detector. Emotions can be inferred from EDA or certain frequency bands within EEG.

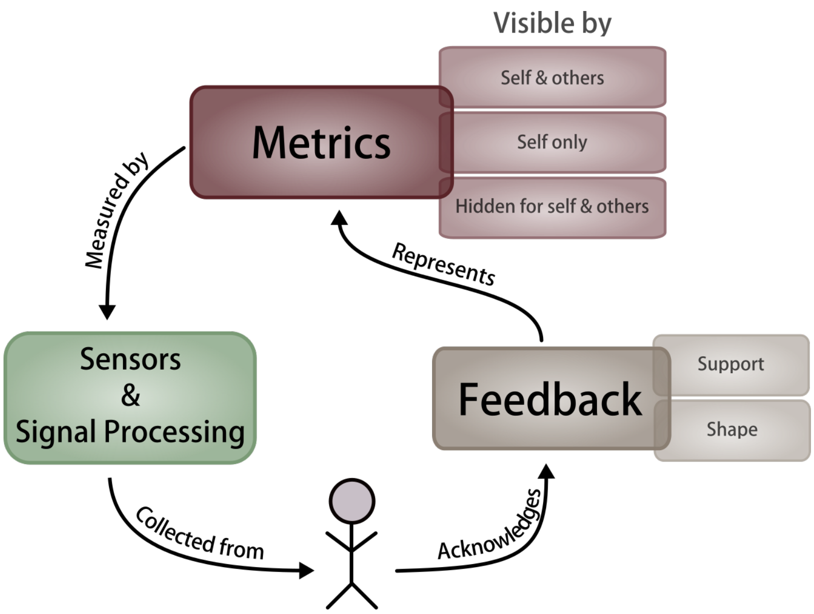

Figure 11.3 – Simplified view of the toolkit that supports Tobe.

11.3.1.2 *Metrics*

There is a continuum in the visibility of the signals and mental states measured from physiological sensors, i.e. metrics. We categorized those metrics in three different levels, depending on who can perceive them without technological help. Each level relates and links differently the 3 actors involved, i.e. the user, the observer and the personae (Tobe).

1. Perceived by self and others, e.g. eye blinks. Even if those signals may sometimes appear redundant as one may directly look at the person in order to see them, they are crucial in associating a feedback to a user.

2. Perceived only by self, e.g. heart rate or breathing. Mirroring these signals provides presence towards the feedback (e.g. similarity-attraction effect, chapter 8).

3. Hidden to both self and others, e.g. mental states such as cognitive workload. This type of metrics holds the most promising applications since they are mostly unexplored.

Lower levels (1 & 2) help to breath life into a proxy used to mediate the inner state of the user. These metrics are accessible to our conscious selves; they are likely to increase the social presence of the proxy (see chapter 8) and they participate in the “minimal features” that an embodied agent should possess in order to create human presence [Sumioka et al., 2014]. On the other hand, level 3 metrics are little known and are hard to conceptualize for the general public [Nisbett and Wilson, 1977] and would benefit the most of a system enabling their visualization.

11.3.2 Support: 3D printing

3D printing a tangible avatar is a powerful incentive for customization. While the version of the system that we deployed in the scientific museum used an already modeled and 3D printed incarnation of Tobe because of time constraints, a user of the system could change the parametric model in order to obtain an avatar that pleases her. The process would be similar to how the appearance of a Nintendo "Mii" can be tuned, except for the tangibility. An automatic feature extraction could even occur to match some of the puppet's traits with the user (e.g. body shape, nose, chin, ...). As a tradeoff between preparation time and customization, we prototyped a "Mr. Potato Head" version of Tobe, with parts ready to be assembled (Figure 11.8).

11.3.3 Sensors: Wearables

Figure 11.4 – Wearables: coat embedding ECG sensors.

Metrics were acquired from five physiological signals. We measured thoracic circumference for breathing, ECG for heart rate, EDA for arousal, electrooculography (EOG, eyes activity) for eye blinks, and electroencephalography (EEG, brain activity) for most high-level mental states.

We created the sensors with a wearable form factor in mind. Since we used Tobe in public settings, it was important that the sensors were non-invasive (no need to remove clothes or apply gel to the skin) and be quick to install and remove, while being able to acquire a reliable signal, and no products on the market could entirely respond to our needs. With the setup described in this section, we were able to equip the users and record physiological signals in less than two minutes.

The different sensors were embedded inside a lab coat (Figure 11.10) which could be put on quickly over daily clothes. This form factor

Figure 11.5 – Wearables: fingerless glove measuring EDA.

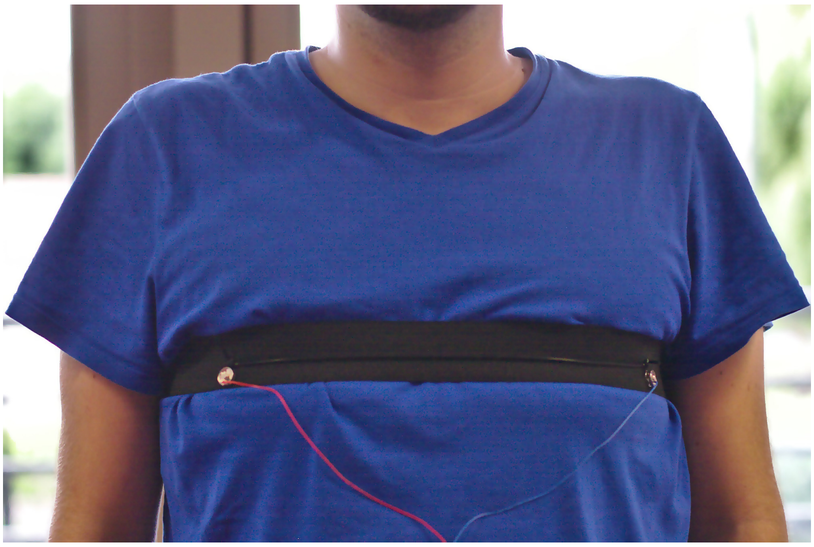

Figure 11.6 – Wearables: breathing belt.

provides enough room in the sleeves and the pockets to take care of the wiring and electronic components storage. The recording of the low-level physiological signals (i.e. everything except EEG) is done using the BITalino board, an Arduino-based recording device. It contains modules that amplifies various physiological signals and embeds a Bluetooth adapter as well as a battery to work in ambulatory settings.

11.3.3.1 *ECG*

We chose to use ECG for heart rate activity as it is more accurate than light emission-based methods to detect individual heartbeats [Kranjec et al., 2014]. Existing solutions for ECG require electrodes to be put directly on the chest, e.g. heart rate monitor belts. We instead

TDE-201 from FRI are the "pins free" version of TDE-200, itself sold as "EL120" by Biopac.

opted for installing TDE-201 Ag/AgCl electrodes from Florida Research Instruments (FRI) on both wrists of the user (ECG needs two electrodes diametrically opposed to sense heart electrical activity). The electrodes were attached to an elastic band sewed inside the end of the lab coat sleeves which could be tightened with velcro straps (Figure 11.4). ECG was recorded with the dedicated ECG module of the BITalino.

11.3.3.2 *EDA*

When measuring EDA, most accurate readings can be obtained from the tip of the fingers. However, since it is difficult to manipulate a tangible interface and controls while having hardware attached to one's fingers, we acquire the signal from the palm of a single hand instead – the palm is good tradeoff between accuracy and practicality, compared to fingers or wrist [Prasad, 2013]. We assess skin conductance from two small conductive thread patches sewn inside a fingerless glove (Figure 11.5). Because the BITalino EDA amplifier was not sensitive enough for signals acquired from the palm we made our own, replicating the schematics described in [Poh et al., 2010].

11.3.3.3 *Breathing*

For breathing, we built a belt based on a stretch sensor (Figure 11.6). A conductive rubber band was mounted as a voltage divider and connected to an instrumentation amplifier (Texas Instruments INA128). As opposed to piezoelectric components, that are sensitive to momentous speed instead of position, stretch sensors can directly map users' chest inflation onto their avatar. An alternate solution consists in measuring air flow near the mouth of the nose, but it would have been too obtrusive.

IV.11

11.3.3.4 *EEG and Eye Blinks (EOG)*

We built our own EEG helmet based on the open hardware OpenBCI board. To shorten setup time we used dry electrodes – the same TDE-201 as for ECG for the forehead, and elsewhere TDE-200 electrodes, which possess small protuberance that could go through the hair. Using a stretchable headband, we restrained electrodes' locations to the rim of the scalp to avoid difficulties with long-haired people. In the 10-20 system, electrodes were positioned at O1, P7, F7, FP1, F8, T8, P8 and O2 locations – reference at T7, ground at FP2. Appendix C is dedicated to the problematic surrounding the construction of such practical headset, which required more thorough engineering compared to the other sensors presented in this chapter.

In earlier iterations of the system we tested the use of an Emotiv EPOC headset to account for brain activity. The EPOC is a consumers-oriented EEG device, easier to install than medical headsets that use gel. However, it still requires a saline solution that tends to dry over time, causing additional installation time between users. Moreover, good signal quality was next to impossible to obtain with long haired persons.

11.3.4 Signal processing

Consumers-oriented EEG headsets usually conceal signal processing behind proprietary algorithms, with little scientific evidence on what is truly measured. While building a tailored EEG helmet, we took the upper hand on the whole pipeline. With access to raw EEG signals, we looked into the literature to match the inner state we wanted to measure with actual neurological markers.

Here our approach somewhat varies from the studies conducted in part II. During the evaluation of HCI components, the experimental setup involved participants willing to spend at least half an hour for calibrating the system, and only in a second phase did we use machine learning to assess their mental states. With Tobe, on the contrary, we could not require users to endure such procedure during our preliminary investigations. Our interventions were meant to be "wear and play". Therefore, instead of pipelines made of external probes, N-back tasks or classifiers, we have used the following features (see chapter 3 for more background about those constructs):

- Vigilance: appoints for the ability to maintain attention over time. We use the ratio between beta frequency band (15-20Hz) and theta + low alpha frequency band (4-10Hz) for all electrodes [Oken et al., 2006].

- Workload: increases with the amount of mental effort required to complete a task. We use the ratio between delta + theta band (1-8Hz) in frontal cortex (F7, FP1, F8, T8) and wide alpha band (8-14Hz) in parietal + occipital cortex (P8, P7, O2, O1) [Antonenko et al., 2010, Schober et al., 1995].

- Meditation: we used the same computations as for Teegi, e.g. instantaneous phase locking value [Lachaux et al., 2000] between front (FP1, F7, F8) and rear (O1, P7, P8) parts of the brain in alpha + beta bands (7-28Hz) [Lehmann et al., 2012] – mindfulness and body focus practices decrease the synchronization while transcendental practice increases it.

- Valence: designates the hedonic tone of an emotion and varies from positive to negative (e.g. *frustrated* vs *pleasant*). We use the ratio between the EEG signal power in the left (F7, P7, O1) and right (F8, P8, O2) cortex in the alpha band (8-12Hz) [Molina et al., 2009].

- Arousal: relates to the intensity of an emotion and varies from calm to excited (e.g. *satisfied* vs *happy*). We use the EDA readings (even though there is actually more than a single psychological process behind this signal [Figner and Murphy, 2011]).

We used OpenViBE to analyze physiological data in real time. EEG signals were re-referenced using a common average reference. Mentioned frequencies were extracted with a band-pass filter, taking the log of the power of signals in order to normalize indices.

Those features constitute a rough estimation of the inner state of participants. Actual applications, as the one we describe in section 11.5, require more thorough computations, notably a calibration phase and machine learning to adapt features to each user and acquire truly meaningful signals (e.g. part II).

Eye blinks were detected when the signal, after DC drift removal, exceeded 4 times the variance in the F8 electrode. The detection of heart beats, hence of QRS waves recorded from ECG, was achieved by combining a similar automatic threshold with a 1-20Hz band-pass filter and a first-order derivative.

Each physiological signal or mental state index was sent to the other stage of the toolkit using LSL, a network protocol dedicated to physiological recordings that possesses implementations in many programming languages.

11.3.5 Shape: Augmentation

Mostly the same as Teegi, except for the calibration.

The visualization of users' signals are displayed onto Tobe using Spatial Augmented Reality (SAR), as introduced by Raskar et al. [Raskar et al., 2001]. SAR adds dynamic graphics to real-world surfaces by the means of projected light. Despite external hardware – i.e. a projector and eventually a tracking device (Figure 11.10) – SAR is an easy solution to prototype a system, faster to deploy than putting actual screens in users' surroundings. For instance we were able to switch instantaneously back and forth between a beating heart onto Tobe chest and a pulsing circle projected onto the table around it. The augmentation occurred within vvvv[5], a software that uses real-time visual programming to render 3D scenes. As for hardware, we used a LG PF80G projector of resolution 1920x1080 and the tracking of Tobe was achieved with an OptiTrack V120:Trio, running a 120 FPS with an overall latency of 8.3ms and a precision of 0.8mm. The projector was calibrated with the OptiTrack using OpenCV's camera calibration function.

As an alternative to SAR, Tobe can be embedded with small screens, LEDs, actuators and small electronics components so that it represents a standalone unit. We already have a proof of concept of such an implementation thanks to the easiness and accessibility of the building blocks that go with the Arduino platform and the Raspberry Pi (Figure

[5] http://vvvv.org/

11.9). This approach participates in the trend around ubiquitous computing [Weiser, 1993] and the Internet of Things – e.g. Flotilla[6], Printoo[7], the Airboard[8], Node-RED[9] (for software), and so on.

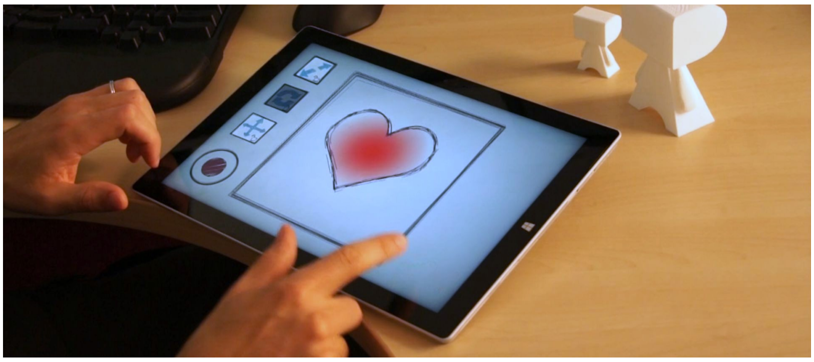

Figure 11.7 – Simple multitouch animator allowing users to create and animate visual feedback.

Figure 11.8 – Customizing the tangible support of Tobe can be achieved using modular body pieces.

11.3.6 Feedback Customization

We conceived a GUI that let users draw a picture and animate it according to their wishes. The animator is touch based; users press a "record" button and animate the picture with gestures (Figure 11.7). Once done, the animation's timeline is automatically mapped to the chosen signal. This animator is kept simple on purpose, it is designed for novice users

[6] http://flotil.la/
[7] http://www.printoo.pt/
[8] http://www.theairboard.cc/
[9] http://nodered.org/

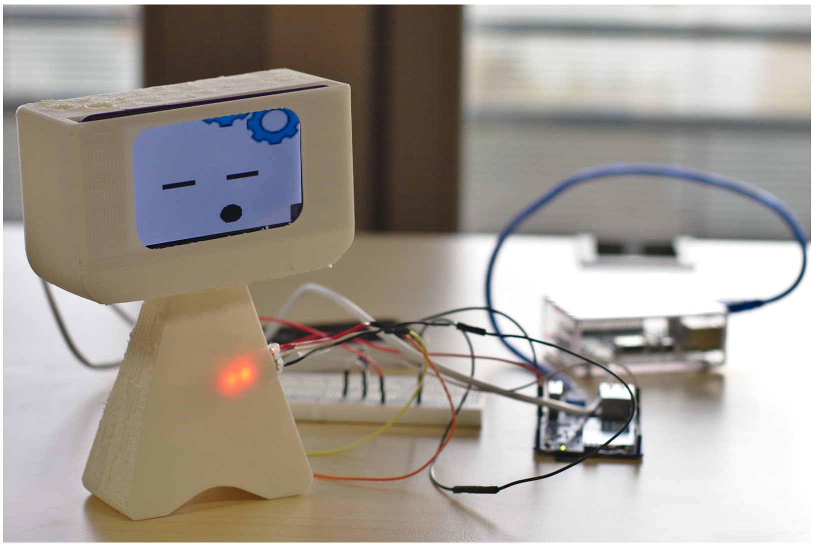

Figure 11.9 – It is possible to embedded electronics inside the support to have a standalone Tobe.

and as such must remain easy to understand and operate for someone not familiar with animation. Only three basic operations are currently supported – scaling, rotation and translation – and yet it is sufficient to generate meaningful animations. For example, scaling makes a heart beat, translation moves a cloud along respiration and rotation spins cogs faster as workload increases. An advanced tool such as Photoshop has already been integrated as a proof of concept, but the simplicity of the current GUI does not impede users' creativity and already is sufficient to enable a tailored feedback.

11.4 Tobe in the Wild

We used and tested Tobe in two different applications cases: as a demonstration in a scientific museum and as a multi-user biofeedback device for relaxation and empathy.

11.4.1 Tobe in a public exhibition

Using a co-design approach, we intervened in a scientific museum over two half days, proposing to passersby to try out Tobe (Figure 11.10). We built the sensors and prepared the signal processing beforehand because these steps require hardware and expertise. Five high-level metrics were selected: workload, vigilance, meditation, valence and arousal. These metrics were chosen because the wide public showed interest into them (meditation and emotions) or because they could benefit from being better known (workload and vigilance). Due to the short duration of our exhibitions, we also set the corresponding

Figure 11.10 – In a scientific museum, various activities were proposed to visitors in order to prompt self-investigation. The setup consists of a projector handling the augmentation and an OptiTrack for the tracking.

feedback (both support and shape), according to the outcome of the questionnaires about people's representations.

After we equipped participants, we gave them "activity cards", a collection of scenarios that were likely to modify their inner state and that prompted self-investigation (Figure 11.11). There were riddles, arithmetic problems, cute and *less* cute images, a breathing exercise and a "Where's Waldo?" picture. Implicitly the activity cards targeted in this order workload, valence, arousal, meditation and vigilance, but participants were free to test whatever they wanted. These sole cards, inspired by those made for Teegi (chapter 10), sufficed to engage participants for a few tens minutes without our intervention. Participants had also at their disposal "definition cards", very brief definitions of each one of the mental state that could be measured.

IV.11

We created the activity cards after our first intervention in the museum. There were some candies left at disposal next to Tobe to lure museum's visitors to our booth. At some point, one user wanted to see how different tastes affected the emotional valence that was displayed on Tobe. This proved to be a fun activity for him – and for the people around. Having such goal in mind was an effective way to drive participants. This is the upside to going in the field: we learn from our errors – we were short on incentives for self-investigation – *and* new ideas emerge.

One degree of freedom was left to users by the mean of a graphical interface (see Figure 11.1). They had to manipulate the GUI on a nearby tablet to drag and drop visualizations on predefined anchor points. Users could customize some of Tobe's aspects (eyes and heart rate feedback) and among the 5 high-level metrics available, they selected which one to study at a particular time. When at first we tested Tobe

Who never lost hours customizing the appearance of an avatar in a RPG video game?

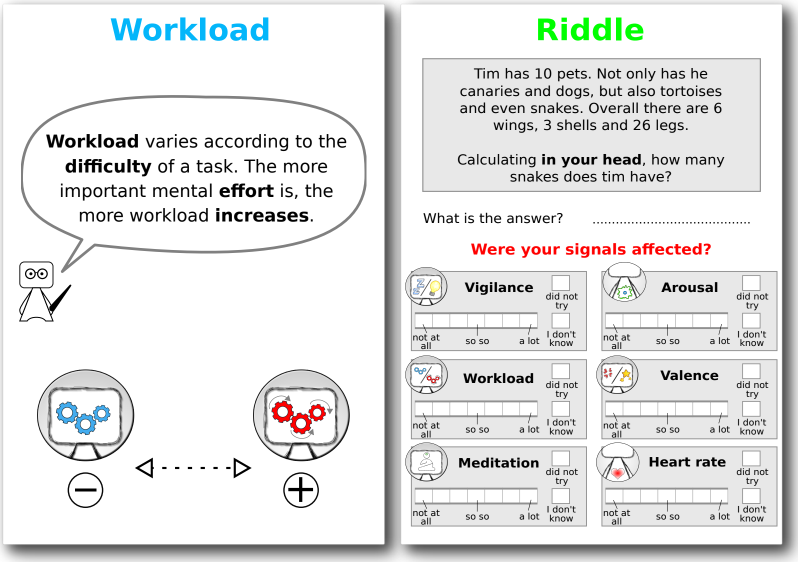

Figure 11.11 – An example of the activities and definitions proposed to participants in order to prompt self-investigation.

with *no* degree of freedom – i.e. all metrics were displayed altogether – we realized that users were too passive and quickly overwhelmed. The GUI helped to focus and engage users.

To further engage users, Tobe was tracked and participants were asked to put Tobe on a spotlight to "awake" it – i.e. to start physiological signals' streams. The action of bringing life to an inanimate puppet goes well with making the world "magical" again [Rose, 2014], that is to say to use the power of abstraction of modern computer science in order to bring back awe. The aim is not to take benefit of ignorance but to strengthen the amazement that technology can offer. We were ourselves pleasantly disturbed and surprised when we happened to hold in our hands a representation of our beating heart during some routine test[10]. Suddenly the relationship with the digital content felt different, truly tangible.

The "wake up" or "summoning" action could be use at start to choose one puppet among several possible models, thus bonding with the personae through touch.

The "wake up zone" was also an opportunity to implement the ambient feedback described in appendix E: the spotlight got pixelated and noisy depending on the quality of EEG signals'. By computing an index based on the upper beta band (25-45Hz) [van de Velde et al., 1998] we could detect artifacts produced by strong muscle activity – e.g. teeth clenching – or by participants' movements and give them a gentle feedback about the noise they were provoking.

[10] Truth is, that occurred back during the study of chapter 9.

11.4.2 Tobe for multi-users relaxation

We tested Tobe as a relaxation device for two users (Figure 11.12). The objective was to see if Tobe could be used both as a biofeedback tool and for collaboration.

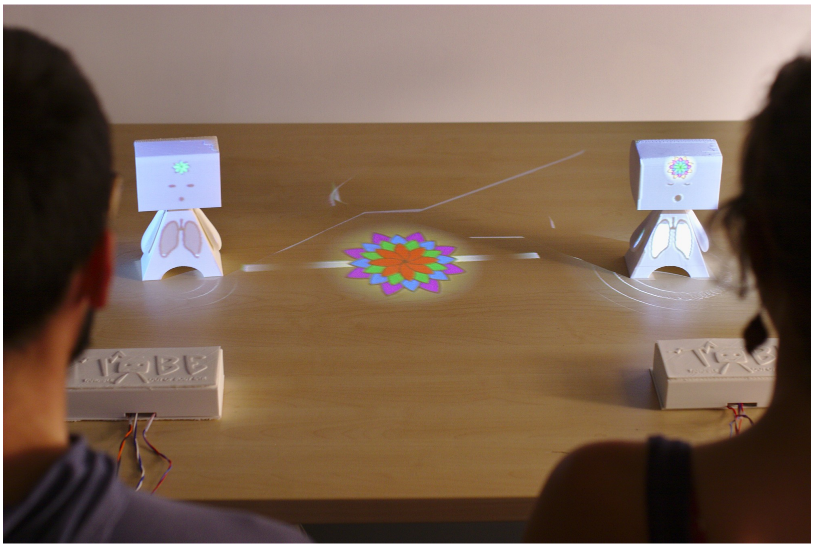

Figure 11.12 – Multi-users application: relaxation through cardiac coherence.

11.4.2.1 *Implementation*

This version of Tobe relies only on respiration and heart rate variability. It relates to cardiac coherence: when someone takes deep breaths, slowly ($\approx$ 10s periods) and regularly, her or his heart rate (HR) varies accordingly and the resulting state has positive impact on well-being [McCraty et al., 2009]. During cardiac coherence, HR increases slightly when one inhales and decreases as much when one exhales. We took the magnitude squared coherence between HR and breathing signals over 10s time windows as a "relaxation" index.

Sensors consisted in a breathing belt and in a pair of elastic bands around the wrists to measure ECG. We used OpenBCI instead of BITalino to measure ECG and breathing in order to get more accurate readings. Indeed, the OpenBCI amplifier has a resolution of 24 bits instead of 10 for the BITalino.

There were two Tobes on the table, one for each participant. They were not tracked. Breathing activity was pictured with inflating lungs onto the torso; cardiac coherence with a blooming flower onto the forehead. The synchronicity *between* participants – users' heart rates varying at the same pace – was represented with a similar but bigger flower projected between both Tobes. Additionally, "ripples" on the table, around Tobes' feet, matched heart beats.

11.4.2.2 *Protocol*

We asked 14 participants, by pairs, to come and use Tobe to reach cardiac coherence – 6 females, 8 males, mean age 25.3 (SD=2.8). Participants were coworkers from the same research institute and already knew each other. Participants were seated on each side of a screen and instructed to not talk to each other. We presented them the cardiac coherence activity as a relaxation exercise. Afterwards, we equipped them and turned the system on.

The experiment comprised of three sessions of 5 minutes. During the first session, participants had to individually learn how to reach cardiac coherence. A smaller screen on the table prevented them to see each other's Tobe. They had to imitate the breathing pattern given by a gauge going up and down in 5s cycles onto Tobe's body. The lights of the room were dimmed to facilitate a relaxation state and each participant was given headphones playing back rain sounds.

After the training session, the screen separating the two Tobes was removed. Participants were then instructed to repeat the same exercise as before, but without the help of the gauge. They could see their colleague's Tobe. However, there was no interaction between them at this stage – it served as a transition between a self-centered task and a collaboration task.

During the third session, participants were instructed to synchronize their *hearts*. In order to do so, they had to both reach cardiac coherence while breathing on the same rhythm – with no other way to communicate than using their Tobes.

After this final session, we gave questionnaires to participants and conducted informal interviews with them to gather feedback about their experience.

11.4.2.3 *Results & Discussion*

From the questionnaires, that took the form of 5-points Likert scales, participants reported that they were more relaxed after the end of the session: 4.36 on a scale ranging from 1 "much *less* relaxed" to 5 "much *more* relaxed" (SD=0.74). Beside the fact that Tobe acted as an effective biofeedback device, the experiment was also a chance to introduce participants to activities centered around well-being, as few of them were practicing relaxation or meditation in their daily life – 1.93 score (SD=1.44) with 1 "never" and 5 "regularly".

During the interviews, the participants reported that they appreciated the feedback, saying that it formed a coherent experience – e.g. ripples on the table and sounds of rain. Among the few that were practicing yoga regularly, one praised how Tobe favors learning-by-doing over wordy and disrupting instructions but had troubles to follow the 10s breathing cycle since it differed from his usual practice. We had mixed reviews about the visualization associated to breathing, mostly due to the mapping between Tobe's lungs and the measured thoracic circumference being dynamically adapted over time rather than calibrated per user with a min/max. Because of that, some users had to draw their attention away from the breathing patterns in order to achieve cardiac coherence. These two last issues could be resolved by giving users access to the signal processing through our toolkit.

We received comments about how a qualitative and ambient feedback (blooming flower) fostered a better focus on the activity compared to the use of quantitative metrics which are an incentive for competition. Indeed, apart from some comparisons made during the second session, participants did use their Tobes for collaboration. Users described how they use the respiration of their partner to get in sync during the third stage – usually by waiting before inhaling. One participant described how she tried to "help" her companion when he struggled to follow. Another retold how she quickly resumed her regular breathing when she saw that a brief hold troubled her colleague. More playful, a participant laughed afterwards at how he purposely "tricked" twice his partner. Even with a communication channel as basic as the display of thoracic circumference, rich interactions emerged between participants over a short period – 5 minutes that felt like less for many of them.

Overall these findings suggest that Tobe could be employed as a proxy for interpersonal communications and that it has an interesting potential for enhancing well-being.

11.5 Applications (design space)

We drew usages for Tobe by exploring different dimensions: on the one hand the number of users, Tobes and external observers involved, and, on the other hand the time and space separating the feedback and the recordings.

On a handful occasions our design space intersects with existing research projects or prototype, indicating that the framework built around Tobe is able to unite – and to extend – the emerging usages that come out of physiological computing.

11.5.1 One User

Tobe can be used as a biofeedback device with a specific goal – e.g. reduce stress – or to gain knowledge about one self. A feedback about workload and vigilance would prevent overwork, as to work *too* hard for too long results in efforts becoming counterproductive. A "cute" form-factor may be an incentive to take care of Tobe… hence of ourselves.

Insights gathered from an introspection session with Tobe could be employed not only to become better, but also to *act* better. Try to recall those times when you answered a bit too harshly to a beloved one because you were irritated by something completely unrelated and did not realize it... An angry-looking Tobe could have reminded you of your inner state before you answered harshly the wrong word.

11.5.2 One User and Observer(s)

Scientists involved in stroke rehabilitation research suggested that Tobe could be used in a medical context. Indeed, patients with motor disabilities may regain mobility after long and difficult sessions of reeducation. However, occasional drawbacks may create anxiety and a counterproductive attitude towards therapy, which leads to even more anxiety. A Tobe could help patients and therapists acknowledge this affective state and break this vicious circle.

Autistic persons could also benefit from using Tobe since it is difficult for them and their relatives to gauge their inner state. To make explicit arousal could help their integration into society, as envisioned in [Picard, 2009] and experimented offline – i.e. after signals were recorded – in [Hedman et al., 2012].

11.5.3 Multiple Users and Tobes

Using Tobe as an alternate communication channel during casual interactions would help to explore connections with relatives, blossom friendships, discover and learn from strangers or improve collaboration and efficiency with coworkers. This has been partially explored through the “Reflect Table”, which gives a feedback about the affective state of meeting participants [Bachour, 2010]; and a bicycle helmet that displays the heart rate of the wearer to the other cyclists nearby has been proposed to support social interactions during physical efforts [Walmink et al., 2014].

I’m most looking forward to these use – many users with one Tobe each – so as to increase social presence and facilitate social interactions.

11.5.4 Archetype of a Group

Tobe could summarize the state of a group. A real-time feedback from the audience would be a valuable tool for every speaker or performer – and for example Google Glasses could be used to monitor its affective state [Hernandez and Picard, 2014]. EDA has been used to analyze afterwards the emotional state of spectators while a company performed in [Wang et al., 2014].

To pace a course, a teacher could use one Tobe as an overall index that aggregates the attention level of every student in the classroom. Through behavioral measures and with a feedback given afterwards, this was investigated in [Raca and Dillenbourg, 2013].

11.5.5 Time and Space

One could want to analyze or to recall inner states after an event. Tobe could replay how one actually felt alongside a video of a cherished moment, a more vivid picture than plots.

If it is not time but space that separates a Tobe from its owner, imagine a distant relationship where the Tobe on your desk slowly awakens as the sun rises in the timezone of your beloved one – and you would wait for Tobe's vigilance to increase to a sufficient level before you pick up your phone for a chat, knowing that your soul mate is a bit grumpy at the beginning of the day. Besides this theoretical view, it has been hinted that even low-level physiological signals could enhance telepresence [Lee et al., 2014].

Finally, if both time and space are different, we could imagine a trail left during a journey as the "neuro tagging map" project from Neurowear[11].

11.6 Conclusion

We have presented an open system aimed at externalizing physiological signals and mental states in order to offer users a shared "out-of-body experience". This system covers the entire pipeline, from signals' acquisition to their visualization. Our framework being customizable and modular, it can adapt to the context where it is applied and to the many views people do have about physiological signals and mental states. Its open nature may be used to introduce STEM discipline to the general public through inquiry-based learning, while end usages can steer them to cognitive science, psychology and humanities, bridging the gap between "hard" and "soft" sciences. Even if the modules we chose promote the inclusion of novices – e.g. visual programming that could be easily extended in OpenViBE or vvvv, they can be switched to other components that would suit more experienced users – e.g. Matlab for signal processing. The system is not reduced to a set of tools, though, and we emphasized how such device is aimed at knowing better ourselves and others.

We put the focus on *one* implementation of the system that consists in a tangible puppet, Tobe, onto which signals are displayed. Its anthropomorphic shape eases users' identification, improves readability and enhances likability. We tested how Tobe affected positively social interactions in a 2-users scenario centered around a relaxation activity. Our co-design approach relied on two interventions that occurred in a scientific museum, as well as on surveys assessing how people relate to physiological signals and how they represent themselves various mental states.

[11] http://www.neurowear.com/projects_detail/neuro_tagging_map.html

We have identified design dimensions that we used to propose potential applications for our system. Supporting rehabilitation in medical care and facilitating the life in society of individuals with sensory challenges such as autism or ADHD – with the possibility to include the therapists in the loop – are use cases that could benefit from a friendly way to expose inner states. Moreover, it would be interesting to investigate how such a system could be used to ease social interaction and collaboration or to foster empathy towards others.

Should be also studied how the form factor of the proxy influences relationships, between a user and a Tobe from one hand, and between a Tobe and observers on the other hand. Tobe is an anthropomorphic avatar at the moment whereas more stereotypes and fantasies may be associated to animal figures, even the more when tangibility anchors the bond between the user and the proxy. There is already a literature surrounding digital environments that could point out to research directions on how we relate to avatars, e.g. [Jin, 2012].

Future work will include testing Tobe in classrooms or public workshops where users will be invited to build their own self-representation from the ground up, including the tangible support, sensors and the feedback design. Longer usages of the toolkit, over multiple days or weeks, will also be the opportunity to strengthen signal processing in order to provide more reliable mental states that could be displayed between users – e.g. BCI settings as in part II. Giving users the tools and manuals to investigate their own bodies and mind is a good way to empower them and prompt self-reflection.

CODA

Where we say a few words about the journey me made, sketch the next routes and stare at a crystal ball.

Conclusion

After 4 acts, this is the end.

Parts I and II can be summarized in one sentence: EEG may be used as a complementary method for the evaluation of human-computer interactions.

Indeed, it is always difficult to assess the benefits and downsides of a completely novel interaction technique; the quality of a tool heavily relies on the type of glasses you wear while looking at it, hence on the evaluation method. Imagine your are the chief executive officer of a big company caring about 3D contents. Despite the well-known discrepancy between the degrees of freedom of the input and output devices (typically 2 *vs* 6), your employees have been using for years the same tools – mouses and keyboards are not ideal but keep the work done. While now and then you hear about revolutionizing and jaw-dropping new devices, you would need more that nice words to get convinced and change your habits. In order to make such impacting decision you need proof, you need studies. Inquiries do not suffice you – you witnessed too many times your children getting excited for new toys, harassing you for months before Christmas, only for said toys to be put in the attic and get dust before new year's eve has passed. At home you endure this stoically each year because you love your little daemons no matter what, but damn you if at work you let your feelings cloud your judgment. If you have to push something new to your assembly line, you want your proofs, a taste of how the new device impacts users – both how they feel while using it and how it affects their work. You need something *objective* to chew on.

You should rather use the term "exocentric" instead, but we forgive you. More importantly, we *understand* you. The evaluation of human-computer interaction with physiological sensors responds to a common need by the high ranked deciders: to give them reliable ammunition so they could decide important stuff. We got that covered; we have shown how measures of brain activity based on EEG could estimate users' visual comfort with stereoscopic displays and workload during 3D manipulation tasks. We have also validated the use of EEG to assess at the same time workload, attention and error recognition; comparing two interaction techniques during a navigation task. Future work will have to replicate those results with other user interfaces and consider even more constructs. This way we will get closer to a continuous measure of user experience, which in turn will help to shape better interfaces.

Neuroergonomics benefits from the availability and affordability of the hardware; a solution based on the open hardware board OpenBCI is a good candidate to disseminate the use of EEG. During a preliminary study – see Appendices – we showed that the signals measured by this device are close to what could be recorded with a medical grade equipment. This lowers the barrier between BCI applications and the general public. In return, new usages may emerge, driven by end users that will finally have access to a reliable source to sense their brain activity. For example it did not require much to craft a headset that could account for a mental state such has workload. Cheaper and more practical sensors will not only improve what already exists, it will also unveil brand new applications, beyond human-computer interaction.

In part III we saw how heart rate could support playful interactions such as board games. Thanks to remote sensing, users did not have to wear any equipment to enjoy the proposed application. It was a "seat and play" setting; ideal to integrate seamlessly physiological computing to everyday life. Physiological sensors could intercede between any persons, whether or not they have a medical condition, even if they do not seek to improve in sport activities or if they are not interested in self empowerment, no matter if computers are in the environment or not. Sharing physiological signals, even as simple as cardiac activity, may help to increase social presence. We unveiled how the "physiological similarity-attraction" effect could foster empathy toward embodied agents (even poorly animated faces with a synthetic voice!). We built on these various premises to create tangible avatars that display physiological activities and mental states, one step further toward the removal of technological artifacts – computers are at their best when you forget they're even there.

We proposed a toolkit to let people explore freely their physiology and mental states in part IV. Not only do users' physiological signals reflects on Tobe, but high-level mental state can be selected and overtly displayed, giving insights about processes that are usually hard – if not impossible – to acknowledge. Teegi, which makes EEG signals *tangible* for the purpose of scientific outreach, can be assimilated as one implementation of the Tobe system. We envisioned multi-users scenarios where each user have a distinct Tobe. When we tested a relaxation exercise with 2-users, we saw the emergence of rich interactions that were based only on signals associated to breathing. This suggests that Tobe could be indeed employed as a proxy for interpersonal communications. Technology eventually stopped to *diminish* social presence, it is now on the verge to augment it with alternate communication channels.

Challenges and perspectives

Although several leads have already been suggested for future works all along the various chapters of this thesis, some broader challenges cover more than a single work. These long-term goals are split in four items. We will question how much sensors may disrupt what they measure, awe before the new wonders of the big data, get suspicious one moment about others' intentions but eventually remain hopeful for brighter social interactions.

Influence of the sensors

We shaped new scenarios and toyed with more practical devices, but no matter how lightweight they are, EEG and physiological sensors change the way people behave and users interact. Movements could be restrained by the devices and users could perceive a more stressful context, potentially biasing their experience.

Some studies inquired the acceptability and usability of EEG headsets from end-users perspective, e.g. [Nijboer et al., 2015]. More generally, we can wonder how many sensors users could handle before the user experience falters and which form factor is best suited. When we intervened in a scientific museum in chapter, 11 we tried our best to craft wearables that would not impede users' comfort. The informal interviews we had with the participants afterwards suggested that the latter were not bothered by our "instrumented" lab coat, but we intervened only over short periods of time and there was no proper evaluation.

We started to establish a protocol to investigate whether or not physiological sensors are a burden, and if so to which extent. During the work that took place along the "3D maze" project in chapter 7, we gave to our participants at mid-experiment a questionnaire inquiring about immersion. Indeed, in virtual reality one common metric to assess by how much the environment and the sensory modalities of the simulation favour immersion consists in measuring how much users are still aware of the *real* reality when they are interacting. The Immersive Experience Questionnaire (IEQ) that we gave possesses items that specifically inquire the perception of stimuli non congruent to the virtual environment. The more users perceive the external world, the less they are immersed. We hypothesized that an EEG cap or an experimental setup too disturbing or disruptive would translate into poor IEQ scores.

The objective was to make between-subjects comparisons, with a group enduring the regular experiment, another one playing the same levels (including calibration tasks) but with no EEG recordings and a last group with still the same protocol but additional physiological sensors on top of the EEG headset – EDA and ECG used during the evaluation of 3D tasks in chapter 6. Because of time constrains we could not proceed with the last two groups.

Hence, we only have IEQ scores from a group wearing the EEG headset, with no possible comparisons. For the record, we computed IEQ scores as described in [Jennett, 2010]. Those scores are calculated by summing up the items of the 7-point Likert scale questionnaires. Across our 12 participants – see chapter 7 for apparatus – the mean immersion score was 142.50 (SD: 23.77). This score can vary between 31 and 217 (7 $\times$ 31 items); the bigger it is the more immersed players feel. The "real world dissociation factor" score was 24.17 (SD: 8.00). This score is the most interesting factor for assessing sensors' influence; it represents how much players are both aware of their surroundings (e.g. external stimuli) and their "real life" (e.g. everyday concerns). It varies between 6 (totally aware of the real world) to 42 (totally forgetful of the real world).

Other scores means and SD: cognitive involvement: 53.33 (9.48); emotional involvement: 54.08 (8.95); challenge: 20.83 (3.38); control: 29.58 (7.18); single measure of immersion: 6.50 (2.35).

The scores we measured seem on par with the various immersive conditions that were tested in [Jennett, 2010] – where no physiological sensors were involved. This may suggest that even a tedious equipment such as a full EEG headset composed of 32 wet electrodes does not necessarily impede user experience, or at least not blatantly. Of course, even if the IEQ questionnaire may serve as a standard for measuring sensors' hindrance, the virtual environment that we utilized was different from the literature and many factors could have influenced positively or negatively the scores we report here.

Our group, or others, will have to carry on with these investigations and delimit the boundaries within which it is acceptable to use physiological sensors to account for users' mental states.

Real-time measures from the many

When dealing with physiological signals in general and brain recordings in particular, one should remain cautious about the interpretations of the results. Confounding factors may obfuscate the mental states that are inferred from brain activity [Gerjets et al., 2014]. This is why we ended up using a dedicated virtual environment in chapter 7, with difficulty levels that were validated beforehand through questionnaires. This way we confirmed that a task such as the N-back could be used to calibrate the monitoring of workload during richer interactions. We did so to compare interaction techniques, but retrospectively it also strengthens the results observed in chapter 6 about the strain that complex 3D manipulation tasks induce .

EEG can be added to the repertoire of HCI evaluation methods. It can be put into practice to assess the quality of a product – a new user interface or interaction technique – during its conception. Brain signals can assess more than one mental state at the same time, which is another advantage over inquiries beside continuous and egocentric measures. However, on the other hand, inquiries are qualitative measures that help to determine which precise aspect of the user experience is being evaluated. It would be interesting, then, to combine those two forms of evaluation methods within the same framework.

For instance, if real-time measures are made instead of offline analyses, we could imagine experts monitoring users while they are interacting, and asking them directly questions about what they are experiencing if the physiological signals prompt for further investigations. For example, an unexpected peak in workload or a sudden burst of interaction errors – which may look like artifacts otherwise – may be due to the fact that the user momentously lost sight of the task. We had somewhat of a similar situation with one of our participants that were playing with our “3D maze”. At some point the reactions of the participant within the game seemed completely random. It was only by talking to him directly that we could understand he had trouble with the instructions. Only for commodity did we not try to implement “real” passive BCI, with real-time measures, but with such tool at disposal a focus group could greatly enhance the overall quality of the evaluation.

Another great potential for HCI evaluation lies in the trend that goes along big data and cloud computing. With the 3D maze still, we have a version of the software that can be played directly within the browser. It is not a technological marvel, nowadays with a development platform alike Unity, it is a matter of a few clicks to deploy an application on the web. Had we pushed into this direction, we could have gathered an incredible amount of behavioral measures, way more than the 12 data points that we harvested over a week of experiments. In fact, the company behind Unity itself now advertises on the fact that developers could gather metrics about players’ habits thanks to their platform.

A true game changer would consist in benefiting from such approach with physiological computing. We mentioned how sensors embedded into wearables may one day become mainstream, but as of today, with devices that users possess, it is already possible to evaluate HCI through physiological signals and through the web – an opportunity to scale up *a little* the collected data. Indeed, appendix D shows that a simple video feed could be used to monitor cardiac activity, and we already saw in chapter 3 that from heart rate and heart rate variability mental states can be inferred – e.g. valence or workload. Knowing that a good proportion of the computers that are sold already possess a webcam – look-up at the bezel that surrounds the shiny screen of your laptop – there would be little left to accomplish before we could have physiological computing in the cloud.

Even the computing power required to synchronize and process such data may not be an issue. Indeed, a framework supporting cloud-based *BCI* – much more resource consuming than heart rate – is well described in [Zao et al., 2014]. There, the data streams were encapsulated in a real-time messaging protocol that ensured that no packets were lost during the transmission over Internet (MQTT) and off-site data centers were used to process signals. While this study involved few users, the company Affectiva[12] is currently deploying a toolkit called "Affdex" that lets content providers and advertisers measure through Internet the emotional state of their audience using facial recognition, showing that real-time processing of the many is doable.

We drew a particular usage for HCI evaluation, but the prospects of physiological *cloud* computing – or, if I may, "brain cloud interfaces" – go far beyond this scope. It would be an answer to the cumbersomeness of having the software and the knowledge to use BCI. You put on the device, request for an application, and that's it. Interaxon, the company behind the consumer-oriented MUSE EEG headset, is collecting and harvesting data from thousands of users across the world. Allegedly to do research and improve their measures. Collecting data across individuals do help to improve measures, for example in [Huang et al., 2014] various physiological signals from 250 participants were gathered to study emotions, but it is doubtful that private companies will open anytime soon their framework. A community-driven effort toward cloud-based EEG signal processing is actually emerging with CloudBrain[13] and may be better suited to support pipelines agnostic to applications.

It is hard to anticipate what outcome will emerged from big data applied to physiological states – or even if any of these solutions will really take off. More practicality? Better algorithms? Better understanding of our mental processes? New insights about the patterns that drive entire societies? One should stay cautious, though, and even prior to reaching those scales there are ethical concerns that surround physiological computing.

---

[12] `http://www.affectiva.com/`

[13] `https://github.com/marionleborgne/cloudbrain`

Ethics

We saw how physiological computing could benefit human-computer interaction and social presence (if not, I have something to worry about). However, although I am thrilled by the possibilities unveiled by physiological computing regarding casual interactions we, as scientists, should raise the awareness of users toward the signals that could be measured from their physiology, and help to prevent controversial uses. For instance, despite people being more and more concerned about their privacy, it seems that the general public still sees activity tracker as being inoffensive [Motti and Caine, 2015]. Yet, with smartwatches constantly connected to smartphones, there is little technological barriers against an application that would covertly data mine physiological reactions for profiling and advertisement. The question is even more problematic when remote sensing is involved. Physiological sensors should be a tool that give more control to them, not one that takes power away.

Hence, users' awareness and active participation should be encouraged. Such principle presided over our choices when we built Tobe. The use-cases we studied with Tobe ensure that users have the upper hand on what they share about themselves, for example they *choose* which signals are displayed on their puppet. I think that symmetry is an important factor for the acceptability of physiological sensors. In a given space, that everyone shares her or his Tobe ensures equal terms between users.

In the end, I do not insist on social interactions solely because I think that a tool such as Tobe could trigger a shift in what computer science has to offer to everyday people. I do so also because physiological computing brings applications and scenarios that are ethically debatable. I am not comfortable with usages that seek to create something alike a "lie detector" or that use body monitoring for advertisement and neuromarketing. Others will disagree, and they have the perfect right to push into these directions. However, I believe that Tobe, which displays in a constrained space and time what is inferred from physiological signals, is an opportunity to give users both awareness about what is measured and control over what they let out – signals are not recorded, not broadcasted, there is no intention to conceal some monitoring behind a one-time authorization.

For example, during the board-game scenario in chapter 9, even though the "bluffing" game mechanism was *de facto* an incentive for the players to probe heart rates and guess the real hand of their adversaries, this was one information among others concealed within the signal; they knew the outcome and made fun of it. They volunteered, and it was when the symmetry was broken – when players had no feedback about their own internal state – that stressful situation tented to arise.

Only with careful precautions about how and why physiological data is utilized could proxies alike Tobe become *social prostheses* that question and extend the boundaries of our self.

Tobeegi

Let's put aside the technological and practical aspects of physiological computing and focus instead on its applications.

In chapters 10 and 11 we saw how such signals could be put into practice to mediate people. The simple fact of reflecting on ourselves and know better our own body – i.e. improving interoception – could suffice to favour a better health [Farb et al., 2015]. One may benefit from a device that puts physiological activity into context. In [Rennert and Karapanos, 2013], users prone to social anxiety wear a mobile device that detects when heart rate increases to tag and geolocalize stressful events; "lifelogging" through the day helps them to acknowledge the causes of their turmoil and deal better with those situations.

With Tobe, people can build from scratch their own representation of their inner self. Thanks to the building block we developed along those last years – both hardware and software – it is now feasible for the general public to assemble and run by themselves a tangible avatar that is linked to their physiology, a "mini-me" of some sort. Guidance is still required, of course, but the process can well occur within a fablab; we have working prototypes that are based on off-the shelf components and we intend to document every step.

At the time I'm wrapping up my thesis, I have the chance to visit a laboratory in Montréal. Not that I'm particularly inclined to unveil what's going on in my life, but here I witness firsthand the emergence of a fast growing community that evolves around BCI technologies. Within a year, enthusiasts from all around the city gathered and started to meet frequently. They have tight links with local scientists from the field, they organize events to introduce the general public to the subject, they hack their way through the various devices at their disposal. Moreover, they attempt to build on the long term, spurring initiatives among students, looking after projects that could involve a variety of expertise. Now under the flag NeuroTechX[14] several branches all over the world are federating. CloudBrain is one realization that came from this working force.

I do not want to give away free advertising space in this section, but I wonder how far such community could go. Their approach, that seeks popularization and nurturing, is very close to the views we had been sharing within the Potioc team – in the end this thesis is but an emanation of a whole scientific group. There is one caveat, though: by far the majority of community has a background in engineering or computer sciences. I may had ventured into cognitive science, but I

[14] http://neurotechx.com/

have, too, the mark of the geek imprinted in my flesh. People that have no particular interest in gadgets and computers, or that cannot stand *obscure* and *stupid* machines – long way to go before seamless HCI – should not be left aside.

We framed how proxies displaying physiological activity and mental states can foster new kind of social interactions and augment social presence. Physiological computing is an opportunity to put back the focus on the human. Using physiology to mediate oneself is a long-term achievement that may or may not take place in a structure alike NeuroTechX, but that will definitely comes from the people.

First step on the road, In the near future I hope to investigate more of multi-users scenarios, using passive BCI settings to provide reliable measures; craft a “Tobeegi” of some sort that could present both low and high level signals to users. And eventually put it on a drone to make it fly, obviously.

Appendices

TOWARD PRACTICAL SENSORS

In the following appendices are compiled more technical aspects of my work as well as insights that I have yet to validate. Indeed, as an intermission between applications of physiological computing oriented from the one hand toward HCI evaluation and, from the other hand, toward social interactions, I took a closer look at the hardware that supports physiological sensing.

Here, concerning neuroimaging, we investigate how new EEG devices, more affordable and open, compete against traditional equipment. As for heart rate measures, we describe how remote sensing could be implemented with a simple webcam thanks to photoplethysmography. Finally, we draw some guidelines that could help to reduce the amount of artifacts by relying on non-obtrusive feedback about signals' quality.

A

Manifesto for an ideal EEG world

We mentioned in part I how new EEG devices appeared on the market in the last years, oriented toward a larger public, with a lower price tag and a more comfortable use. As opposed to "wet" electrodes employed for medical research, "dry" electrodes are faster to set-up (no more conductive gel) but are less sensitive – see [Blankertz et al., 2010], sec. 2.1. Hence, some companies, while trying to transform EEG into a mass-product, bring less reliable technology to the market. Those devices often posses fewer electrodes. Lastly, without a helmet the electrodes are difficult to place in a standardized position on the scalp.

While there are already many devices that are practical to use thanks to dry sensors and wireless connection to computer – a must-have for ambulatory use-cases – it is not always easy to verify their claims concerning their accuracy. The use of proprietary or close software also prevent users from freely choosing their workflow, e.g. one may have to switch to another operating system because of poor support from the manufacturer[1].

Having the possibility to place electrodes at specific locations on the scalp is essential, especially because it would require very expensive and cumbersome hardware in order to cover at once all the positions. Unfortunately, most of the EEG headsets that are appealing from a user perspective are an all-inclusive solution were we must comply with the choice of the manufacturer – that sometimes differ from standard

[1] any resemblance with situations where I had to run a computer on windows in order to feed a pipeline under Linux is completely assumed

positioning [David Hairston et al., 2014]. Often, as it is the case with motor imagery, it would be preferable to have 8 electrodes packed around a narrow area of interest (i.e. C3/Cz/C3) than 16 electrodes trying to cover the whole head. The poor spatial resolution of EEG could be solved in part by the versatility of electrodes position.

This is why we defined three requirements for an ideal EEG solution:

- Open-source driver in order to process freely and knowingly data (I strongly disagree with a business model that make pay 100 extra dollars just to provide a SDK for another *distribution* of linux. Don't temp me to give away the brand I'm thinking about.)

- Open-source hardware in order to acknowledge *beforehand* the reliability of measured signals. When the specification and the schematics of hardware are a known variable it is easier to assess the precision of the measures. Characteristics such as the signal-to-noise ratio of the amplifier or the resolution of the analog-to-digital converter (ADC) are critical. For example, even if you can acquire the electrical currents originating from the heart with a pair of clips – true story! – it would be delusional to measure anything with a 8bit ADC: since the system likely runs at 3.3V (or 5V), the resulting precision is $\frac{3.3V}{2^8} = 12mV$, way bigger than the few millivolts range of ECG signals. And if you plug an amplifier to boost the input signal, then it would be impossible to measure with the same component ECG and EEG signals, the latter being within the µV range, a thousand times less. Even though traditional manufacturers give the overall specifications of their amplifiers, the electronic circuitry is the result of complex interactions and in practice many components influence devices' performance [Usakli, 2010].

- Customizable electrode positioning, that is to said the possibility to place electrodes anywhere on the scalp. Moreover, this positioning should follow the 10-20 international system, or even better the 10-5 system (more refined separations along the sagittal reference curve), so that measures and findings could be compared together and with the existing scientific knowledge.

To which we add new items for a *practical* solution:

- Affordable, not only for every lab to be able to use them, or to make possible to use many different EEG headsets at the same time, but also – and maybe especially – for the technology to reach the general public. More often that we may realize science is driven by the usages, and even if scientists posses many qualities, the most important pool of creativity lies in everyday women and men. Give them a tool that could not hurt them, see how they handle it, what they want to do with it, and help them to do better.

- Lightweight, so it could be used in ambulatory settings. That implies a technology that is not a hindrance nor a health risk when it is used during a prolonged period. Little is studied in BCI with users that could use EEG during days or weeks – instead of hours and minutes. Again, new usages, and new processing (constant background auto-calibration?), could arise.

- Wireless; it seems obvious that EEG recordings with no strings attached is better for the freedom of movements, but is poses new constraints on extensive uses, for example energy consumption – a problem often occurring with "wearables", computers and devices that users could wear and use all day long.

- No preparation – having especially the *skin* preparation in mind. Because it is the one nightmare of someone coming in the lab to do a BCI experiment, and the one tedious step for the experimenter at the same time: fill electrode one by one with gel or other conducting solution, wait for good measures, sometimes check the impedance and correct accordingly the position. Such operation could last for dozens of minutes if there are many electrodes. As such, dry electrodes seem the obvious choice. At the same time as individual electrode setup, the positioning of the cap according to the 10-20 system should be made as fast as possible, e.g. with easy to handle markers that do not require to measure by hand the distance between the nasion and the inion or between the ears.

With all these observations in mind, we welcomed happily the appearance of the OpenBCI project on the participative funding plateforme Kickstarter during December 2013. We almost immediately jumped in and were among the firsts to received this new generation of amplifier, aimed at the many, once the campaign and the manufacturing process ended, by November 2014.

The OpenBCI project by itself was fulfilling many of the items on our list for achieving a perfect and practical EEG device. It's open-hardware, the firmware and the proposed software suite is open-source, it was nearly a *hundred* times cheaper than the hardware that we used in the lab, and both the form-factor design and the proposed 3D printable headset oversaw quite a handful of possible use.

There were few downsides, of course. First it has only 16 electrodes available, which is less than the 32 that we were used to. The more channels we have the better, since a better scalp coverage enable more accurate measures, but as mentioned at the beginning of this section, when the EEG markers are known, a good placement can overcome this difficulty. Second, the electronics such as the core chip, the Texas Instrument ADS1299, were suited on the paper for precise and reliable measure, but they needed to be tested. This is why it was necessary to compare it with our existing solution: After some pre-test (P300 speller and motor imagery, team's specialty) we conceived a protocol to conduct a proper evaluation of the system that we describe below in appendix B. In the appendix C that follows, we had the opportunity to craft our own headset, that we could attach to it.

B

Evaluating the OpenBCI board

To our knowledge, this preliminary study is the first which attempts to investigate formally the reliability of the OpenBCI board. In this chapter we compare side by side the OpenBCI to the g.tec g.USBamp amplifier. For this purpose, we employed an original montage, based on the simultaneous recording of the same set of electrodes. Two set of recordings were performed. During the first experiment a simple adapter with a direct connection between the amplifiers and the electrodes was used. Then, in a second experiment, we attempted to discard any possible interference that one amplifier could cause to the other by adding "ideal" diodes to the adapter.

Both spectral and temporal features were tested – the former with a N-back task, the latter with an P300 speller. Overall, the results suggest that the OpenBCI board could be an effective alternative to traditional EEG devices. Even though a medical grade equipment still outperforms the OpenBCI board, the latter gives very close EEG readings, resulting in practice in a classification accuracy that may be suitable for popular interactions. We conclude the chapter with several leads for further improvements of the open-hardware solution.

An summarized version of this work was published in [Frey, 2016a].

I thank here and now Thibault Laine for the technical support – more below in section Credits.4.

B.1 Introduction

We have shaped what form would take more practical EEG devices, that could be used in the field while remaining reliable enough to it could account for actual brain signals. The main component of an EEG device is its amplifier, the circuit board that will seek electrical currents few millionths volt and give comprehensive readings to the computer. It is a critical step in the chain. We highlighted a community driven and open hardware project that aroused recently, the OpenBCI board, based on Texas Instrument ADS1299 chips.

In this chapter we investigate how this board compares to medical grade equipment that is commonly used in research laboratories dealing with EEG. For instance, we compare side by side OpenBCI with the hardware used all along the works described in part II, the g.tec g.USBamp amplifier. The price tag of the g.tec solution is around 20 thousands euros, 25 times more expensive than the 800 euros of 16 channels version of the OpenBCI board. Besides the price, the g.USBAmp is also more bulky; bigger, heavier and does not run on batteries.

There are of course alternatives, even within g.tec products, that would be closer to what OpenBCI has to offer on the paper. For instance the g.USBamp does *not* aim at wearables. Still, one could wonder if there is a catch, if in practice, beside the specs, there are differences in using a solution alike OpenBCI. Here the question is not to assess which device is the best *per se*. Instead, we investigate if in a context of popular interactions – a narrow scope compared to the possibilities that offers the g.USBamp – it is conceivable for researchers from the field or (well equipped) enthusiasts to make the leap. To which extend should we employ devices coming from the DIY community for actual BCI applications?

To answers this question, we adopted an approach somewhat different to what exists in the literature. Many papers deal with the comparison of electrodes, e.g. wet *vs* dry. To do so, authors try to optimize the placement of both sets of sensors in order to get measures that originate from the same spots. However, no matter their efforts they could not merge sensors, and even clever montages, with electrodes of one sort positioned between electrodes of the other sort [Tautan et al., 2013], are not ideal. It will produce a slight offset, hence a slight inaccuracy. Another alternative is to make separate measures by repeating the recordings with each system [Nijboer et al., 2015], but once again the conditions could not be exactly the same.

In the present study we do not attempt to assess the quality of electrodes, but the behavior of amplifiers that are attached to them. Not a whole system, *only* the amplifiers. Therefore, we would not mind using the *same* electrodes during simultaneous recordings. This setup would ensure that the signal coming in each amplifier's inputs is exactly the same, avoiding any bias regarding the source of the measures.

We made that possible by crafting a dedicated adapter, one that basically splits in two the electrodes' wires. Such parallel measurement works because the amplifiers have high impedance circuits, that is to say that they are designed to not draw any amount of current from their source. As such, when one amplifier is connected, the readings of the other stay the same. Of course an infinite impedance cannot be achieved, and no matter the precautions this setup may cause a very slight difference compared to separate recordings. This is why in a second time we added to our adapter a circuit that prevents any interference between the two amplifiers, using ideal diodes to block current flows in one direction.

During the two experience described bellow, we covered the two types of EEG features that we employed during part II. A task monitoring workload aimed at assessing spectral information, and an oddball task sought temporal information. For each amplifier we measured the performance of a classifier based on those recordings, and additionally we compared both by correlating the signals that they recorded. No matter the financial aspects, the qualities of the g.USBamp amplifier make it the perfect baseline to gauge new challengers. This is also true for the electrodes developed by its manufacturer; in this study we are using g.tec wet and active electrodes.

B.2 First experiment: direct connections

B.2.1 Experimental setup

We acquired 16 EEG channels using the active g.Ladybird electrodes from g.tec. In this system, the electrodes are attached to a box that powers their electrical components and retrieves the signal; the g.GAMMAbox. After studying the wiring of the g.GAMMAbox, we designed a printed circuit board (PCB) to connect both amplifiers. Our adapter plugs on one end to the D-sub 26 connector of the g.GAMMAbox. Thanks to a pinout composed of 2.54mm connectors that gave access to all the channels (16 EEG + reference + ground), we attached the OpenBCI board to the adapter. On the other end of the adapter there was a D-sub 26 female connector, onto which we could plug the g.USBamp amplifier as if it were the regular end of the g.GAMMAbox. The schematics of the 2 layers PCB and a view of the setup are presented in Figure B.1.

Newer g.GAMMAboxes posses proprietary connectors instead of D-sub.

B

The EEG channels were positioned according to the 10-20 system at AFz, Fz, FCz, C3, C1, Cz, C2, C4, CPz, P3, Pz, P4, POz, O1, Oz and O2 – ground at FPz, reference on the left earlobe. Since the measures between both amplifiers were identical, only one recording session occurred, with one participant – there were no factors to counterbalance with repeated measures. The signals of both amplifier were acquired using OpenViBE 1.0.1; at 512Hz sampling rate for the g.USBamp and 125Hz sampling rate for the OpenBCI board.

As advised by the documentation, the ground channel was attached to the "BIAS" pin of the OpenBCI.

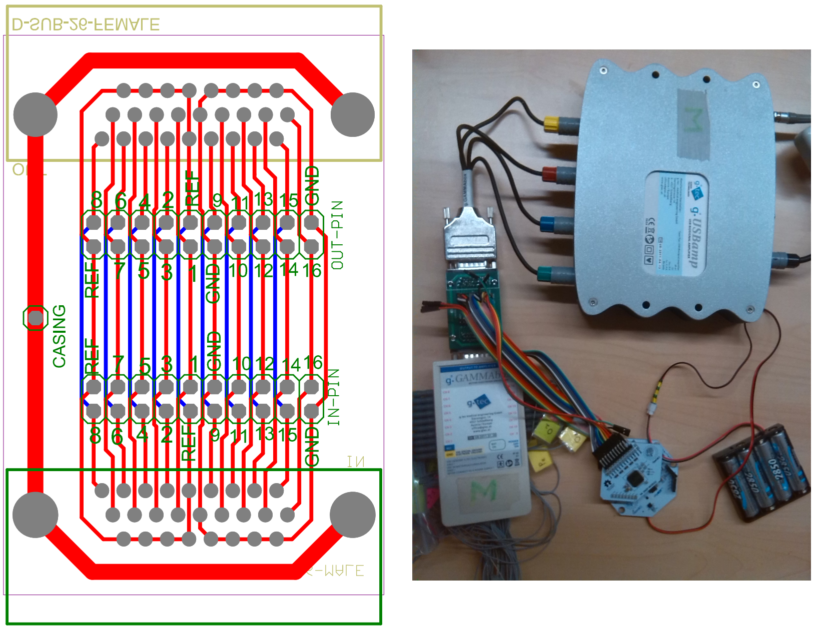

Figure B.1 – *Left*: schematics of the direct adapter. *Right*: Corresponding view of the setup.

The spectral features were investigated using the very same N-back task used to calibrate workload in part II. There were 360 trials presented during 6 blocks of alternate difficulty levels – see chapter 6 for details. The recording session lasted approximately 12 minutes.

The temporal features were investigated using an oddball task directly implemented within OpenViBE with a visual P300 speller (see chapter 4). During the recordings a matrix of 6 by 6 letters and digits was displayed in full screen on a 24-inch display. Only the calibration session occurred, during which one need to focus one's attention on a predefined sequence of letters. 32 letters composing a pangram were mentally "spelled" this way. The sentence was, without spaces, "pack my box with five dozen liquor jugs". Letters were flashing for 0.2s. There were 24 flashes per letter (12 times the row, 12 times the column), hence due to the matrix disposition there were in total 4608 trials, among which 768 were targets – "odd" trials, i.e. the letters of the target sentence were flashing. The recording session lasted approximately 30 minutes.

The acquisition of both amplifiers' signals and the P300 application occurred within the same OpenViBE scenario (script). The recordings of each amplifier were synchronized with the appropriate events and exported in separate GDF files for later analyses. There was also only one scenario involved in the synchronization of all signals and events in the case of the N-back task; stimulation from the python script supporting this latter task were retrieved using the LSL protocol.

B.2.2 Signal processing

Two kinds of analyses were performed. One aimed at assessing if and how the amplifiers differ in practice, when used for classification. The second then looked at the correlation between the acquired signals.

B.2.2.1 *Classification*

The signal processing of the data acquired during the N-back task is identical to what was employed in previous works, i.e. 2s time windows, 5 frequency bands – delta (1-3 Hz), theta (4-6 Hz), alpha (7-13 Hz), beta (14-25 Hz) and gamma (26-40 Hz) – and spatial filters. Because there was no point in transferring the calibration task to another context, we did not use the stationary subspace CSP introduced in chapter 6, but the "regular" common spatial patterns spatial filters to reduce the 16 channels to 6 "virtual" channels more discriminant between the workload conditions. Additionally, we also tested the 3 frequency bands version of our pipeline, that consider only the lower frequencies, less prone to muscular artifacts – delta, theta and alpha.

Concerning the oddball task, we kept the same signal processing as the one described in chapter 7 for interaction errors and audio probes. That is to say that we band-passed the signal between 0.5Hz and 40Hz, downsampled it by a factor 32 using the "decimate" Matlab function – by a factor 8 for OpenBCI because of the reduced sampling rate –, and applied a REFSF spatial filter to reduce channels' dimension from 16 to 5. We used 1s time windows after stimuli onsets – letters' flashes – to epoch ("slice") our signal. However, in order to prevent data to overlap between consecutive stimuli due to the rapid pace of the flashes, after a first pass of epoching we discarded overlapping time windows from further analyses. This ensured that no part of the signal could be seen twice by the classifier between the training phase and the testing phase and bias the accuracy. The procedure was automatic, the first non-overlapping epoch in order of appearance being kept. As a result, in the end we obtained 48 target trials and 240 distractor trials for classification, identical between the g.USBamp and the OpenBCI recordings.

The safeguard against overlapping time windows was in fact present during every single analysis described in this thesis, but only useful here.

Both for the workload and the P300 speller tasks, we used shrinkage LDA for classification and 4-fold cross-validation to assess the accuracy of the system, computing AUROCC scores (refer to chapter 7 for more details). In order to make statistical comparisons between both amplifiers for each type of features that we studied, we ran 10 times the analyses – the trials were selected randomly for cross-validation.

B.2.2.2 *Correlations*

We compared, on the one hand, the frequency spectra associated to the different workload conditions and, on the other hand, the time course

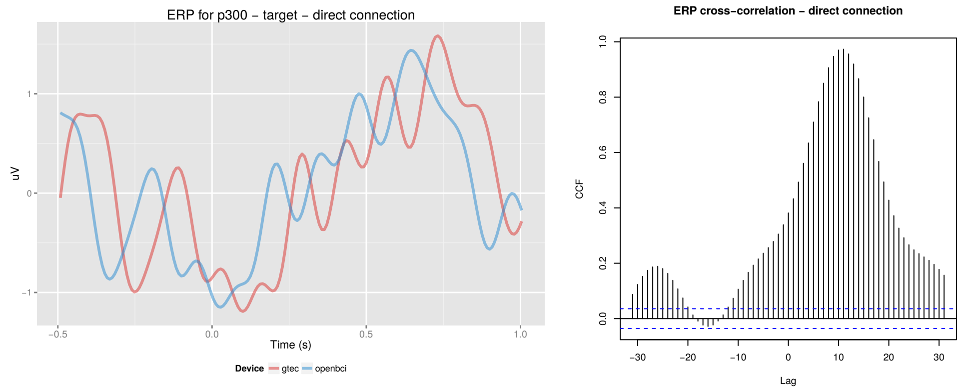

Figure B.2 – *Left*: Averaged ERP across channels of the target trials during the oddball task, before time shift correction. *Right*: Cross-correlation between the amplifiers. The computed lag of 11 data points corresponds to 88ms. (Direct connection.)

of the ERP that were caused by the flashing target letters. To do so, we used Pearson correlations, on par with the literature for similar analyses – e.g. [Zander et al., 2011]. In order to ensure a 1-to-1 correspondence between our sets of data, the recordings from the g.USBamp were downsampled to 125Hz – same sampling rate as for the OpenBCI – using the "resample" function from Matlab R2014a signal processing toolbox.

Concerning the workload task, we first aggregated the 2s time-windows corresponding to each condition (0-back and 2-back). Then we used the "spectopo" function of the EEGLAB toolbox (version 13.4.4b) to compute the grand average power spectral between 1Hz and 40Hz, for each channel. The output of the function was then passed on to R (version 3.0.2) to compute correlations through the "rcorr" function from the "Hmisc" package.

For the oddball task, we first band-passed the signals between 1Hz and 8Hz – the approximate frequency band used for classification. Then we extracted time epochs starting 0.5s prior to the flashing of the target letters and ending 1s after stimuli onset. Contrary to what occurred for classification, we did not prune overlapping epochs in the oddball task when we compute the averaged ERP – there was no bias that could have been induced here. Finally, we averaged the ERP per channel before exporting the time points to the R environment.

B

B.2.3 Results

B.2.3.1 *Classification*

The results regarding classification accuracy are presented in Table B.1, with the AUROCC scores for each one of the 10 repetitions, for both amplifiers and both tasks – including the 3 and 5 frequency bands pipeline for workload.

We tested for significance using Wilcoxon signed-rank tests. There was a significant difference between amplifiers for the P300 tasks ($p <$

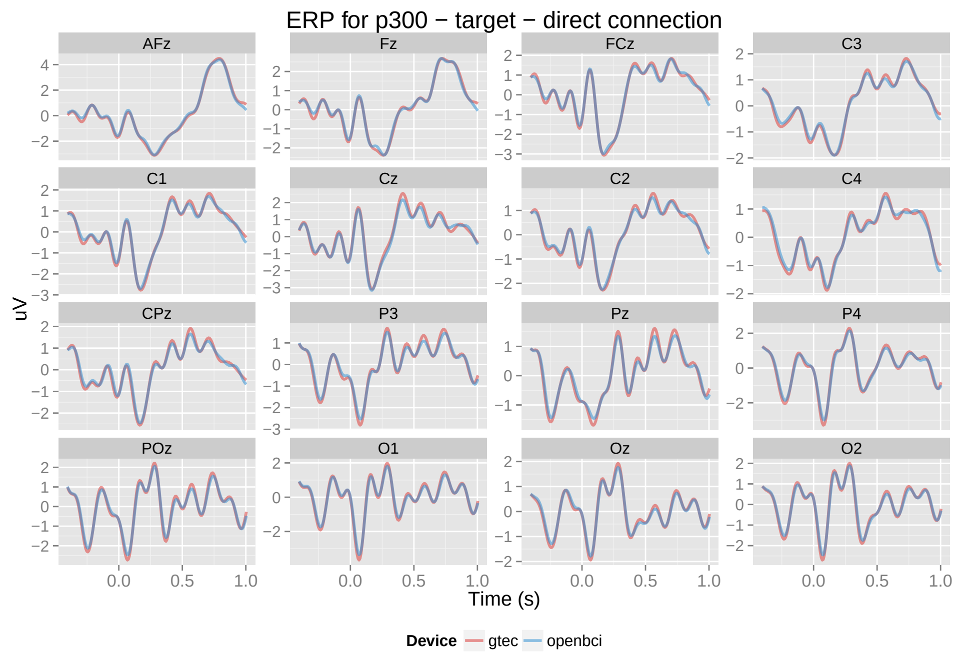

Figure B.3 – Averaged ERP for the target trials of the oddball task (direct connection).

0.01). The AUROCC mean score for the g.USBamp was 0.961 *vs* 0.918 for the OpenBCI. There were however no significance but tendencies concerning the workload task, with mean AUROCC scores between 0.85 and 0.86 for the 3 bands pipeline and between 0.89 and 0.90 for the 5 bands task – see Table B.1 for details.

B.2.3.2 *Correlations*

When we first analyzed our data to seek correlations regarding the oddball tasks, we realized that a shift occurred during the recordings, as denoted in Figure B.2 by the grand average of the ERP for target trials across channels. This may have been caused by a software issue (see Discussion). In order to correct the shift and conduct proper comparisons between both amplifiers' measures, we used a cross-correlation to estimate the time shift, using the "ccf" function from the R "stats" package. We found a delay of 88ms between the two signals – 11 data points at 125Hz, see Figure B.2.

In Figure B.3, the averaged ERP were shifted by as much for each channel. Corresponding Pearson correlation R scores, that were computed using the "rcorr" function, are presented in Table B.2. The mean R score is 0.9965 and is statistically significant ($p < 0.001$).

Truth is, rcorr returned a "0" p-value for all the correlations mentioned in this chapter.

There was also a significant correlation ($p < 0.001$) for the spectral features, with a mean R score of 0.9983 for the 0-back condition and 0.9979 the 2-back condition (see Table B.2 for details). Among the brain signals patterns that could be expected during the completion of a difficult task, the decrease nearby the alpha frequency band during the

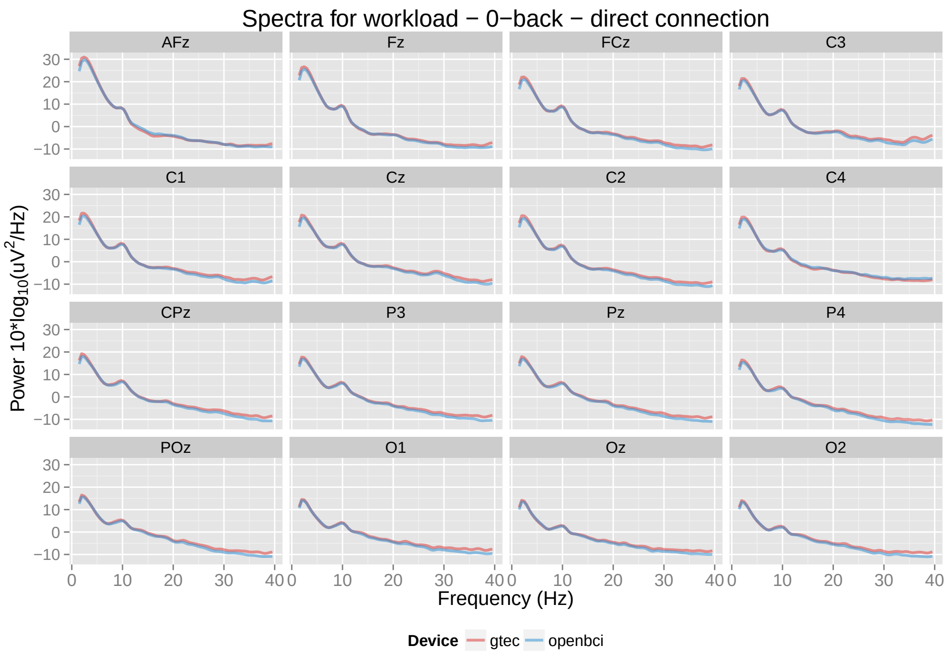

Figure B.4 – Averaged spectra for the 0-back trials of the N-bak task (direct connection).

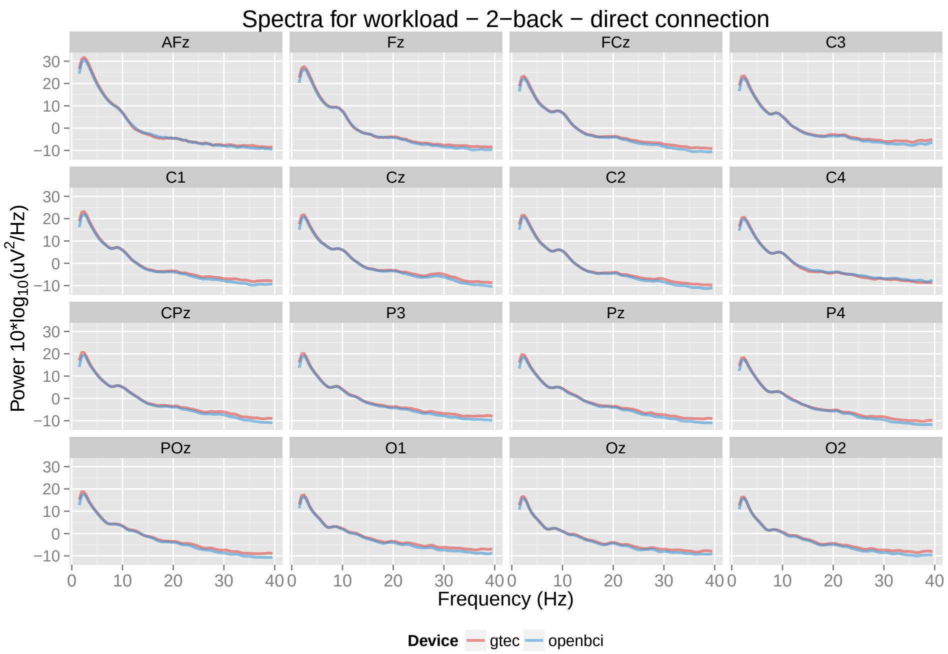

Figure B.5 – Averaged spectra for the 2-back trials of the N-bak task (direct connection).

B

2-task condition can be observed within per-channel spectra presented in Figures B.4 and B.5. Note that we did not correct time shifts prior to workload analyses due to the nature of the features – i.e. spectral and not temporal.

B.2.4 Discussion

The correlation between both temporal and spectral features tends so show that the signals acquired by the g.USBamp and the OpenBCI are, if not identical, very closely related. For every condition and channel tested, the Pearson R score was greater than 0.99.

There were however more dissimilarities in the classification accuracy obtained during the corresponding tasks. While there were hardly a difference between the AUROCC scores computed from both amplifiers with the N-back tasks, the g.USBamp performed significantly better than the OpenBCI during the P300 speller task. The time shift observed afterwards between the two amplifiers may partially explain this difference. Indeed, the detection of ERP is particularly sensitive to signals' latency, and a shift between events' timestamp and signal's acquisition could result in such degradation of performance when temporal features are involved.

The radio transmission between the wireless OpenBCI board and the dongle plugged to the computer may be one of the cause of the situation. The problem could also originate from the software. As a matter of fact, the OpenViBE acquisition driver of the OpenBCI board was released no so long before our experiment, and was still labelled as "unstable" as for version 1.0.1 of the software. One "oddity" that may further highlight the youth of OpenBCI software integration: we realized during our analysis that the recorded signals were completely inverted on the Y axis. The voltage reported by the board were the opposite of what g.USBamp was claiming. Since on numerous occasions we acknowledged the accuracy of g.tec devices readings, it is the OpenBCI's signals that we inverted back to "normal" prior to correlation analyses.

Beside time shifts issues, as mentioned during the introduction we needed to strengthen those first insights by discarding the eventuality that both EEG signals may have influence each other due to the direct wiring with the electrodes.

B

B.3 Second experiment: isolated connections

The second set of recordings is very similar to the what was described during the first study. The second experiment only differs by the nature of the adapter that was employed. As such we will only discuss the changes that were made to the hardware and quickly dive into the results.

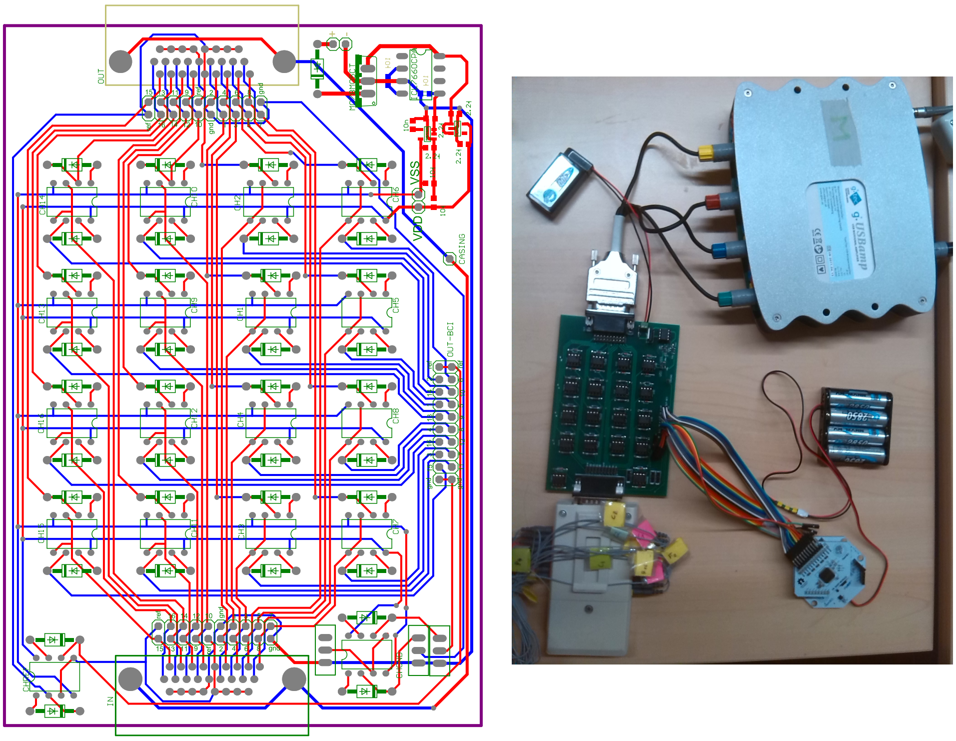

Figure B.6 – *Left*: schematics of the adapter with the ideal diodes montage. Note that there is a set of ideal diodes on the ground channel, but they were shorted with a jumper during our experiment. *Right*: Correspoding view of the setup.

B.3.1 Ideal adapter

We modified the adapter that connects the amplifiers to the g.GAMMAbox – and by extent to the EEG electrodes. Instead of a direct connection between each amplifier's inputs and the EEG channels, we interposed "ideal" (or "super") diodes on the branches of the "Y" wiring.

Diodes are electrical components that let the current flow in only one direction, the "forward" direction. Hence, this type of montage ensures that no current could travel directly from one amplifier to the other, contaminating the recordings. However, regular diodes cause a voltage drop. The voltage drop varies depending on the materials used for their construction, but it is at least 0.3V. Meaning that if the current coming in the forward direction is lesser than 0.3V, no signal will pass through. 0.3V is an order of magnitude superior to the range of EEG signal – $\approx$ a thousand time, therefore regular diode could not be used.

To circumvent this problem, we utilized a particular montage that involved operational amplifiers (op-amp). Op-amps are components widely used in electrical circuits, acting as sorts of "building blocks". Notably, in combination with a regular diode, one could use a precision rectifier configuration to obtain an "ideal" diode. This particular montage is also known as a "super" diode, since there will always be a

The regular diode is placed between the output and the negative input of the op-amp, i.e. on the feedback loop.

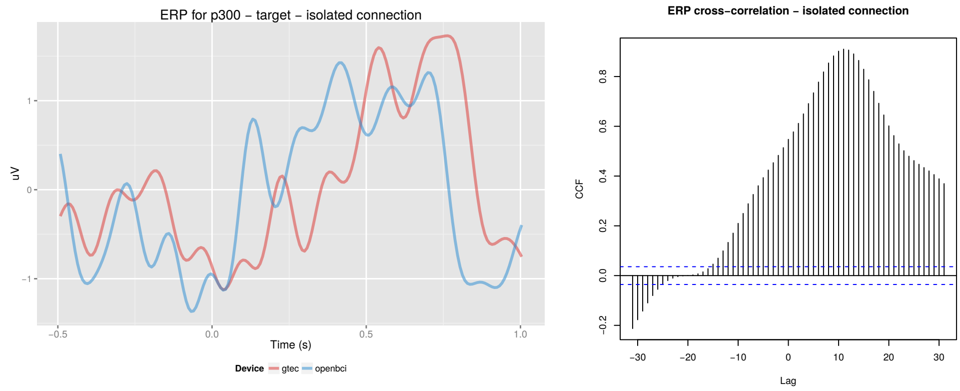

Figure B.7 – *Left*: Averaged ERP across channels of the target trials during the oddball task, before time shift correction. *Right*: Cross-correlation between the amplifiers. The computed lag of 11 data points corresponds to 88ms. (Isolated connection.)

slight voltage drop, but in this case, thanks to the gain of the op-amp, it becomes negligible.

We mounted 36 of such ideal diodes on the adapter. One on each end of the "Y" section associated to the 16 EEG channels, plus 2 for the reference. Due to the nature of the electrical recordings, only the ground was left without such circuit. We utilized Texas Instrument op-amps, model TLC2272ACPE4. The TLC227xA series are more indicated for precision application, and with 2 op-amps per chip we could limit the overall size of the adapter. The operational amplifiers were powered by an external circuit with regulated -2.5 / +2.5 voltage. The schematics and a view of the adapter – also a 2 layers 2 layers PCB – are presented in Figure B.6.

B.3.2 Results

The signal processing and the analyses were strictly identical to the first experiment detailed above, refer to the previous section for related information.

B.3.2.1 *Classification*

As with the first study, the results regarding classification accuracy are presented in Table B.3, with the AUROCC scores for each one of the 10 repetitions, for both amplifiers and both tasks – including the 3 and 5 frequency bands pipeline for workload. We tested for significance using Wilcoxon signed-rank tests. No matter the task there was no significant difference, although the 5% threshold was nearly reached for spectral features; the p-value was 0.051 for the 3 bands version of the workload pipeline.

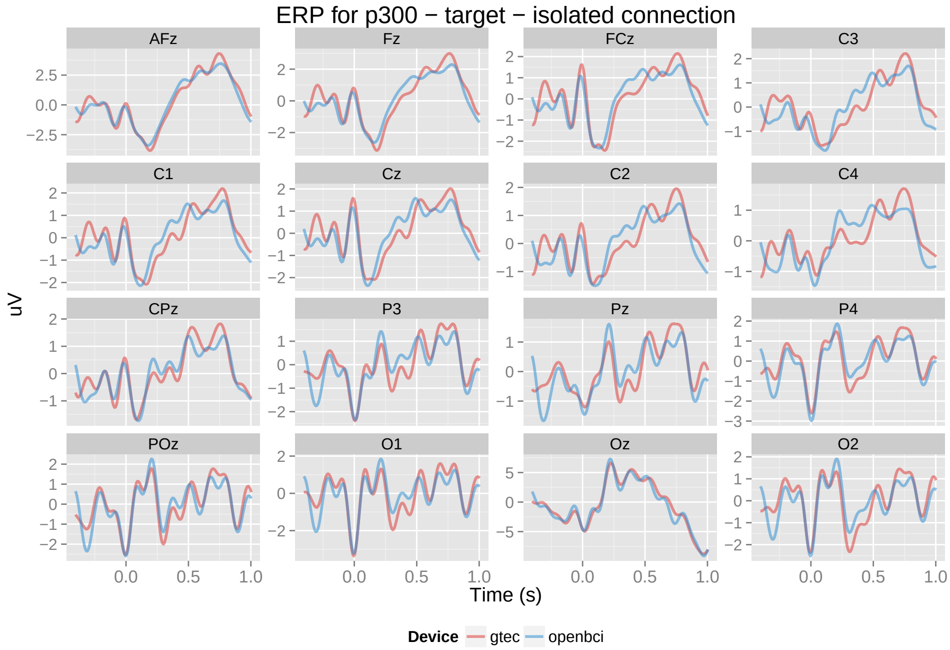

Figure B.8 – Averaged ERP for the target trials of the oddball task (isolated connection).

B.3.2.2 *Correlations*

Concerning the P300 oddball task, there was a offset of 88ms as well between the recordings of both amplifier with the isolated connection – see Figure B.7 for the grand ERP average and the cross correlation. The per-channel averaged ERP are plotted in figure B.8. Corresponding Pearson correlation R scores are presented in Table B.4. The mean R score is 0.8847 and is statistically significant ($p < 0.001$).

There was also a significant correlation ($p < 0.001$) for the spectral features, with a mean R score of 0.9976 for the 0-back condition and 0.9987 the 2-back condition (see Table B.4 for details). The per-channel spectra are presented in Figures B.9 and B.10. As with the direct connection, the band frequency changes between the 0-back and the 2-back conditions can be observed in the spectra.

B.3.3 Discussion

The results with the isolated connections are not that different from what was obtained during the first experiment. This would suggest that directly connecting two high impedance amplifiers to the same EEG electrodes could be a viable montage for a side-by-side comparison.

Since with both types of connector there was only one set of recordings, we could not draw any conclusion about the lower classification accuracy obtained with the isolated montage. The vigilance level of the participant alone could explain these performances.

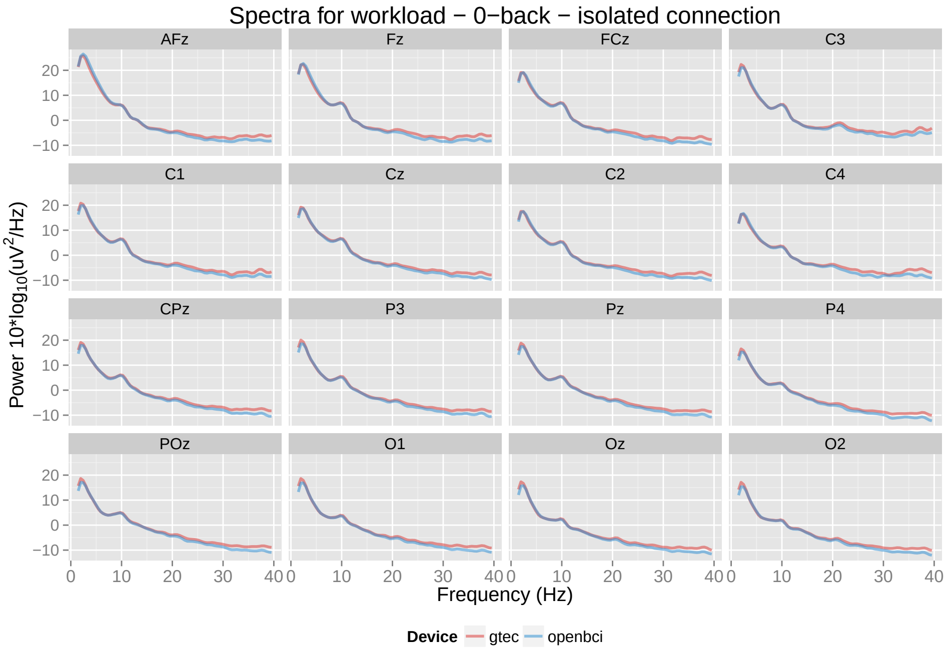

Figure B.9 – Averaged spectra for the 0-back trials of the N-bak task (isolated connection).

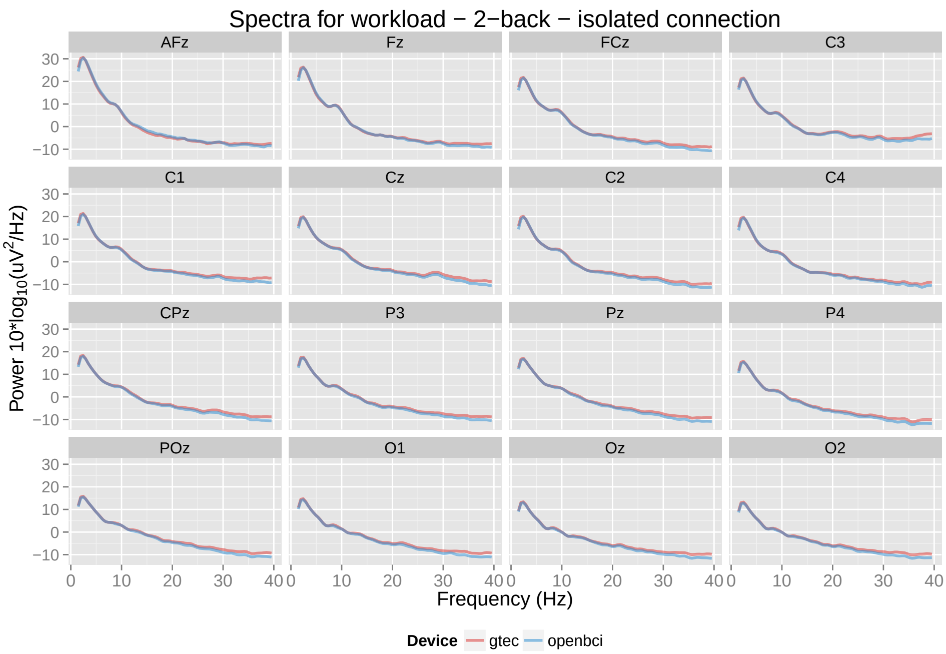

Figure B.10 – Averaged spectra for the 2-back trials of the N-bak task (isolated connection).

B

Thanks to signals' correlations, however, we may infer that noise was added to the system due to the presence of additional electrical components in the adapter. Indeed, while the spectra were once again strongly correlated, the averaged ERP achieved "only" a mean R score of 0.88. Here external factors such as the metal state of the participant or the quality of electrodes contacts could not have influenced one amplifier rather than the other. Since temporal features are more sensitive than spectral features to signal quality – e.g. one "peak" in the signal *vs* oscillatory patterns over several seconds –, it is instead more plausible that the difference with the first experiment comes from the adapter.

Nonetheless, even though the ideal diode montage did not produce ideal signals, those results still advocate for a close proximity between the g.USBamp and the OpenBCI. No device behaved "better" than the other, because no matter the lower correlation between averaged ERP, the classification accuracy is in practice comparable between both amplifiers. Each one probably endured different fluctuations since each had a dedicated set of ideal diodes.

B.4 Conclusion

During this preliminary study, we compared the OpenBCI board to the g.tec g.USBamp amplifier. We employed an original montage, based on the simultaneous recording of the same set of electrodes. While as a first approach we used a simple adapter with a direct connection between the amplifiers and the electrodes, in a second experiment we attempted to discard any possible interference that one amplifier could cause to the other.

To do so, we built an adapter that embedded "ideal" diodes, components that prevented electrical currents to flow "backward". This ensured that we could test both devices in isolation. We did not try to compare both adapters as the purpose was simply to gather more insights about the possibility of simultaneous recordings – this was a precaution to detect a possible bias.

Overall, the results strongly suggest that the OpenBCI board – or a similar solution – could indeed be an effective alternative to traditional EEG devices. Even though a medical grade equipment still outperforms the OpenBCI board, the latter gives very close EEG readings, resulting in practice in a classification accuracy that may be suitable for popular interactions.

Again, that is not to say that the OpenBCI could replace anyhow an equipment such as the g.USBamp. For example, the open-hardware initiative does not aim at medical applications, nor should it be employed in sensitive contexts. It does not possess any certification; one reason why so many cheap EEG devices are wireless is not for practicality, but to avoid any hazard due to power supply. Connecting somehow a body to the power grid requires extra precautions and a certified isolation, moreover when the impedance between the electrodes and the brain is intentionally lowered.

Beside the scope of application, we also stumbled on few issues with the current state of the OpenBCI project. One concerns the sampling rate of the board. While 125Hz may be enough for our use-cases – no frequencies beyond 40Hz are used during this thesis – it may not suffice others. The limitation of the sampling rate is caused by the wireless protocol used for data transmission. OpenBCI can deliver 250Hz signals to the computer, but only on 8 channels instead of 16. Note that this may be optimized in the future by updating the firmware or using alternate communications – as far as the board itself is concerned, the documentation of the ADS1299 claims a sampling rate up to 16,000Hz. The bandwidth is too limited to increase the sampling rate.

More problematic, there were unexpected behaviors regarding signals synchronization. However, here I have to make my *mea culpa*. Indeed, for the better or for the worse, I'm the one who implemented OpenBCI support within OpenViBE. Hence, I may be also the one to blame for any concerns regarding this particular software, even though I have been precocious and programmed safeguards against data loss. I will not try to dodge my own critics, but I also observed a difference of a fraction of hertz in the actual sampling rate of the OpenBCI board I tested, as well as packets that were lost here and then during recordings – another source of respectively drift and artifacts. Fortunately there are alternatives to OpenViBE and a wide variety of development kit available to acquire OpenBCI signals, not mentioning the board firmware that can bee freely reprogrammed. Consequently, there is room for improvement thanks to the open-source nature of the project, and is likely that the few issues raised regarding temporal features will be fixed in the future.

Now that we have demonstrated that an affordable device could give reliable measures, in the next chapter we take a closer look to the form factor aspect of the EEG device. At the same time we consider more practical and more accessible electrodes – the active electrodes employed in this chapter require a box twice as big as the OpenBCI board and costs twice its price. We investigate to which extent one could craft a dry headband with mostly off the self components.

Table B.1 – Classification accuracy (AUROCC scores) for the P300 and workload tasks studied during the first experiment – direct connection between the electrodes and the amplifiers. The 4-fold cross validations were repeated 10 times. Two pipelines are presented for the workload: 3 frequency bands ($\delta + \theta + \alpha$) as well as 5 frequency bands pipeline ($\delta + \theta + \alpha + \beta + \gamma$). Significance was tested using Wilcoxon signed-rank tests.

Condition	Amplifier	1	2	3	4	5	6	7	8	9	10	**mean**	**SD**	*p-value*
P300	g.USBamp	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.97	**0.961**	**0.003**	*< 0.01*
	OpenBCI	0.92	0.92	0.91	0.91	0.92	0.92	0.93	0.92	0.92	0.91	**0.918**	**0.006**	
Workload 3 bands	g.USBamp	0.85	0.85	0.85	0.86	0.85	0.87	0.87	0.87	0.87	0.86	**0.860**	**0.009**	*0.095*
	OpenBCI	0.86	0.86	0.85	0.86	0.84	0.86	0.85	0.85	0.85	0.85	**0.853**	**0.007**	
Workload 5 bands	g.USBamp	0.90	0.89	0.90	0.89	0.91	0.90	0.89	0.90	0.89	0.90	**0.897**	**0.007**	*0.079*
	OpenBCI	0.91	0.89	0.87	0.88	0.89	0.88	0.89	0.89	0.89	0.90	**0.889**	**0.011**	

B

Table B.2 – Pearson correlation R scores between g.USBamp and OpenBCI recordings at the 16 different electrode locations with a direct connection. The "P300 target" condition corresponds to temporal features (ERP averaged across trials) and the workload conditions to spectral features.

	AFz	*Fz*	*FCz*	*C3*	*C1*	*Cz*	*C2*	*C4*	*CPz*
P300 target	0.998	0.997	0.997	0.997	0.997	0.994	0.996	0.992	0.994
Workload 0-back	0.999	0.999	0.998	0.998	0.998	0.998	0.998	0.999	0.998
Workload 2-back	0.999	0.999	0.998	0.998	0.998	0.998	0.998	0.999	0.997
	P3	*Pz*	*P4*	*POz*	*O1*	*Oz*	*O2*	***Mean***	***SD***
P300 target	0.995	0.994	0.996	0.996	0.996	0.994	0.995	**0.9965**	**0.0015**
Workload 0-back	0.998	0.998	0.998	0.999	0.998	0.998	0.998	**0.9983**	**0.0003**
Workload 2-back	0.998	0.998	0.998	0.998	0.997	0.998	0.998	**0.9979**	**0.0005**

B

Table B.3 – Classification accuracy (AUROCC scores) for the P300 and workload tasks studied during the second experiment – isolated connection between the electrodes and the amplifiers. The 4-fold cross validations were repeated 10 times. Two pipelines are presented for the workload: 3 frequency bands ($\delta+\theta+\alpha$) as well as 5 frequency bands pipeline ($\delta+\theta+\alpha+\beta+\gamma$). Significance was tested using Wilcoxon signed-rank tests.

Condition	Amplifier	1	2	3	4	5	6	7	8	9	10	**mean**	**SD**	*p-value*
P300	g.USBamp	0.84	0.82	0.83	0.83	0.81	0.85	0.83	0.84	0.83	0.84	**0.832**	**0.011**	*0.157*
	OpenBCI	0.82	0.83	0.83	0.82	0.81	0.82	0.84	0.82	0.83	0.83	**0.825**	**0.008**	
Workload 3 bands	g.USBamp	0.88	0.88	0.88	0.91	0.89	0.90	0.89	0.88	0.89	0.88	**0.888**	**0.100**	*0.051*
	OpenBCI	0.88	0.88	0.89	0.89	0.88	0.89	0.87	0.88	0.88	0.87	**0.881**	**0.007**	
Workload 5 bands	g.USBamp	0.92	0.91	0.92	0.90	0.91	0.89	0.90	0.91	0.91	0.90	**0.907**	**0.009**	*0.286*
	OpenBCI	0.90	0.92	0.92	0.91	0.92	0.90	0.91	0.91	0.90	0.91	**0.910**	**0.008**	

B

Table B.4 – Pearson correlation R scores between g.USBamp and OpenBCI recordings at the 16 different electrode locations with an isolated connection. The "P300 target" condition corresponds to temporal features (ERP averaged across trials) and the workload conditions to spectral features.

	AFz	*Fz*	*FCz*	*C3*	*C1*	*Cz*	*C2*	*C4*	*CPz*
P300 target	0.976	0.934	0.892	0.846	0.872	0.881	0.838	0.811	0.912
Workload 0-back	0.999	0.998	0.998	0.997	0.998	0.998	0.998	0.996	0.998
Workload 2-back	0.999	0.999	0.999	0.998	0.998	0.999	0.999	0.999	0.999
	P3	*Pz*	*P4*	*POz*	*O1*	*Oz*	*O2*	***Mean***	***SD***
P300 target	0.849	0.823	0.896	0.910	0.878	0.982	0.853	**0.8847**	**0.0483**
Workload 0-back	0.998	0.998	0.998	0.997	0.997	0.997	0.997	**0.9976**	**0.0007**
Workload 2-back	0.999	0.999	0.999	0.999	0.999	0.998	0.998	**0.9987**	**0.0004**

B

C

Popular EEG headset

In this chapter we investigate the creation of an EEG headband which could be used during public exhibitions for a fast installation of an EEG device. Indeed, while we were developing our "Tobe" platform (chapter 11), we had the opportunity to bring EEG to a scientific museum and let everyday people grasp their brain activity. In the earlier iterations of the system we tested the use of an Emotiv EPOC headset. The EPOC was already easier to install than medical headsets that use gel. However, it still required a saline solution that tended to dry over time, causing additional installation time between users. Since the Tobe project aims also at popularizing physiological computing by letting people craft their own systems, we built an EEG device that suited our need and that could at the same time be reproduced by the general public within a fablab.

Amplifiers are usually agnostic to electrodes, and it is the case for the OpenBCI device. There are standard connectors, and as soon as an electrode could be connected with one pin and deliver a voltage, it is good to go. The standard kit on OpenBCI comprises passive electrodes made with gold. Compared to the electrode that we already possessed (g.tec active g.Ladybird), passive electrodes are more sensitive to noise (notably movements or magnetic disturbance). "Active" means that a pre-amplifier is located right next to the recording site, onto the electrode, so the signal that is transmitted through the cable to the main amplifier is stronger. Besides, with active electrode the skin impedance is less of a problem. The other handicap consists in the material: gold may have a better reputations for jewels, but for conductance the best metal is silver, more precisely a combination of silver and silver chloride (Ag/AgCl) [Tallgren et al., 2005].

Electrodes shipped with OpenBCI adopted a "cup" format and needed a paste to make contact with the scalp. This was maybe its only advantage over the g.tec active Ag/AgCl electrode, as Ten20 product which was delivered altogether is more reliable than other conductive solutions [Tallgren et al., 2005]. The passive/active antagonism was the most problematic to circumvent. Adapting existing solution to the OpenBCI board would have withdrawn one of its benefits, as this small and lightweight main amplifier would have required the addition of proprietary cases (i.e. g.GAMMAbox) to power the active electrodes and convert their connector to the more standard touchproof one. And building our own active circuits would have required much engineering and increased the coast of an affordable system, even if some paper describe elegant solutions [Degen and Jäckel, 2006, Chi et al., 2010].

Mixing materials could result in ions slowly drifting from one type of coating to the other because of differences in electrical charges between electrodes.

Fortunately, when we studied furthermore the differences between our two sets of electrodes, we realized that in this particular situation, and against many messages conveyed by manufacturers, passive electrodes are not dead yet. Indeed, the length of the cable and their dangling, two of the three causes of poor signal quality, did not really apply to an OpenBCI board that is supposed to be embedded in a headset – e.g. attached to the rear part. With the main amplifier at the back of the skull, the cables' length leading to the electrodes could be greatly reduced, and since all cables are tightly attached to the headset, their movements when users move is also reduced. And as for the third variable, skin impedance, it is often a requirement of the past, a reminiscence of a time when scientists were happily scratching their subjects' scalps for the sake of good signals. Excepts that since a few years now, the quality of EEG amplifiers have increased, and their (very) high input-impedance – that is to say their capacity to not disrupt the incoming electrical current while it is being measured – deals way better with scalp-electrode impedance [Ferree et al., 2001].

This is why we took this opportunity to build our own passive electrodes, that use Ag/AgCl and, because we stumbled on cheap electrodes that possess "fingers" to go through hair, we were also tempted to test a dry solution. The result is of course not perfect, either regarding the quality or the ease of use but we think we have been close enough for a thesis that is not labelled as electrical engineering – I am eagerly waiting for contact-free [Chi et al., 2010] or almost invisible [Norton et al., 2015] electrodes.

C.1 Electrodes

We used a set of Ag/AgCl electrodes (reference TDE-201) from Florida

These electrodes are also sold by Biopac, reference EL120

Research Instruments (FRI) in order to build our passive solution. These electrodes can be bought with no leads at a low price: less than 50 euros (shipping included) for 20 electrodes. Their coating is made in Ag/AgCl and they possess 12 pins that can go through the hair and reach the surface of the scalp with no skin preparation. For the areas not covered by hair – i.e. forehead – we used TDE-200 electrodes, that do not possess pins.

Figure C.1 – Various examples of commercially available dry electrodes. *Left*: TD-201 from Florida Research Instruments that we used in our headset ("EL120" in Biopac catalogue). *Middle*: Cognionics Flex. *Right*: g.tec g.SAHARA. Pictures from manufacturers. Note: scales mismatch between images.

The geometry of the TDE-201 is not the *best* suited for the conception of dry electrodes. In [Nathan and Jafari, 2014] authors studied how the number of pins influence hair penetration, the surface of the coating in contact with the scalp and at last EEG signal quality. They describe guidelines about the best pins density to facilitate contact. They conclude that an increase in a number of pins do not necessarily increase the quality of the signal. The effect could even be the opposite: when the number of pins is too important it could impede hair penetration. The TDE-201 do not posses a pins array too dense, however since the pins are located only on the outer ring of the electrode, their disposition may not be ideal. In case the dry contact is not enough – e.g. very long hair – a hole occupy the center part of the electrode and can be filled with gel if needed (see Figure C.1).

The length of the pins of the TDE-201 prevent its use when the hair is too furnished. The pins are 2mm long, much less than traditional dry electrodes. For example the g.SAHARA system from g.tec comes in two lengths: 7mm for the traditional electrodes and 16mm for the extended version that aims at people with more hair than the average, or people with a thick hairstyle. Finally, there is on the market novel components that enable the creation of flexible electrodes that could adapt to the pressure applied onto them and penetrate the hair even more effectively, such as the electrodes conceived by Cognionics[1].

All in all, that could make appear the FRI TDE-201 as a poor candidate for the conception of dry electrode. Indeed, the specifications of the

[1] http://www.cognionics.com/

manufacturer are unclear on the real purpose of those electrodes. Dry recordings may not be exactly their intended use case, even though it as been use as such in several studies in the past [Degen et al., 2007, Ye et al., 2014, Tautan et al., 2013].

The main reason why we chose the TDE-201 is because of their price and their availability, but they do possess a concrete advantage over the other candidates: the fact that there is a space located in the inner space of the electrode let the possibility for the experimenter to use conductive solution in situations that prevent direct skin contact or to improve signal quality. Other dry electrodes use a button-like contact with the cable lead. It facilitates the installation but prevent the addition of conductive solution once the electrodes are positioned.

FRI sells dedicated cable leads to be used with the EL120 but we constructed our own cable lead in order to have more flexibility using, regular wires.

C.2 Headset and cable leads

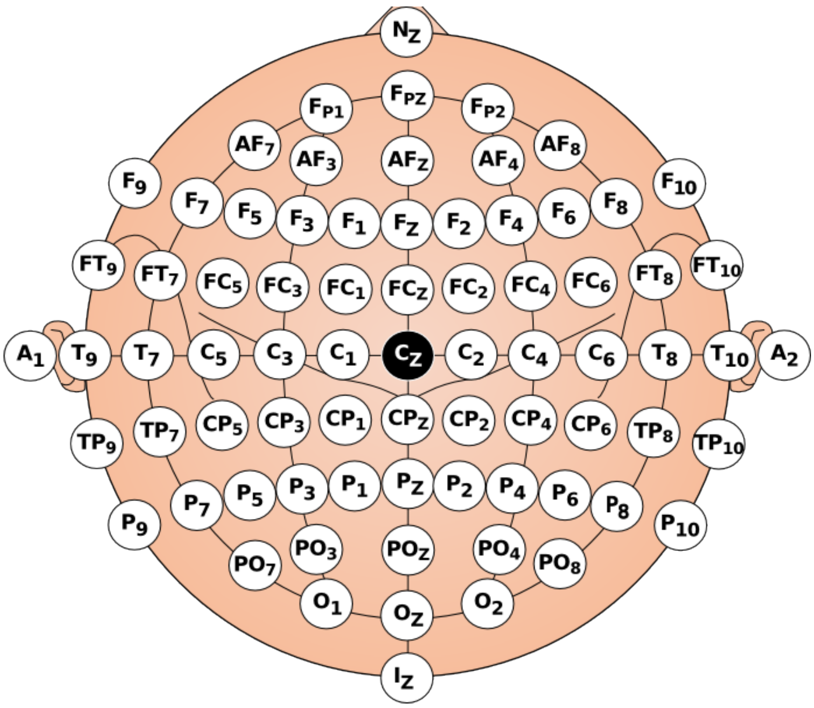

Figure C.2 – The 10-20 system used to standardized electrodes location. The head is assimilated as a sphere that is centered in Cz – half the distance between nasion and the inion in the sagittal axis and half the distance between the ears in the frontal axis. Electrodes are separated onto a that sphere by steps of 10 or 20 degrees. If you think this is the best picture ever of the 10-20 system, go fetch and re-use the SVG there: https://github.com/jfrey-xx/10-20_system_svg.

C

EEG electrodes that use conductive solution such as the Ten20 paste could be directly position onto the scalp. The viscosity of the paste can hold electrodes at the same time contact is made with the skin. However, part of the reason why cap are used with EEG – besides facilitating and securing the installation – is to make sure that the electrodes are positioned according to standardized locations, usually according to the 10-20 international system. Once the cap is installed along skull marks – Cz is positioned at half the distance between nasion and inion on one axis at half the distance between the ears, see Figure C.2 – then electrodes are locked onto standardized positions. Some manufacturers even propose to locate in space the exact position of each electrode using a tracking system similar to what is used to motion tracking or spatial augmented reality (see part IV to know more about the latter), e.g. Cephalon[2] and their "Xsensor 3D electrode digitizer".

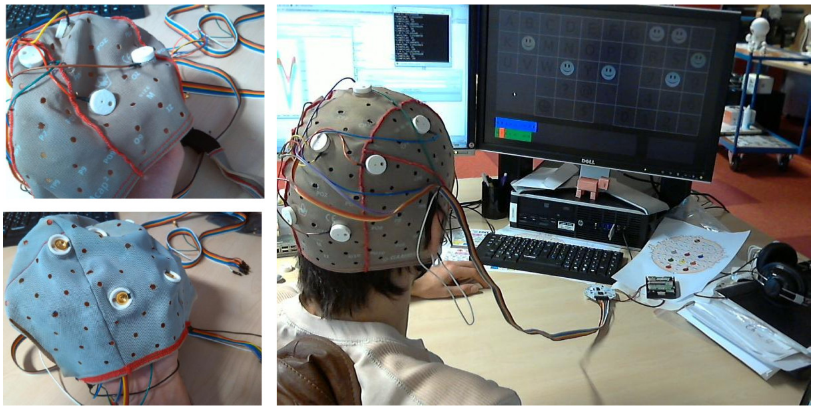

Figure C.3 – *Left*: 3D-printed holders used to mount regular gold cup electrodes on medical grade EEG cap. Our design. *Right*: the setup was tested successfully in combination with the OpenBCI amplifier with a P300 speller application (see http://blog.jfrey.info/2015/02/04/openbci-p300-coadapt/).

Regular EEG cap are made out of an elastic fabric. While we developed and 3D-printed custom holders in order to use 3rd parties electrodes with the g.tec g.GAMMAcap (Figure C.3), we chose to based our solution on a headband instead. That way we shorten even further the installation time of our dry electrodes. We did not need a high spatial resolution with our intended use case in the scientific museum, and restraining electrodes' locations to the rim of the scalp also avoided difficulties with long-haired people. Finally, we wanted to investigate alternatives to pricey EEG caps – they range in the hundreds euros.

Using a stretchable headband, electrodes were positioned at O1, P7, F7, FP1, F8, T8, P8 and O2 locations – reference at T7, ground at FP2. The headband was attached to users with a Velcro fastener and an 8 channels 32bit OpenBCI board was clipped with a dedicated 3D printed case at

We prevented the possibility to charge and use at the same the EEG board with a switch. You do not want your brain to be at one chip distance of a household outlet.

[2] http://www.cephalon.dk/

C

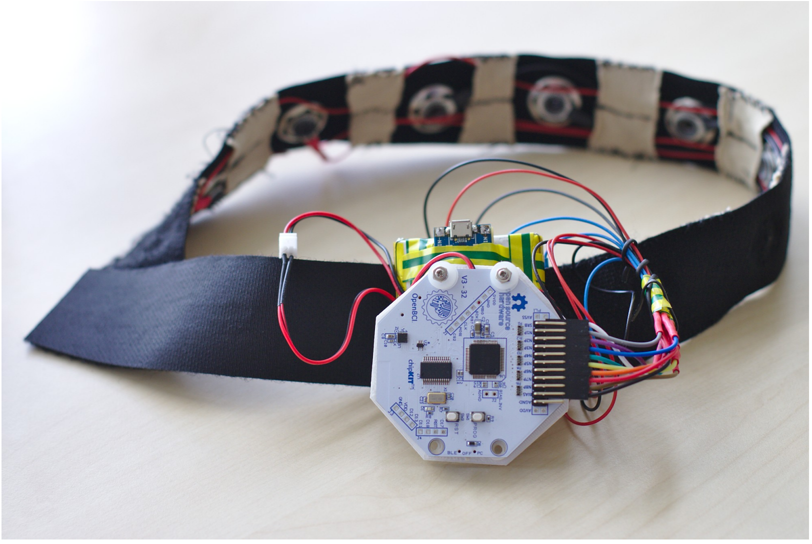

Figure C.4 – EEG headband created for public exhibitions.

the rear of it (Figure C.4). The case contained a 1400mAh lithium-ion polymer (LiPo) battery – light enough not to disturb participants and yet enough power to supply the system for hours. The battery was strapped to a charging circuit to facilitate maintenance.

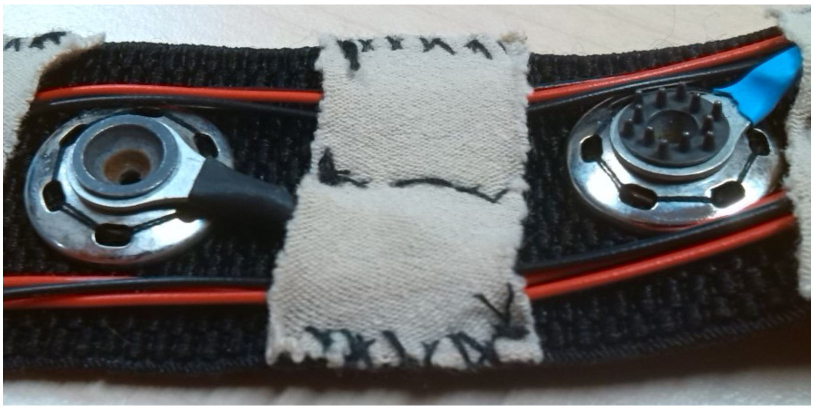

Figure C.5 – Close-up of the attach system. EL120 electrodes are snapped to a size 10 Dritz and contact with the cable leads is made through crimp rings. *Middle*: a man that knows his sewing. Or not.

Snaps that fit the size of the dry electrodes were sewed on the chosen 10-20 locations – we used nickel Dritz size 10, reference D80N-21 – so that the EL120 could be easily replaced should the Ag/AgCl coating wears off. We did not solder the cable lead directly on the sew-on snaps but on a tin plated crimp ring M2 stud size (Figure C.5). Even though the ring were squeezed between the electrode and the snap, there was a tiny wiggle room. Because of that additional noise was added to the system when users were moving, but we wanted to keep a modular system for this proof of concept. As a “failsafe”, we left tiny holes in the headband, aligned to the ones of the Dritz snaps and electrodes, so that gel could be added later on to improve readings.

C.3 Validation

In the previous chapter we validated the amplifier used in our dry headband, showing that in our experimental setup it could be compared to medical grade equipment. Yet we were using tried and tested active electrodes and a regular EEG cap. There is yet to see if the do-it-yourself solution that we describe here can record actual signals.

To sense so, we investigated how the headband behaves with spectral information, as those features were used in the field, in chapter 11.

We used once again the N-back task to induce workload. Because we did not try to compare the headband to another device – or use the trained classifier with unknown data as in part II – we setup a basic experiment. One participant was recorded during the N-back task. There were 6 blocks of 60 trials, blocks alternated between the 0-back and the 2-back tasks. We repeated the procedure twice in a row and conducted two separate analyses. The signal processing is analogous to what was described at length previously. In Table C.1 we present the results both for the 5 frequency bands version of our pipeline – delta (1-3 Hz), theta (4-6 Hz), alpha (7-13 Hz), beta (14-25 Hz) and gamma (26-40 Hz) – and the 3 frequency bands version that is less prone to muscular artifacts – delta, theta and alpha.

Table C.1 – Cross-validation accuracies (%) to descriminate workload levels (4 folds). Two runs, 360 trials per run – 160 for 0-back condition (easy) and 160 for 2-back condition (hard).

Frequency bands	Session 1	Session 2	Average
$\delta + \theta + \alpha$	65.56	71.11	68.33
$\delta + \theta + \alpha + \beta + \gamma$	64.72	71.39	68.06

While the average accuracy ($\approx$ 68%) is worse that what was obtained with a medical grade amplifier used in combination with 32 gel-based active electrodes placed all over the scalp – $\approx$ 84% in chapter 6) –, these scores are better than chance (according to [Müller-Putz et al., 2008]). This suggests that cheap, custom-made and quickly crafted alternatives could indeed retrieve useful information, which may be employed at least in non sensitive contexts, such as entertainment or for demonstration purposes.

C.4 Conclusion

In this section we documented the design of a cheap and lightweight EEG headset that could be installed in seconds and that we used during a public exhibition (see chapter 11). This proof of concept – very rough compared to other works more oriented toward research [Zander et al., 2011] – was the opportunity to explore the conception of a home-made equipment. While we did not try nor seek to obtain the cleanest EEG signals – the critical factor back then being the practicality of the system – we did manage to obtain a classification accuracy better than chance on spectral features. We can draw various directions to improve our prototype and broaden the possible applications of such device outside the laboratory.

We used a passive electrode design, more sensitive to noise than active electrodes but easier to craft and less cumbersome. Between those two solutions a third alternative exists: driven shields [Rich, 1983, Fraden, 2010]. With this technique, also known as active shielding, a shielded cable – such as RG-174 coaxial cable – is used to limit noise due electromagnetic disturbances and movements. The shielding is not simply grounded – otherwise it could produce *more* noise – but a guard circuit is used to "inject" back in the shield the signal measured from the electrode (inverting buffer). Various implementations has been successfully reported with physiological recordings in general and EEG in particular [Gargiulo et al., 2011, Usakli, 2010] – detailed schematics and PCB [Matsuzaka et al., 2012]. The takeaway message: the electronics is deported to the end of the cable lead, next to the amplifier – there are no components on the electrodes, as opposed to active solutions – hence headsets equipped with this system remain lightweight.

Concerning electrodes location, a 3D printed headset such as the one published along with OpenBCI[3] or the mesh described in [Giacometti and Diamond, 2013] could combine affordability and flexibility. Flexibility of the positioning – a parametric model could adapt to each morphology – but also of the material: nowadays flexible filament can be printed out regular 3D printers (e.g. Ninjaflex[4] worked well with our Makerbot Replicator 2[5]). With dual extruders – i.e. two "heads" that could print two filaments at the same time – it may be even possible to have solid parts for rigidity and flexible parts for comfort, both at once in the same piece. And why not use a *triple* extruder and conductive filament in order to print directly *electrodes*? At least alpha waves seem to be observable with first prototypes[6].

Alas, even though dry electrodes may match one day the reliability of wet ones, they are not as comfortable to wear because of the constant pressure needed onto the scalp to make contact [Chen et al., 2014]. Even the flexible polymer described by this latter paper leaves marks on the skin after a prolonged period. At least dry electrodes, well, don't dry.

Reduce signal noise, improve practicality, make longer recordings: EEG may be an old technique – nearly a 100 years [Haas, 2003] – but it

[3] https://github.com/OpenBCI/Ultracortex

[4] http://www.ninjaflex3d.com/

[5] http://www.makerbot.com

[6] http://conorrussomanno.com/2015/02/16/3d-printed-eeg-electrodes/

is still in the making. New devices appear regularly, with some that could even be used while running [Reis et al., 2014]. The future of the hardware is still ahead, however one should not loose sight that for a technology to become mainstream, there is more than good numbers on the paper. There is more at stake than good-looking devices – although end users' preference do matter [David Hairston et al., 2014] – or even comfortable headsets – oddly, spiky electrodes do not please every one [Nijboer et al., 2015]. For a technology to spread, it has to fulfill a need. Such as bettering everyday life by strengthening the social fabric, bringing people closer. Just a random thought, nothing that could be possibly be done in the scope of this manuscript. Part IV, anyone?

D

Heart rate through video feed

Over the course of my thesis I had the opportunity setup an application that let people interact with their physiology during a casual interaction, for instance a card board game. The idea was to let players play as usual… except for their heart rate, that would be displayed live within the gaming area. I could have decided to use some of our physiological sensors to record hearts' activity, but since I wanted almost no setup time and no intrusive recordings – not counting the fact that I wanted to have at least 3 players – there really was no regular technology at our disposal that could fulfill my requirements. Even though I tested ECG recordings of several users with one OpenBCI board (shared ground and a "bipolar" montage that could accommodate up to 16 persons with our model), skin contact was mandatory and I was afraid this sole setup, even if restrained to the wrists, would have disrupted the interaction. Indeed, physiology was not the focus of the interaction here, it was really the game.

This is why I decided to have a try at *remote* heart rate monitoring, using video feeds. A covert solution, scalable and that could be easily deployed since the required hardware is already widespread. Although the card board game, as an example of social interaction that could be augmented with physiology, is described in part III, the implementa-

D

tion of such remote sensing is a fit for a part that aims at diffusing physiological computing through tools.

This chapter describes the implemented pipeline, starting from webcam acquisition and stopping at heart rate feedback. The end-use described in chapter 9 being actually under review, no source code is available yet. The entire pipeline – summarized by Figure D.1 – will eventually be released as an opensource software.

I thank Hereiti Hatitio, Anta Mbaye, Jean-Baptiste Rey and Maxime Vincent for their help – a few more words in section Credits.5.

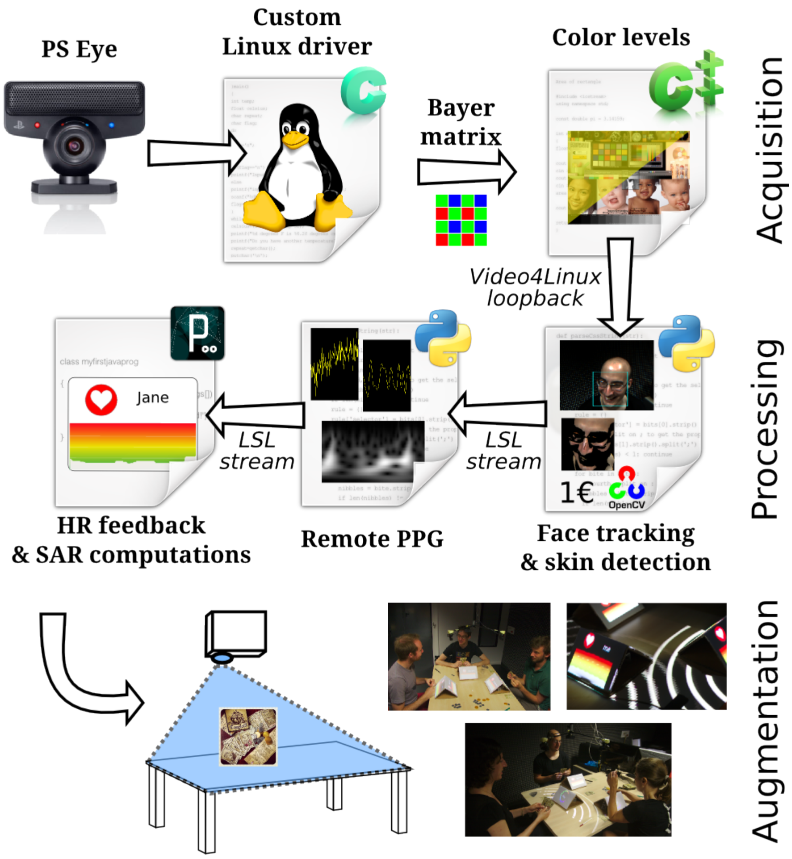

Figure D.1 – The pipeline composing the system measuring remotely heart rate. While signal acquisition and processing is detailed appendix D, the augmentation and use-case are recounted in chapter 9.

D.1 Heart rate measures

The eulerian video magnification method [Wu et al., 2012] had a significant impact within the computer vision community. It demonstrated how subtle color changes in a video could be amplified to the point that the variations of skin pigmentation occurring along each heartbeat became visible. This method works almost at the pixel level, but processing the average color of the entire skin suffices to compute a heart rate. An algorithm that takes as input values averaged from a

D

region of interest (ROI) is also less computationally demanding – the processing described in [Wu et al., 2012] and in the follow-up papers [Wadhwa et al., 2014] would be difficult to implement in real time, even more if 3 players should be recorded simultaneously.

The optical measuring of the volumetric variation of an organ – such as the heart – is dubbed as photoplethysmography (PPG). Several methods have been described in the past years that use PPG and a video feed to monitor specifically heart rate activity [Kranjec et al., 2014]. We chose to implement the PPG algorithm described in [Bousefsaf et al., 2013] because it uses a regular webcam and combines good accuracy and simplicity.

Instead of the green channel – often employed in PPG because of the green photosites being twice as numerous in Bayer arrangement, it uses the "u" channel of the Luv color space as a basis for its signal processing. This channel has a noise level nearly as low as the green channel of the RGB color space, is less sensible to motion artifacts, and contains color wavelengths matching those of hemoglobin absorption [Bousefsaf et al., 2013].

We tuned some of the signal processing steps described by Bousefsaf and al. [Bousefsaf et al., 2013] to improve the performance of our system. Notably, to speedup the filtering we detrended the signal and used a 5th order Butterworth band-pass filter between 0.6 and 4Hz. Then we kept the continuous wavelet transform (CWT), using Morlet wavelet, to make the spectral analysis and extract mean heart rates from 10 seconds sliding windows – compared to Fourier transform, there is no trade-off between the size of the window and the frequency resolution with CWT. We managed to obtain at least 10Hz in heart rate measures for a 3 players setup.

D.2 Video capture

We used a set of 3 PlayStation Eyes to record the faces of the players – they were hanging from the ceiling, slightly above the head. These webcams are cheap – around 10 dollars the unit – and yet provide satisfactory video quality. With PPG, color accuracy and consistency is more important than resolution; the algorithm we based our work on used 320 by 240 pixels images at 30 FPS. We set the PlayStation Eye to an even better video mode, with a 640x480 resolution. Note that the PlayStation Eye is capable to achieve 60 FPS, but we had no use of such high framerate – the noise would have increased and the computations would have been more important, with no benefits for players' heart rates that could hardly go faster 2Hz (120 BPM).

Another reason to use the PlaySation Eye lies in the availability of its electrical components' datasheets. We managed to modify its Linux driver (kernel 3.13) in order to access directly to the raw images of the webcam, i.e. to the Bayer matrix, before any demosaicing occurred.

With the raw images, we by-passed most of the automatic corrections that could parasite PPG recordings and that were previously performed by the hardware, such as brightness, contrast or white balance. Only the exposition and the gain remained enabled. Exposition was set to its maximum and the gain to its minimum to keep sensor's noise to its lowest. Without on-chip color correction, we relied on software algorithms upon which we had total control – e.g. accurate skin tones, no fluctuations of the color once the video feed was initiated.

As for most of the image processing undergone in this work, we used OpenCV [Bradski, 2000] to convert the raw images from the Bayer matrix to RGB colorspace – in our situation we did not need any advanced demosaicing that would preserve edges sharpening or prevent moiré artifacts. The white balance and the color levels were adjusted once, before the start of the PPG recordings, using "Simplest Color Balance" algorithm with 1% clipping [Limare et al., 2011] and a test image picturing purposely various faces and color scales[[1]. We performed colors adjustment in a 32 bits color space to avoid information loss.

Processing raw images ensured that we got the best possible video quality out of the PlayStation Eye, but our pipeline is not limited to this webcam. Indeed, once decoded and adjusted by our program, the raw images were fed to the Linux video API (V4L2) through v4l2loopback[2], so that the rest of our processing could accommodate any other peripherals, as we demonstrate during the validation study.

D.3 Face tracking and skin detection

For each one of the video feed, the OpenCV implementation of Haar Feature-based Cascade Classifiers [Viola and Jones, 2001] extracted the position of users' faces. To keep-up with the 30 FPS of the video, the Haar classifiers were computed once every 5 frames – the players were seated and their heads had a limited motion speed. In the use-case scenario of this multi-users remote PPG system, payers interact with each others. And not only would it be illusory to hope for them to remain perfectly still, but it would degrade significantly their game experience if they were instructed to do so. The measures have to adapt to ecological settings, and in order to stabilize the ROIs detected by the classifiers and reduce PPG artifacts, we filtered the face tracking with the 1€ filter [Casiez et al., 2012] (mincutoff = 0.01, β = 0.01, dcutoff = 1.0). If the face of a player was not detected at a particular frame, the last known position was used.

A threshold in the YC_bC_r color space [Mahmoud, 2008, Bousefsaf et al., 2013] was used to create a mask from the square containing the face – see Equation D.1. For skin types beyond 4 on Fitzpatrick scale

[1] http://www.gballard.net/photoshop/pdi_download/

[2] https://github.com/umlaeute/v4l2loopback

D

[Fitzpatrick, 1988] – i.e. the darkest tones – we adjusted the formula with $Y > 0$.

$$\begin{cases} Y > 80 \\ 77 < C_b < 127 \\ 133 < C_r < 173 \end{cases} \tag{D.1}$$

Once we got the mask, we were able to compute the average color of the detected skin and to process it with the PPG algorithm described previously. The communication between the module handling skin detection and the module detecting the heart rate was made through the Lab Streaming Layer (LSL) network protocol. The LSL protocol ensures a good synchronization and avoids data loss. It was also used to link the heart rate module to the display module.

In a playful interaction, as the final test case described in chapter 9, the practicality of the sensors takes precedence. We did not equip our board game players with contact sensors in order to assess the reliability of the resulting system. It would have disrupted the whole interaction. Instead, we conducted a separate study to assess the quality of our implementation.

D.4 Validation

We compared the measures produced by our implementation of remote PPG to a ground truth obtained with an electrocardiogram (ECG). No only did we assess the correlation between PPG and ECG, but we also took this opportunity to test on-the-fly other consumer-oriented webcams beside the PlayStation Eye.

D.4.1 Apparatus

The study involved one participant whose physiological activity was recorded over the course of three sessions – skin type 3 on Fitzpatrick scale [Fitzpatrick, 1988]. Each recording session lasted 10 minutes and was preceded by 5 minutes of aerobic exercise. This way the heart activity was expected to vary in two ways: the average heart rate would decrease down to a basal state after the activity stopped, and the instantaneous heart would vary according to the breathing patterns – heart-rate tends to decrease when one breathes in and decrease when one breathes out.

ECG was recorded at 512Hz with a g.tec g.USBAmp amplifier. Three passive Ag/AgCl electrodes with conductive gel applied on them were taped to the torso. A three leads montage was used, ground placed on the lower part at the left arm side, main channel upper part same side, reference placed on the upper part at the right arm side.

During the recording sessions, the participant was seated about 30cm away from the various cameras, that were attached together

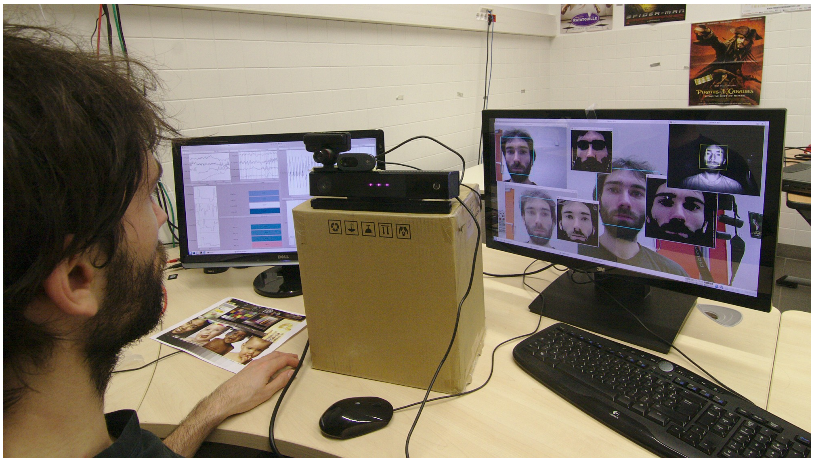

Figure D.2 – Experimental setup used during the experiment. Altogether with recordings from ECG electrodes placed on the torso, 3 webcams were recording subtle changes in skin color.

at level with the face – see Figure D.2. There were 3 cameras: the PlayStation Eye, a Logitech C270 and a Kinect2. The experiment took place in a room well lit – fluorescent light from the ceiling that was somewhat diffused by the white walls and floor.

D.4.2 Signal processing

The video feed of the Logitech C270 was set to a resolution of 640x480 pixels. All parameters were set to default – including automatic corrections, such as white balance or exposition, that were left activated. The Kinect 2 possesses two video feeds; a high resolution RGB feed (1910x1080 pixels) and an infra-red (IR) feed of 512×424 pixels. The acquisition of the cameras as well as the signal processing were performed on an Alienware Aurora R4 running Kubuntu 14.04 operating system. To retrieve the video feeds of the Kinect 2 under linux, the libfreenect2 library[3] was used in combination with v4l2loopback. The video feed of the PlayStation Eye was computed from the raw Bayer matrix with a resolution of 640x480 pixels, as previously described.

All four video feeds were acquired at 30 FPS. In case of RGB images, we used the same feature extraction method to retrieve PPG signals – face tracking, skin pixels, "u" channel of the Luv color space. However, because the Kinect 2 IR video feed is monochromatic, in this latter case we could not use the same techniques to extract skin pixels. Instead, we restrained the region of interest detected by the face tracking algorithm to a square half as wide and two-thirds as tall – an approximation to discard pixels not related to the face (see Figure D.2). The average of this region was passed on to the next stage of the signal processing, as with the other three video feeds.

[3] https://github.com/OpenKinect/libfreenect2

D

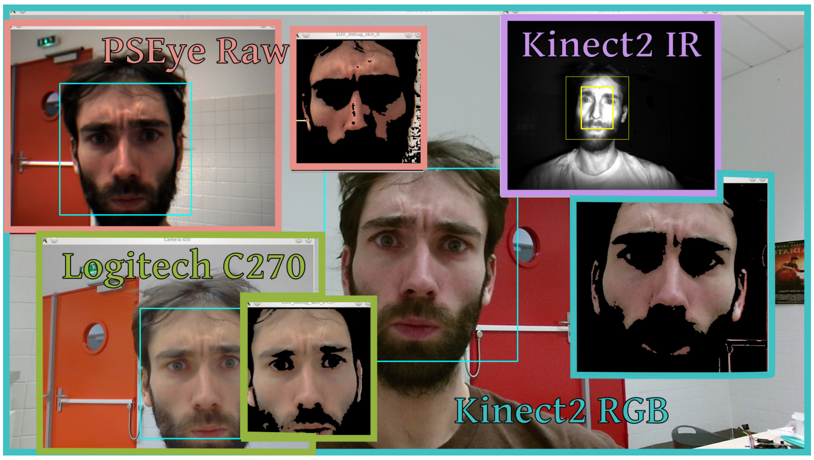

Figure D.3 – Snapshot of the video feeds that were simultaneously recorded during the experiment, including skin pixels masks for color feeds. Maybe by the look you could tell it was the last experiment that took place in the scope of this thesis.

Because we wanted to stress the quality of the recordings and see if we were able to retrieve heart-rate variability from remote PPG measures, in this study we replaced the last stage of our pipeline by a heartbeats detector. Indeed, using a continuous wavelet transform over 10s windows may be less prone to artifacts in casual settings that require no more than an average heart rate – e.g. chapter 9 –, it is also less sensitive to instantaneous heart rate – heart rates values are smoothed.

Hence, all PPG signals – as well as the ECG signal – were retrieved and processed in OpenViBE (version 1.0.1). The ECG signals were filtered with a Butterworth band-bass filter between 1Hz and 20Hz. The ECG being loud and clear, heartbeat detection was achieved with a simple threshold applied to the power of the signals.

On the other hand, PPG signals were more noisy – even though we could clearly see the pulsatile component in the raw signals (Figure D.4). PPG signals were smoothed other a sliding windows of 0.25s and then a 5Hz, low-pass filter was applied. Note that because we needed instantaneous heart rate in real time, we had to use a low filter order – 2 for instance. Indeed, while the 512Hz sampling rate of the ECG ensures chunks of data big enough for OpenViBE to apply higher orders, with PPG streams sampled at 30Hz there are only 7 data points aggregated over 0.25s. This is also why we used a derivative instead of a low-pass filter to remote signals drift. After filtering, a heart beat was detected when the derivative changed from positive to negative values, that is to say when the "peak" of the pulsatile component was reached.

True formula for max heart rate: 208 - 0.7 x age.

Finally, whatever the source of the signal (ECG or PPG), the instantaneous heart rates were computed by measuring the elapsed time between two consecutive heartbeats. To discard values that would make

D

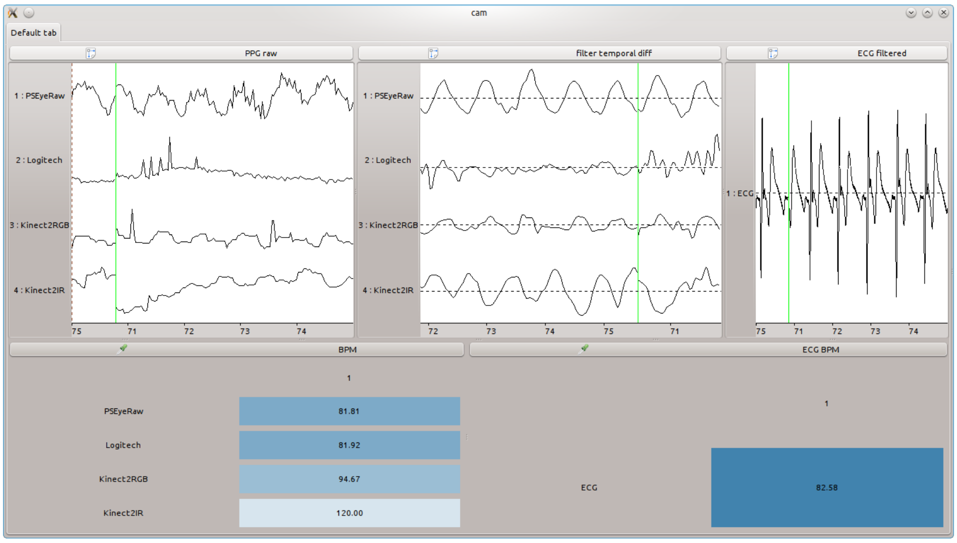

Figure D.4 – Screen capture of the PPG signals processed in real time within OpenViBE. *Top left*: raw signals exctacted extracted from video feeds. *Top middle*: filtered signals. *Top right*: ECG. Bottom: corresponding instantaneous heart rates. Pulsatile components can bee clearly seen within the raw signals of the PlayStation Eye (top left, first plot).

no sense physiologically, one last filtering occurred. Heart rates were cropped between 30 beats per minute (BPM) and 190 BPM [Tanaka et al., 2001] and could not variate by more than 45 BPM in one second [O'Brien et al., 1986]. The heart rates acquired within OpenViBE were then exported to the hard drive for later statistical analysis.

D.4.3 Results

The instantaneous heart rates measured from each PPG stream are plotted against the ground truth in Figure D.5.

R version 3.0.2 was used to perform statistical analyses. The instantaneous heart rates collected from PPG and the data retrieved from ECG were compared using a Pearson correlation ("rcorr" function of the "Hmisc" package). The correlation coefficients are presented in Table D.1; the corresponding p-values – corrected for multiple comparisons using false rate discovery (FDR) – are in Table D.2. Table D.3 lists the heart rate grand averages over the sessions.

D.4.4 Discussion

The data acquired with the PlayStation Eye camera is the closest to the ground truth. The heart rate patterns fit EGC measures; the variability induced by breathing could be observed in particular during the second recording session (Figure D.5). These findings are confirmed by the correlation scores. With a grand average which is only 1.5 BPM off the

D

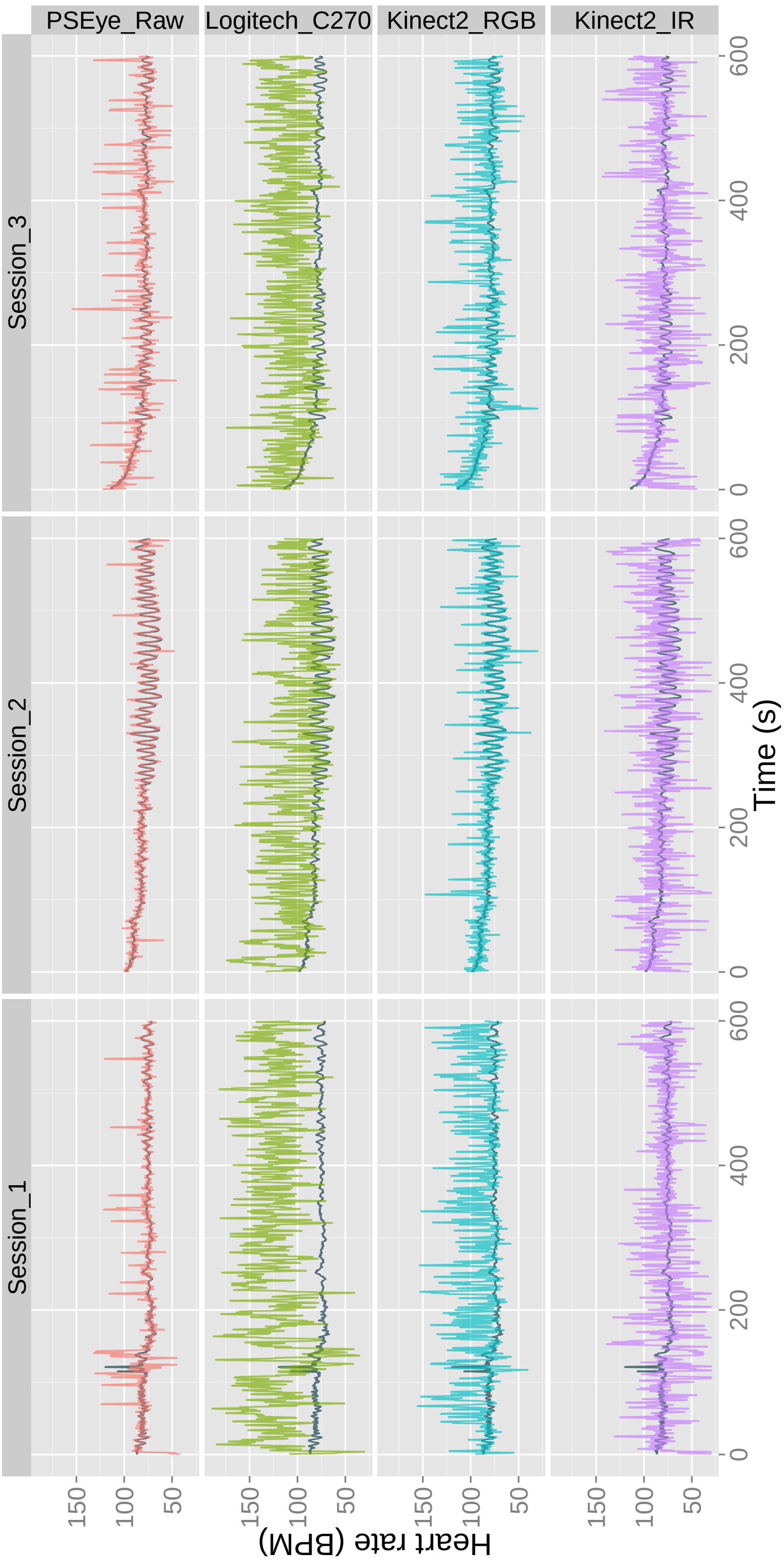

Figure D.5 – Instantaneous heart rate from video feeds (in color) against the measure obtained with ECG (dark line). Note that as opposed to figure D.4 the oscillations that could be observed are *not* pulsatile components, but the heart rate variability caused by breathing patterns.

Table D.1 – Pearson correlation coefficients between the instantaneous heart rates obtained from the webcam and the measures acquired from the ECG.

	PSEye Raw	Logitech C270	Kinect 2 RGB	Kinect 2 IR
Session 1	0.301	-0.079	0.025	-0.031
Session 2	0.811	0.316	0.540	0.000
Session 3	0.463	0.085	0.333	0.011
Mean	0.525	0.107	0.300	-0.007

Table D.2 – p-values – FDR corrected – corresponding to the Pearson correlation coefficients presented in Table D.1.

	PSEye Raw	Logitech C270	Kinect 2 RGB	Kinect 2 IR
Session 1	< 0.001	< 0.001	0.103	0.044
Session 2	< 0.001	< 0.001	< 0.001	0.980
Session 3	< 0.001	< 0.001	< 0.001	0.500

Table D.3 – Mean heart rate over the entire recording sessions.

	ECG	PSEye Raw	Logitech C270	Kinect 2 RGB	Kinect 2 IR
S1	76.38	77.81	122.6	90.49	76.92
S2	79.52	79.68	100.80	81.74	83.90
S3	80.31	82.60	107.80	85.75	83.31
All	78.74	80.03	110.40	85.99	81.38

ground truth, the results suggest that these remote measures could indeed account for heart rates.

The numbers are less conclusive concerning the other webcams we had the opportunity to test, however in this case we did not try to optimize anyhow signal acquisition. Notably, the IR video feed obtained with the Kinect 2 could benefit from more robust algorithms (e.g. [Cennini et al., 2010]). The RGB video feed of the Kinect 2 ranks second in the correlation scores. While this device is at the moment an order of magnitude more expensive than the PlayStation Eye camera – ≈ ten times more –, its high resolution, combined with its wide angle lense may be put to use to cover a bigger area, or to measure multiple users. The Logitech C270, which belongs to an intermediate category, shows how an image that may be more flattering to the eye due to sharpening or auto color correction (Figure D.3) fails to deliver proper signals. There was little correlation between this camera and ECG, and the average HR was nearly 60% more important than it should be.

Sensitivity to motion is the main reason why there were so many artifacts that pushed instantaneous HR to high values. During our real use-case scenario (chapter 9), in combination to a heart rate extraction method less sensitive to noise, we used a confidence index that discards values when motion is detected [Bousefsaf, 2015].

D.5 Conclusion

In this chapter we propose an implementation of a heart rate measure that does not require to equip users. Off the shelf cameras could be employed to measure the instantaneous heart rate of multiple users in real time.

On a technical point of view, it is worth noting that the framework we developed could accommodate other devices besides the PlayStation Eye. For instance, we were able to incorporate the Microsoft Kinect 2 to our pipeline. Even though the first measures we obtained with this device are not on par with the carefully optimized pipeline of the PlayStation Eye, the Kinect 2 is interesting because of the wide angle of its lens. We have a prototype that uses a single video feed to record the heart rate of several users at the same time. Besides, the infrared camera that is integrated may be used to improve the accuracy of the physiological measures once robust algorithms will be integrated. The Kinect 2 could be used in a "blackjack" placement, with players seated in an arc and facing the camera – whether in this case the game is "video" or not.

Combining remote sensing for input modality and spatial augmented reality for output modality enable the emergence of complex HCI; and yet most of the technological artifacts are hidden to end-users. In the application proposed in chapter 9, no computers lie in the immediate environment of the players. Seamless integration of the technology is another step toward acceptability.

E

Remove artifacts the human way

E.1 Introduction

Physiological sensors in general are sensitive to noise caused by users' motion. With EEG in particular, many muscular artifacts could pollute the data. If the upper part of the body is tensed, the amplitude of the signal notably increase in the upper beta frequency band and beyond (i.e. gamma). Should the user clench his or her teeth, and the corresponding EEG segment is to be discarded.

There are other sources of noise than participants for the instruments, such as the utility frequencies from the power grid that scramble signals around 50hz or 60hz depending on the country, but these external factors are much easier to anticipate and acknowledge. Users' behavior is often more random, more difficult to prevent or fix.

Although the experimenter instruct users to move the least and to remain relaxed during the experiment – with instructions such as "try to blink only between trials", it is tedious and often frustrating to remind user when too often noise is seen within the EEG; the experimenter has to monitor the signals consciously and it also jeopardizes users' immersion.

This was the problematic of the part of the team working with BCI and physiological sensors. I saw a way to remedy to this after fruitful discussions with Renaud[1] about persvasive and ambient feedback – he is

[1] Go read his thesis!

into gentle reminders displayed in the ambient space rather than irritating popups.

What if, instead of explicit warnings, the level of noise is displayed in a supporting and *not* disruptive way to end users, so they could correct their behavior on the go, autonomously? Better still, what if the feedback is so meaningful that no instructions are needed anymore?

To support our idea, we implemented both an "explicit" and an "ambient" feedback. While we had it working with an EEG setup – some users tested an ambient feedback while we exhibit Tobe (chapter 11) – the first real deployment of this type of feedback occurred while we were using remote PPG, as seen in previous chapter. As such, we will use this implementation, aimed at reducing heart rate artifacts caused by head movements, to detail what form could take this feedback.

We had an experiment ready for Teegi during the IIT Techfest but nothing was done in the end because there was too many attendees.

E.2 An ambient feedback to guide users

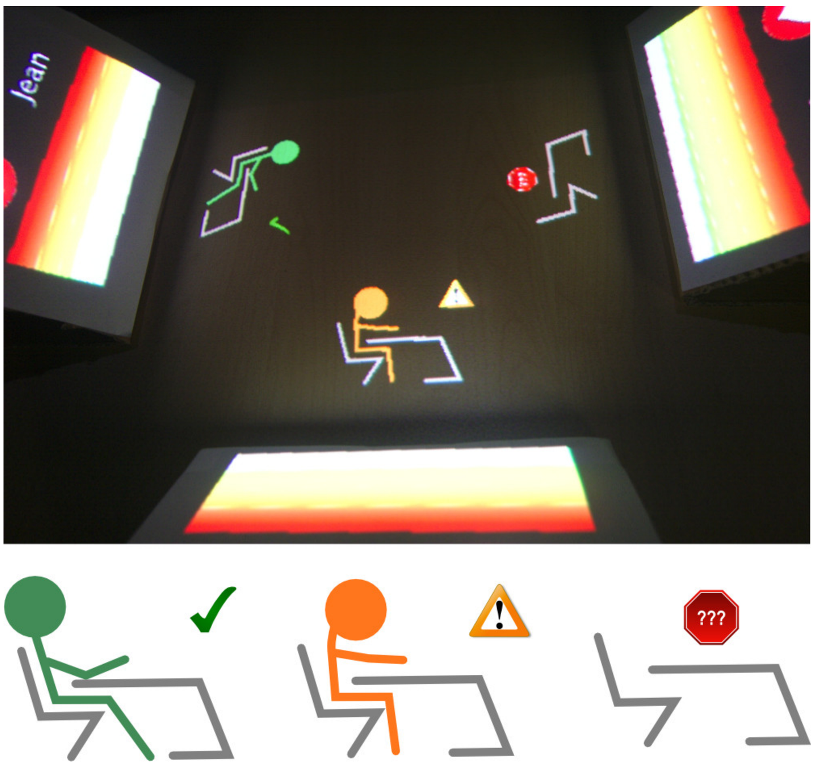

Figure E.1 – Proof of concept of an "explicit" feedback that uses spatial augmented reality to indicate PPG signal quality. *Left*: good signal. *Center*: bad signal. *Right*: the user is not detected.

We took the opportunity of the study supporting remote PPG (chapter 9) to test our hypothesis regarding an ambient feedback that could guide users' behavior during physiological recordings. For instance, PPG is sensitive to motion artifacts.

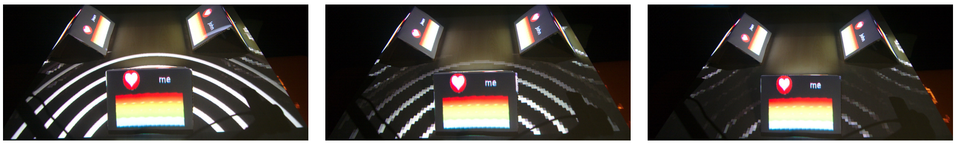

Figure E.2 – Example of an "ambient" feedback, consisting of moving waves, that gives information about PPG signal's quality. *Left*: good signal. *Center*: user is moving too much, waves are pixelated. *Right*: the user is not detected, waves are noisy

The best strategy to prevent such artifacts would be to give an *explicit* feedback to users', that is to say a feedback that they would not miss and that they would understand at first sight. Such feedback could take the form of pictures, visible icons displayed on directly in front of the player – see Figure E.1.

Even though an explicit feedback should be the most effective solution to acquire good signals [Hattie and Timperley, 2007], it may disrupt too much the interaction; during pre-test we realized that displaying an explicit feedback grabbed too much players' attention. Besides, they felt more anxious because of such a visible prompt, which was giving a negative feedback when they were moving too much.

In our game settings, the exchanges between players were more important than the physiological measures, and depending on the intended use of a BCI, it could be important not to draw too much of users' attention. This is why, instead of a feedback explicit and disruptive, we developed an *implicit* feedback. Our approach is similar to the "ambient persuasive feedback" tested in [Maan et al., 2010], where a lamp with changing colors gave feedback about power consumption.

E.3 Apparatus

As mentioned previously, our pilot study took place alongside the board game described in chapter 9. During those gaming sessions, participants were playing to a card game by groups of 3. We recruited 18 participants, i.e. 6 groups of 3 players – 5 females, 13 males, mean age 23.3. We gave to half of the groups an ambient feedback about the quality of remote PPG recordings. The other half acted as control groups and did not have any feedback.

Our implicit feedback took the form of "waves" going from the participants to the game table, using projection. Note that this "ambient feedback" was not overlapping the heart rate visualizations that took place on dedicated display stands. The ambient feedback was projected in the surroundings of the players, the waves smoothly fading in the area covered by the projector.

The waves were pixelated if a participant was moving too much, and were becoming very noisy if the face was not detected (see Figure E.2). For each participant, the state – "still", "moving" or "not detected" –

had to remain unchanged for 2 seconds before we changed the ambient feedback.

To control for participants' motion, a "confidence index" was computed using the formula presented in [Bousefsaf, 2015]. It uses the derivative of the PPG value before filtering as an indicator of players motion, hence as an indicator for the quality of the PPG signals. With no motion the confidence index reaches 100. We set a threshold to 50, below which the ambient feedback was pixelated.

During the card board game used alongside remote PPG, the groups that had an ambient feedback received the same instructions than the groups without. Once, before the study started, they were told to remain in front of the webcam and to limit their head movements for the heart rate to be accurately measured. After this sole disclaimer, they were not remembered the instructions.

E.4 Results and conclusion

The influence of the ambient feedback over players' behavior and signal quality was tested using a between-subjects experimental plan – 2×9 participants. We measured, from the one hand, the ratio of the time during which players' faces were detected, and from the other hand the ratio of the time during which the confidence index remained above our threshold of 50.

We tested for significance using Mann-Whitney test. There was no significant difference between the group that received an ambient feedback and the group that had no feedback, neither for the ratio of face detection nor for the ratio related to the confidence index. Mean for face detection ("ambient" group *vs* "none" group): 0.14 *vs* 0.13; SD: 0.16 *vs* 0.14. Mean for confidence index above 50: 0.21 *vs* 0.23; SD: 0.09 *vs* 0.09. Note that the low ratios are explained by the fact that the recordings took place during whole game sessions, *including between turns*, when players were casually talking to each others.

Even though the size of our sample and the important variance between participants prevented the appearance of any significant difference, we believe that with a bigger population and a less rich environment, this kind of feedback could be a useful addition to physiological monitoring. Notably, out of the 9 players that got the ambient feedback condition, 5 did not notice *at all* that something was projected on the table beside the HR visualization. The one who did thought it was only some cosmetics. The ambient feedback was discrete by design, and we hoped that the self-explanatory relationship "quality of the waves == quality of the signal" would strike players. Had the players been less engaged in the game and their attention drawn to the waves from the beginning, the ambient feedback may have served as a gentle reminder of signal's reliability.

In the future this kind of ambient feedback should be studied more thoroughly, for example it could be combined with a BCI based on motor imagery or arithmetical tasks in order to discreetly remind participant why their mental state is not recognized. Using spatial augmented reality to display information in the surrounding space instead of the screen would avoid distracting users' attention from the main task, hence such method that improve signals reliability would come with little or no additional cost.

Behind Science

A HUMAN STORY

Every chapter of this thesis or so rests on the shoulders of various collaborators – colleagues and students – that accompanied me during those last 3 years. I owe them a debt; the least I could do was to spare a few paragraphs about each one of them.

Credits

Rolls

Credits.1 The backstage behind stereo (chapter 5)

I *have* to write down a huge THANKS to two persons with whom I have collaborated during the evaluation of stereoscopic displays: Léonard Pommereau and Aurélien Appriou. Both have been interns during the first year of their master degree in cognitive science, staying for about 12 weeks in the team.

When Léonard came in 2013 he only had a fuzzy subject: combine stereoscopic displays and EEG. He was given some papers to read and within a couples weeks he managed to come by himself with an idea about how users' visual comfort could be assess with EEG. At that time I was busy with something else, but as soon as he was done with the ~~boring~~ necessary bibliographic stuff my duties magically disappeared (no, I'm not *that* Machiavellian, I really had a deadline ongoing while he was sweating). As his work was very close to the main theme of my thesis I jumped in, of course, and after few brainstorms with the rest of the team we refined the protocol. Léonard managed to do an amazing work in a short time frame. After he was gone I played a little with the data, most of the work consisting in analyzing the EEG, and that led to the publication of a pilot study. It was the first time during my thesis that I had the impression to do *real* science – not that a minor point should I add – and it's already one hell of a good reason to high-five the man.

The "stereo" project, as call it in my file-system, was put on a hold until the year after, when Aurélien arrived from the following vintage, taking over. This time we had a clear view of what needed to be done, the items that needed to be improved in order to finalize the protocol and produce a complete study. Nonetheless, Aurélien *also* had to work by himself for a couples weeks, shortly after he started his internship… while I was lucky enough to fly to Toronto to present the first results we got (weird timing). Aurélien managed to support with the right references the hunches we got, at the same time he got his hands on many mysterious technologies in order to shape the software behind the final study. Surprisingly, he accommodated well to a geekish environment, we were not easy on him. He was also a great asset thanks to his numerous relationships: he recruited 12 participants that were close demographically, perfect for good inferences. All the pieces of the puzzle were eventually there in order to obtain definitive results. Well, I do not intend to write references letters, only to report another great team work – if you have the occasion, buy one or two students from the cognitive science master of Bordeaux University, they're good (and I don't say that because I've been there myself). By that time of course I already was a true and great etc. scientist, the EEG signal processing pipeline was there; but still it was another premiere that took place back then: first time I built a true BCI, borrowing handful scripts from Fabien for features extractions and classification. One more milestone on my personal journey and second big-up, it's free.

I can't end this "behind the scene" retrospective without acknowledging the participation of Jérémy Laviole to this project. We took some of his Processing voodoo and knowledge of how to produce accurate stereoscopy in order to craft our stereoscopic images. It's always great to have someone reliable and obliging around. All in all, maybe it is a bit of a long description to serve as a chapter's incipit, but I did not want to silence or overcome the achievements of those with whom I worked. A PhD is not *one*'s journey – would it still be Science if it were? – if for some the "they" may diminish my apparent contribution, so it is.

Credits.2 The other guy behind 3D tasks' evaluation (chapter 6)

The work described in chapter 6 could not have been done without the strong involvement of Dennis Wobrock. Dennis did a six months (or so) internship in the team during 2014. Contrary to some of the others persons mentioned along this thesis, Dennis does not come from the beloved master in cognitive science of the university of Bordeaux… but from a sibling, an engineering school located in the same city, the ENSC – "École Nationale Supérieure de Cognitique". The idea behind this internship was to put into practice some of the first leads that arose at the beginning of my thesis. In fact, for various reasons it was more than a simple practice, it was an investigation and an extension of what I had projected back then. Dennis carried-on on his own, while I was busy with Teegi (chapter 10) and tying up loose ends regarding the evaluation of stereoscopic displays that you just read – I always got many irons on the fire! I won't hide the fact that Dennis and I had little interaction during this period. I was genuinely interested by the work that was being achieved, but had no real opportunity to give a hand. Not that Dennis would have needed it. He was already well supervised by his many advisers, and managed to keep himself busy for the whole time.

His job was not facilitated by the fact that he shared his time between the laboratory and the company that partnered on the project, Immersion. It became a strength: the many spokespersons involved and the numerous meetings that was necessary to keep everyone on track resulted in various reports that proved valuable afterwards, when the work needed to be taken over. Apart for some piece of code and for being a naive participant (quite helpful to experience first-hand something when one need to describe it later-on), when the study actually took place I was off. I like to think that I could have helped to speedup some part of the process if I tried harder to get my hands dirty, but even with the equipment that I actively voted for I'm not really sure I would have done a better job. The Bitalino was a new addition to the hardware that the team possessed, and in a short amount of time Dennis managed to integrate it within our OpenViBE workflow. It was not the only device that he needed to master within a small time frame, there was also the interacting device developed by Immersion, the "CubTile" that was one of the focus of the study.

Once more, I want to stress where my contribution starts, and when it stops. I had some influence on the background story, even if at the midst of my thesis I was not sure I would spend much effort in physiological sensors beside EEG (silly me!). So, Dennis did most of the stuff, and I am here about to mercilessly steal all of that and integrate *his* work in *my* thesis, is that so? Not exactly, fortunately. The study of the CubTile fits beautifully in the landscape of a thesis that study at large how EEG could enhance HCI evaluation. You're right, it is not enough; it's not because you pass by a nice painting that you could grab it and hang it in your bedroom. In fact I really stepped-in in the project… once Dennis left (and that it was time for me to get back on tracks[1]).

Dennis managed to pass the experiments before the end of his internship, but he had little time to do proper analyses. All in all we end up by a promising study but with no solid evidence about what we could expect from it, just some hints about the kind of help that physiological recordings could give to HCI designers, and one or two figures that seem to indicate that indeed, maybe there was something that was worth it, some interesting curves and peaks that *seemed* to match a purposely difficult interaction.

This is why and where I came in. To rescue the study, to thoroughly analyze data and shed light on the first evaluation of an interaction technique and through physiological sensors.

[1]…and that a close deadline pushed me to dive into Matlab, but shush, this kind of behavior never happens in Science.

Credits.3 The working force behind Maze project (chapter 7)

It may not seems so, but from all the stories collected in the present manuscript, the study described in chapter 7 is by far the one that spanned an over the longest period of time.

The very idea of using a 3-dimensional maze to test various constructs and conceive over-complicated interfaces and poor interaction techniques – not that simple to make bad choices so as to test the good ones – dates back to before I started my thesis, when I was merely an intern in a cognitive neuroscience laboratory. Not that I want to disrespect in any way the beloved participants that took part in this work, but truth said, the at 3D maze comes from a… 2D maze that was – and is – tested with… monkeys (and that comes from a T maze tested with mice, but enough!). Our non-human primates siblings are very good at learning and repeating over and over simple tasks, and their occupation in the maze I'm referring to is inspired by the multi-armed bandit used in decision-making. Curious may refer to [Etienne et al., 2014] to gather some clues about that.

Basal Gang forever!

So I had this idea of transferring the "monkey maze" to a (clever and engaging) 3D environment, with tons of possibilities – study spatial orientation, see which strategies is used between learning by place or learning by targets, add probabilistic choices to the picture, … – but there was one slight problem in the way. I hardly had any skill regarding the programming of such environment. Not to say it was an impossible task with the tools we have nowadays, but it still required a major investment in time and in will. I only managed to achieve a *very* rough prototype with Unity. Then the project went dormant – shortly after I finished my prototype I was irresistibly attracted to a little tangible guy that you have read about (or you *will*, depending on you reading itinerary). There had been talks and meetings here and there, but in the end I never manage to finish what I started, even though it should have been from day one the major accomplishment of my (original) thesis.

One year after my last commit other projects passed by and almost all my hopes were dry. I hardly saw anything but a sad ending. Then, as an angel descending from the sky, came Maxime Daniel. The Immersion company with which we collaborated previously saw some potential in the almost dead project, and they let one of their interns toy a bit with it. Let's say that Maxime did far more than his share. Not afraid by having others things in the pipeline, he brought back to life the maze prototype. A good and hard worker he is – and not the least, no flapping wings that put feathers in every corner.

Thanks to his skills he surpassed what I could have possibly done by myself, adding his own views to the mix, preventing more than a few wrong turns. Beside his technical and theoretical *savoir-faire* (*potpourri*: LSL, audio probes, Sternberg paradigm), he also had a sheer amount of enthusiasm to share. A willpower that never failed nor faded all along those 8 months he had been working on this project. Even when I forced him to implement stupid ideas that we thrown away, even when I had his own agenda and study to finish, even after the official term of his internship, even once he found a PhD of his own, even when he left the city (!), he made sure that the study could be achieved in time – i.e. before the deadline of the 3-letters-conference that shakes the entire human-computer interaction community each year, incidentally before I had to send my manuscript. The "maze energy" owes him one (all the graphics are entirely his fault, too).

So, thank you Maxime, if it were not for you, this thesis would have missed a proper closure. May the EEG goddesses and gods – or any other deity that you revere – favour the PhD thesis you just started.

Credits.4 The missing geek behind OpenBCI evaluation (appendix B)

When I had the idea of the simultaneous recordings for evaluating formally the OpenBCI board, I rapidly became excited by the prospect. A perfect comparison in one pass! Involving only one participant while maintaining a high scientific value is not that common. I dutifully sketched the wiring that could adapt our electrodes to the OpenBCI board, prototyped a first version of the adapter, dreamed of the perfectly isolated montage based on no less perfect electrodes… and that was it. For months and months the item stayed on my todo list. Huge potential but low priority. Most of all, it was such a pain to solder this damn D-sub 26 connector that my eagerness to start the study was balanced by fear to ruin the electronic components. For sure I would have ended with a poor hardware.

Fortunately, a couple months before the term of my PhD a new engineer arrived at Inria, freshly out of school but loaded with the skills that I lacked. In a matter of days Thibault Laine came up with the solution I was desperately after. *Long* days – it was a chance to stay behind while most of the center was on vacation; during this period I was the sole beneficiary of this work force. Digging into specifications, drawing electrical circuits and soldering components that I could barely see were a no-brainer for him. I would have never imagined it was remotely possible to end up with an actual PCB – it seemed nearly *easy* to craft those "Gerber" files while I watched him. A thrill to receive something that looks so much like a "real" product. And, above all, that behaves as intended. I'm particularly grateful that all went right at the first attempt. My schedule was tight and it was literally during my last week in the laboratory that I could finally proceed to the study that I longed – best value for money ever. This thesis would have been one chapter *too* short if it were not for Thibault's committed labour.

Credits.5 The precious students behind remote PPG (appendix D)

I had the chance to teach a bit alongside my PhD, and among my duties I supervised groups of students during software engineering projects. One of those groups chose a subject of mine: measuring heart rate through a video feed. Rings any bells? Back then I was new to the matter and just threw at them some references and a task more challenging than expected. They faced many difficulties but they managed to keep the boat afloat. I must admit that I was sometimes part of the problem; at first I absolutely wanted to use a Kinect 2 on *linux* to do that thing. It required more hacks than expected, they lost weeks trying to please me before I let them focus on what really mattered for them, the code.

Even though in the end very few of their code remains in the software presented here – I'm too paranoid to let students have the last word regarding algorithms! – their motivation drove me to have a closer look at their prototype. When I realized this remote PPG thing could go somewhere, I secretly forked their project in its early stages and started to follow my own path. So, thank you Hereiti Hatitio, Anta Mbaye, Jean-Baptiste Rey and Maxime Vincent; for all the crazy projects I thought of over the course of my thesis, without you I would have not jumped the shark on this one.

Credits.6 Tangible avatars, tribute to a Quebecois (part IV)

You know, those proverbs and expressions that lost their meaning after they have been said so many times out of context, becoming mere *clichés*? It's like an old clothing you keep wearing even years after colors washed out, the days off, at home. At some point there was something written on this T-shirt, a text that now only you could decipher.

"Two sides of the same coin"… how does it sound? Maybe in other circumstances *that* would oversell how an external work complement a thesis; yet I can't think of a better term to describe how the last part of my work – and by extension the very point I'm trying to make in my manuscript – would be shallow, no, *empty*, if it were not for some guy popping in the Potioc team a couple months after I began my PhD.

Renaud Gervais, coming from the far far away Québec, had been doing a PhD at the same time as myself. He started with a subject very unlikely to have anything to do with me, all about creative coding and tangible user interfaces. The fate put both of us in the same open space (speaking about the sneaky influence of architecture) and, even worst, placed our desks few meters apart (supposedly projection mapping needs the right balance of luminosity, between the sun and artificial light… my foot!). So be it.

I was trying to measure mental states with passive brain-computer interfaces, playing with physiological sensors and experimental protocols, interested in social interactions. He was trying to shatter the barrier between digital content and physical surroundings, playing with the idea of customizable objects that you would be compelled to hold and interact with, interested in personal development.

Slowly each one got curious in the work of the other. Before we knew it, our worlds collided. Supported by comprehensive supervisors, who gave us freedom to do so, we ventured into another territory. This part is at the crossroad of our motives, the result of our complementary skills, aftermath of a fruitful emulation; I think you get it.

That is not to say that our theses are the same. Even if it is now difficult, if not impossible, to decide which contribution belongs to one or the other, I believe we did not have the same angle, and not the same agenda either – for example, if you were interested in tangible interaction, SAR, cute things and creativity, you're definitely reading the wrong manuscript.

All I can do is recall some memories from my PhD life before we started to build Teegi together.

I am 5 months into my thesis, I present what BCI are and how they work during "*La semaine digitale*", a public event aimed at promoting computer science that took place in my city. I am myself quite new to the subject – fortunately I have a great mentor – and to conclude my talk I show miscellaneous applications, among them the not *so* insignificant Necomimi[2], a headband with "cat ears" that supposedly move according to brain activity. After a handful of medical applications and signal processing, these examples are here to lighten the mood, make the audience laugh a bit. (You've never encountered genuine comedy until you see me impersonating cat ears with my bare hands.) Despite the fact that with one electrode on the forehead Necomimi may have more to do with facial expression than anything else, since I inserted this picture in my slide I could not stop thinking *what if*? What if such device could come with true brain signals and proper mental states?

Since I began my PhD 13 months have passed. I have the impression to have such a slow start. After a year I finally had *one* paper accepted, but I'm not sure if I should celebrate or if it's just random luck after a third submission; sometimes I doubt I will be able to advance much my PhD subject. (Fear not, young Jérémy, basically it ended up being parts I and II). It's night, we are at Cap Sciences, in the midst of IHM '13 conference, after a keynote talk from Anatole Lécuyer. (Signs of the things to come.) Appetizers and drinks await the replete audience. Usually I'm not last to answer to the free food call, but tonight I don't feel like it. Instead, I'm seating in a corner of the first floor, distant from the animation. In the dark I'm starring at the screen of my laptop and

[2] http://neurowear.com/projects_detail/necomimi.html

my fingers type furiously. During the talk there were works I already knew about and sometimes my thought wandered against my will. Between two BCI and virtual reality applications, suddenly I had a *vision*. And now I have to write everything I can remember of before it vanishes. An avatar representing inner mental states, a device to make explicit the emotions and processes of the mind. Not talking about human but *social* enhancement. What shift in social interactions could that bring? Useful for people bad at empathy – may be easier to interact with an avatar – or to facilitate collaborative work – could gently point out the colleagues in trouble; for sure I'd be a better teaching assistant with such a tool. What form could it take? Animated tatoos? A smartphone around the neck with a cartoon face on it? The event ends, the restaurant awaits, the grandest inspirations fade before my stomach, I get back to reality. The days later I try to find a name for what I want to achieve. Social prosthesis? Physiological homunculus? Brain to social interface? I discover that the slot is not entirely unexplored; their website looks old but I found interesting references in this "Affective Computing Group" thing. (Silly me for having underestimate the fierce MIT media lab back then, as if they are one step ahead for *everything*.) Unfortunately I have to put a hold on my review and focus once again on how to use BCI for 3DUI evaluation. I keep my notes aside but for the better I must forget my impossible fantasies.

And then, just few months later, right after I came back from PhyCS in January 2014, Renaud came forth and wondered what it'd feel like to see from the outside his own brain activity – at that time there was already the Mind Mirror project to think about. And we started to talk, to share. Step by step blurry interrogations became more consistent and Teegi was brought to life, first of his kind, followed by Tobe after a long maturation. I could not have conceived a better platform for the "social prosthesis" (or whatever, I'm bad at branding) described above. It's only the beginning, we have yet to test formally our hypothesis concerning social interactions and empathy, but I never thought I could go sot far already. This manuscript is the "secret plan" flavour of my thesis, shyly disclosed one year ago during a team seminar, when I presented several outlines for my manuscript – including an ambitious one with numerous "work in progress" and "todo" flags. I did not expect *this* version – which somehow hold everything in place (or so I hope!) – to become the final one. Thank you Renaud, you made my dreams come true.

Moreover, you were a valuable companion during 3 years. I met and talk to PhD students outside the team, outside the lab, outside the field. Among the recurrent critics regarding the harsh life during a PhD there was one that stroke me by its strangeness: most of the students felt lonely. Very often they had to work on their own they said, with little interactions with other members of their teams. As far as I can tell communication was not the issue, it was really more about collaboration. Up to the point that many were not sure they wanted to pursue in research. A shame since Science is all about sharing (but a *useful* shame, that means more positions available later!). I sympathized and tried not to brag too much since this was one problem I did not have *at all*. Those testimonies helped me understand once again my good fortune. On many projects I had been closely working with people around, above all with Renaud. So I guess he also saved me from mental breakdowns. Thanks for my sanity, pal! What? This whole section almost feels like a love letter? Why not? They may be made of plastic and a bit dead inside, but we had *babies* together. And here they come.

Credits.7 Tobe, son of many (chapter 11)

I will not mention *again* how much this work owes to Renaud, instead I will briefly put the highlight on a *third* dad for Tobe: Alexis Gay. Our work would have not be same without this intern in Design – who's also got a background in cognitive science and education (3 master degrees largely make up for a PhD!). While we wanted from the start to go public with Tobe, to bring our project in the field, Renaud and I are not very comfortable with strangers when we have to put our hats as scientists, especially when they are many.

Alexis does not have this problem. More than we knew, we needed someone not afraid to put a prototype into the hands of people; as a chill-tattooed-surfing-artist he fitted perfectly the character. He's the kind of guy who go *talk* to people – and who *enjoy* it! Unthinkable, and yet this is quite a precious perk. We went on is own with pages of questionnaires seeking people's representation about mental states and physiology. He came back unharmed but full of data.

Beyond his personality, he spurred his views and his working method – before I met him I had no idea about designers' job, I thought they were making *nice* things, not that it was all about the process, with fancy words such as "participatory design". The methodology for shaping the "design space" below, it's him. He forced us to think Tobe as a product, speeding up work, helping us to face dead-ends.

Valuable addition to a HCI team, Alexis was also our interface with the scientific museum where we intervened. For a long time we longed for a public space where we could expose our works. We were lucky enough to have Cap Sciences as a playground and Didier Laval as an enthusiast interlocutor. We write "scientific museum" to save on words count, but it is truly a *scientific cultural center*, a shared space between the public, makers, scientists and designers. Pay them a visit next time you come to Bordeaux, maybe you'll see us around.

During our journey with Tobe we also had the occasion to talk to colleagues from other fields about our project. Not only did their positive feedback comforted us in our approach at a time when doubts were clouding our judgment, but we also gained new ideas of applications thanks to their ingenuity. On particular, we would like to thank Pierre-Alain Joseph from the laboratory "Handicap & Système Nerveux" (Bordeaux University) and Éric Sorita from the university hospital Pellegrin Bordeaux – using Tobe for stroke rehabilitation, it's all on them – as well as Matthew S. Goodwin from the Bouvé College of Health and Sciences (Northeastern University), who saw the potential of Tobe for teaching STEM. Finally, I would like to thank Christelle Godin. She does not know it yet, but by casually mentioning cardiac coherence on the way back from PhyCS' 15 welcome cocktail, she sparkled a nice multi-users scenario.

For the brightest ideas remain rough and fuzzy concepts without the right people to nurture them, the awesomeness of Tobe stands on the shoulders of many.

BIBLIOGRAPHY

Author's references

[Frey, 2015a] Frey, J. (2015a). Émersions sensorielles. *CORPS*, 13:113–121.

[Frey, 2015b] Frey, J. (2015b). Heart Rate Monitoring as an Easy Way to Increase Engagement in Human-Agent Interaction. In *PhyCS - International Conference on Physiological Computing Systems*. (Cited on page 96)

[Frey, 2016a] Frey, J. (2016a). Comparison of a consumer grade EEG amplifier with medical grade equipment in BCI applications. In *International BCI meeting*. (Cited on page 182)

[Frey, 2016b] Frey, J. (2016b). Remote Heart Rate Sensing and Projection to Renew Traditional Board Games and Foster Social Interactions. In *Proceedings of the extended abstracts of the 34th annual ACM conference on Human factors in computing systems - CHI EA '16*. (Cited on page 111)

[Frey et al., 2015] Frey, J., Appriou, A., Lotte, F., and Hachet, M. (2015). Estimating Visual Comfort in Stereoscopic Displays Using Electroencephalography: A Proof-of-Concept. In *INTERACT '15*. (Cited on pages 21, 28, and 35)

[Frey et al., 2016a] Frey, J., Appriou, A., Lotte, F., and Hachet, M. (2016a). Classifying EEG Signals during Stereoscopic Visualization to Estimate Visual Comfort. *Computational Intelligence and Neuroscience*, 2016:1–11. (Cited on pages 28 and 35)

[Frey et al., 2016b] Frey, J., Daniel, M., Castet, J., Hachet, M., and Lotte, F. (2016b). Framework for electroencephalography-based evaluation of user experience. In *Proceedings of the 34th annual ACM conference on Human factors in computing systems - CHI '16*. (Cited on page 73)

[Frey et al., 2014a] Frey, J., Gervais, R., Fleck, S., Lotte, F., and Hachet, M. (2014a). Teegi: Tangible EEG Interface. In *UIST-ACM User Interface Software and Technology Symposium*. (Cited on pages 122, 124, and 141)

[Frey et al., 2014b] Frey, J., Mühl, C., Lotte, F., and Hachet, M. (2014b). Review of the use of electroencephalography as an evaluation method for human-computer interaction. In *PhyCS - International Conference on Physiological Computing Systems*. (Cited on pages 1 and 145)

[Frey et al., 2014c] Frey, J., Pommereau, L., Lotte, F., and Hachet, M. (2014c). Assessing the zone of comfort in stereoscopic displays using EEG. In *Proceedings of the extended abstracts of the 32nd annual ACM conference on Human factors in computing systems - CHI EA '14*, pages 2041–2046, New York, New York, USA. ACM Press. (Multiple combos on pages 21, 28, 35, and 37)

[Gervais et al., 2016] Gervais, R., Frey, J., Gay, A., Lotte, F., and Hachet, M. (2016). Tobe: Tangible Out-of-Body Experience. In *TEI '16*. (Cited on page 145)

[Gervais et al., 2015] Gervais, R., Frey, J., and Hachet, M. (2015). Pointing in Spatial Augmented Reality from 2D Pointing Devices. In *INTERACT '15*. (Cited on page 3)

[Roy and Frey, 2015] Roy, R. N. and Frey, J. (2015). Neurophysiological markers for passive BCIs (in press). In *Brain Computer Interfaces: Methods, Applications, Perspectives*. Wiley-ISTE.

[Wobrock et al., 2015] Wobrock, D., Frey, J., Graef, D., de la Rivière, J.-B., Castet, J., and Lotte, F. (2015). Continuous Mental Effort Evaluation during 3D Object Manipulation Tasks based on Brain and Physiological Signals. In *INTERACT '15*. (Cited on pages 28 and 52)

References

[Abbass et al., 2014] Abbass, H. A., Tang, J., Amin, R., Ellejmi, M., and Kirby, S. (2014). Augmented Cognition using Real-time EEG-based Adaptive Strategies for Air Traffic Control. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 58(1):230–234. (Cited on page 30)

[Afergan et al., 2014] Afergan, D., Shibata, T., Hincks, S. W., Peck, E. M., Yuksel, B. F., Chang, R., and Jacob, R. J. K. (2014). Brain-Based Target Expansion. In *UIST '14*. (Cited on page 30)

[Afonso et al., 1999] Afonso, V. X., Tompkins, W. J., Nguyen, T. Q., and Luo, S. (1999). ECG beat detection using filter banks. *IEEE transactions on bio-medical engineering*, 46(2):192–202. (Cited on page 62)

[Agelink et al., 2001] Agelink, M. W., Malessa, R., Baumann, B., Majewski, T., Akila, F., Zeit, T., and Ziegler, D. (2001). Standardized tests of heart rate variability: normal ranges obtained from 309 healthy humans, and effects of age, gender, and heart rate. *Clinical Autonomic Research*, 11(2):99–108. (Cited on page 100)

[Akers et al., 2012] Akers, D., Jeffries, R., Simpson, M., and Winograd, T. (2012). Backtracking Events as Indicators of Usability Problems in Creation-Oriented Applications. *ACM Trans. Comput.-Hum. Interact.*, 19(2):1–40. (Cited on pages 16 and 20)

[Allison, 2011] Allison, B. (2011). Trends in BCI research. *XRDS: Crossroads, The ACM Magazine for Students*, 18(1):18. (Cited on page 30)

[Anderson et al., 2011] Anderson, E. W., Potter, K. C., Matzen, L. E., Shepherd, J. F., Preston, G. a., and Silva, C. T. (2011). A user study of visualization effectiveness using EEG and cognitive load. In *Eurographics '11*, pages 791–800. (Cited on page 54)

[Antonenko et al., 2010] Antonenko, P., Paas, F., Grabner, R., and van Gog, T. (2010). Using Electroencephalography to Measure Cognitive Load. *Educational Psychology Review*, 22(4):425–438. (Cited on page 154)

[Audet and Okutomi, 2009] Audet, S. and Okutomi, M. (2009). A user-friendly method to geometrically calibrate projector-camera systems. In *CVPR '09*, pages 47–54. (Cited on page 115)

[Bachour, 2010] Bachour, K. (2010). *Augmenting Face-to-Face Collaboration with Low-Resolution Semi-Ambient Feedback*. PhD thesis, EPFL. (Cited on page 163)

[Bang et al., 2014] Bang, J. W., Heo, H., Choi, J.-S., and Park, K. R. (2014). Assessment of Eye Fatigue Caused by 3D Displays Based on Multimodal Measurements. *Sensors (Basel, Switzerland)*, 14(9):16467–85. (Cited on pages 37 and 50)

[Bangor, 2001] Bangor, A. W. (2001). *Display technology and ambient illumination influences on visual fatigue at VDT workstations*. PhD thesis, Virginia Tech. (Cited on page 36)

[Becker et al., 2005] Becker, C., Prendinger, H., Ishizuka, M., and Wachsmuth, I. (2005). Evaluating Affective Feedback of the 3D Agent Max in a Competitive Cards Game. In *Affective Computing and Intelligent Interaction*, pages 466–473. Springer Berlin Heidelberg. (Cited on page 110)

[Benabid et al., 2009] Benabid, A. L., Chabardes, S., Mitrofanis, J., and Pollak, P. (2009). Deep brain stimulation of the subthalamic nucleus for the treatment of Parkinson's disease. *The Lancet Neurology*, 8(1):67–81. (Cited on page 30)

[Benko et al., 2012] Benko, H., Jota, R., and Wilson, A. (2012). MirageTable: freehand interaction on a projected augmented reality tabletop. In *CHI '12*, page 199, New York, New York, USA. ACM Press. (Cited on page 128)

[Benko et al., 2008] Benko, H., Wilson, A., and Balakrishnan, R. (2008). Sphere: multi-touch interactions on a spherical display. In *UIST '08*, pages 77–86. (Cited on page 128)

[Berka and Levendowski, 2007] Berka, C. and Levendowski, D. (2007). EEG correlates of task engagement and mental workload in vigilance, learning, and memory tasks. *Aviat Space Environ Med.*, 78(5 Suppl):B231–44. (Multiple combos on pages xvii, 3, 12, 14, and 23)

[Berta et al., 2013] Berta, R., Bellotti, F., De Gloria, A., Pranantha, D., and Schatten, C. (2013). Electroencephalogram and Physiological Signal Analysis for Assessing Flow in Games. *IEEE Transactions on Computational Intelligence and AI in Games*, 5(2):164–175. (Cited on pages 23 and 98)

[Bestgen et al., 2004] Bestgen, Y., Fairon, C., and Kerves, L. (2004). Un barometre affectif effectif: Corpus de référence et méthode pour déterminer la valence affective de phrases. *Journées internationales d'analyse statistique des donnés textuelles (JADT)*. (Cited on page 105)

[Blankertz et al., 2011] Blankertz, B., Lemm, S., Treder, M., Haufe, S., and Müller, K.-R. (2011). Single-trial analysis and classification of ERP components–a tutorial. *NeuroImage*, 56(2):814–25. (Cited on page 44)

[Blankertz et al., 2010] Blankertz, B., Tangermann, M., Vidaurre, C., Fazli, S., Sannelli, C., Haufe, S., Maeder, C., Ramsey, L., Sturm, I., Curio, G., and Müller, K.-R. (2010). The Berlin Brain-Computer Interface: Non-Medical Uses of BCI Technology. *Front Neurosci*, 4(December):198. (ULTRA combos on pages 5, 12, 13, 14, 23, 26, and 178)

[Boksem et al., 2005] Boksem, M. a. S., Meijman, T. F., and Lorist, M. M. (2005). Effects of mental fatigue on attention: an ERP study. *Cogn Brain Res*, 25(1):107–16. (Cited on page 13)

[Bos et al., 2011] Bos, D. P.-o., Poel, M., and Nijholt, A. (2011). A study in user-centered design and evaluation of mental tasks for BCI. *Advances in Multimedia Modeling*, pages 122–134. (Cited on page 20)

[Bousefsaf, 2015] Bousefsaf, F. (2015). *Mesure sans contact de l'activité cardiaque par analyse du flux vidéo issu d'une caméra numérique : extraction de paramètres physiologiques et application à l'estimation du stress*. PhD thesis, Université de Lorraine. (Cited on pages 219 and 224)

[Bousefsaf et al., 2013] Bousefsaf, F., Maaoui, C., and Pruski, A. (2013). Continuous wavelet filtering on webcam photoplethysmographic signals to remotely assess the instantaneous heart rate. *Biomedical Signal Processing and Control*, 8(6):568–574. (Cited on pages 212 and 213)

[Bowman et al., 2002] Bowman, D. A., Gabbard, J., and Hix, D. (2002). A survey of usability evaluation in virtual environments: classification and comparison of methods. *Presence-Teleop. Virt.*, 11(4):404–424. (Cited on pages 20 and 53)

[Bowman et al., 2004] Bowman, D. A., Kruijff, E., Laviola, J. J., and Poupyrev, I. (2004). *3D user interfaces : theory and practice*. Addison Wesley Longman Publishing Co. (Cited on page 53)

[Bradley and Lang, 1994] Bradley, M. M. and Lang, P. J. (1994). Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, 25(1):49–59. (Cited on pages 18 and 118)

[Bradski, 2000] Bradski, G. (2000). The OpenCV Library. *Dr. Dobb's Journal of Software Tools*. (Cited on page 213)

[Brandt et al., 1973] Brandt, T., Dichgans, J., and Koenig, E. (1973). Differential effects of central versus peripheral vision on egocentric and exocentric motion perception. *Experimental Brain Research*, 491:476–491. (Cited on page 8)

[Brockmeyer et al., 2013] Brockmeyer, E., Poupyrev, I., and Hudson, S. (2013). PAPILLON: designing curved display surfaces with printed optics. In *UIST '13*, pages 457—-462. (Cited on page 128)

[Brouwer et al., 2012] Brouwer, A.-M., Hogervorst, M. a., van Erp, J. B. F., Heffelaar, T., Zimmerman, P. H., and Oostenveld, R. (2012). Estimating workload using EEG spectral power and ERPs in the n-back task. *Journal of neural engineering*, 9(4):045008. (Cited on page 12)

[Brouwer et al., 2015] Brouwer, A.-M., Zander, T. O., van Erp, J. B. F., Korteling, J. E., and Bronkhorst, A. W. (2015). Using neurophysiological signals that reflect cognitive or affective state: six recommendations to avoid common pitfalls. *Frontiers in Neuroscience*, 9. (Cited on page 30)

[Bruseberg and McDonagh-Philp, 2002] Bruseberg, A. and McDonagh-Philp, D. (2002). Focus groups to support the industrial/product designer: a review based on current literature and designers' feedback. *Applied ergonomics*, 33(1):27–38. (Cited on page 4)

[Burns and Fairclough, 2015] Burns, C. G. and Fairclough, S. H. (2015). Use of auditory event-related potentials to measure immersion during a computer game. *International Journal of Human-Computer Studies*, 73:107–114. (ULTRA combos on pages 14, 22, 23, 78, 87, and 93)

[Cartwright-Finch and Lavie, 2007] Cartwright-Finch, U. and Lavie, N. (2007). The role of perceptual load in inattentional blindness. *Cognition*, 102(3):321–340. (Cited on page 73)

[Casiez et al., 2012] Casiez, G., Roussel, N., and Vogel, D. (2012). 1€ Filter: a Simple Speed-Based Low-Pass Filter for Noisy Input in Interactive Systems. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2527–2530. (Cited on page 213)

[Cennini et al., 2010] Cennini, G., Arguel, J., Akşit, K., and van Leest, A. (2010). Heart rate monitoring via remote photoplethysmography with motion artifacts reduction. *Optics express*, 18(5):4867–75. (Cited on page 219)

[Chalder et al., 1993] Chalder, T., Berelowitz, G., and Pawlikowska, T. (1993). Development of a fatigue scale. *J Psychosom Res*, 37(2):147–153. (Cited on page 13)

[Champion et al., 2004] Champion, R. a., Simmons, D. R., and Mamassian, P. (2004). The influence of object size and surface shape on shape constancy from stereo. *Perception*, 33(2):237–247. (Cited on page 38)

[Chanel and Mühl, 2015] Chanel, G. and Mühl, C. (2015). Connecting Brains and Bodies: Applying Physiological Computing to Support Social Interaction. *Interacting with Computers*. (Cited on pages 112 and 147)

[Chanel et al., 2011] Chanel, G., Rebetez, C., Bétrancourt, M., and Pun, T. (2011). Emotion assessment from physiological signals for adaptation of game difficulty. *IEEE T Syst. Man Cy. A*, 41(6):1052–1063. (Multiple combos on pages 19, 21, 24, and 25)

[Chavarriaga et al., 2010] Chavarriaga, R., Biasiucci, A., Forster, K., Roggen, D., Troster, G., and Millan, J. D. R. (2010). Adaptation of hybrid human-computer interaction systems using EEG error-related potentials. *IEEE Eng Med Biol Soc Ann*, 2010:4226–9. (Cited on page 17)

[Chavarriaga and Millan, 2010] Chavarriaga, R. and Millan, J. D. R. (2010). Learning from EEG error-related potentials in noninvasive brain-computer interfaces. *IEEE Trans. Neural Syst. Rehabil. Eng.*, 18(4):381–8. (Cited on page 16)

[Chavarriaga et al., 2014] Chavarriaga, R., Sobolewski, A., and Millán, J. D. R. (2014). Errare machinale est: The use of error-related potentials in brain-machine interfaces. *Frontiers in Neuroscience*, 8(8 JUL):1–13. (Cited on page 17)

[Chen et al., 2013] Chen, C., Li, K., Wu, Q., Wang, H., Qian, Z., and Sudlow, G. (2013). EEG-based detection and evaluation of fatigue caused by watching 3DTV. *Displays*, 34(2):81–88. (Cited on page 37)

[Chen and Bargh, 1999] Chen, M. and Bargh, J. a. (1999). Consequences of Automatic Evaluation: Immediate Behavioral Predispositions to Approach or Avoid the Stimulus. *Personality and Social Psychology Bulletin*, 25(2):215–224. (Cited on page 18)

[Chen et al., 2014] Chen, Y.-H., de Beeck, M., Vanderheyden, L., Carrette, E., Mihajlović, V., Vanstreels, K., Grundlehner, B., Gadeyne, S., Boon, P., and Van Hoof, C. (2014). Soft, Comfortable Polymer Dry Electrodes for High Quality ECG and EEG Recording. *Sensors*, 14:23758–23780. (Cited on page 208)

[Chi et al., 2010] Chi, Y. M., Ng, P., Kang, E., Kang, J., Fang, J., and Cauwenberghs, G. (2010). Wireless non-contact cardiac and neural monitoring. *WH '10*, page 15. (Cited on page 202)

[Cho et al., 2012] Cho, H., Kang, M.-K., Yoon, K.-J., and Jun, S. C. (2012). Feasibility study for visual discomfort assessment on stereo images using EEG. *2012 International Conference on 3D Imaging (IC3D)*, pages 1–6. (Cited on page 37)

[Cohen, 2001] Cohen, J. (2001). Defining Identification: A Theoretical Look at the Identification of Audiences with Media Characters. *Mass Communication & Society*, 4(3):245–264. (Cited on page 140)

[Cooney et al., 2014] Cooney, M. D., Nishio, S., and Ishiguro, H. (2014). Designing Robots for Well-being: Theoretical Background and Visual Scenes of Affectionate Play with a Small Humanoid Robot Lovotics Designing Robots for Well-being. *Lovotics*, 1. (Cited on page 148)

[Coull, 1998] Coull, J. T. (1998). Neural correlates of attention and arousal: insights from electrophysiology, functional neuroimaging and psychopharmacology. *Progress in neurobiology*, 55(4):343–61. (Cited on page 78)

[Cowley et al., 2009] Cowley, B., Charles, D., Black, M., and Hickey, R. (2009). Analyzing player behavior in Pacman using feature-driven decision theoretic predictive modeling. *CIG '09*, pages 170–177. (Cited on page 3)

[Cutellic and Lotte, 2013] Cutellic, P. and Lotte, F. (2013). Integrating Neural Activity in Evolutionary Computation for Design. In *eCAADe 2013*, volume 2013. (Cited on page 31)

[Cutrell and Tan, 2008] Cutrell, E. and Tan, D. (2008). BCI for passive input in HCI. *CHI '08 Workshops*, pages 1–3. (Cited on page xviii)

[Damasio, 1994] Damasio, A. R. (1994). *Descartes' error: emotion, reason, and the human brain*. Putnam Publishing. (Cited on pages 17 and 139)

[Dang and André, 2010] Dang, C. T. and André, E. (2010). Surface-poker: multimodality in tabletop games. In *ITS '10*, pages 251–252. (Cited on page 113)

[David Hairston et al., 2014] David Hairston, W., Whitaker, K. W., Ries, A. J., Vettel, J. M., Cortney Bradford, J., Kerick, S. E., and McDowell, K. (2014). Usability of four commercially-oriented EEG systems. *Journal of neural engineering*, 11(4):046018. (Cited on pages 179 and 209)

[de Kort, 2007] de Kort, Y. A. W. (2007). Digital Games as Social Presence Technology : Development of the Social Presence in Gaming Questionnaire (SPGQ). *Presence*, pages 1–9. (Cited on page 118)

[Degen and Jäckel, 2006] Degen, T. and Jäckel, H. (2006). A pseudodifferential amplifier for bioelectric events with DC-offset compensation using two-wired amplifying electrodes. *IEEE transactions on biomedical engineering*, 53(2):300–10. (Cited on page 202)

[Degen et al., 2007] Degen, T., Torrent, S., and Jäckel, H. (2007). Low-noise two-wired buffer electrodes for bioelectric amplifiers. *IEEE transactions on bio-medical engineering*, 54(7):1328–32. (Cited on page 204)

[Dekker et al., 2012] Dekker, S., Lee, N. C., Howard-Jones, P., and Jolles, J. (2012). Neuromyths in education: Prevalence and predictors of misconceptions among teachers. *Frontiers in Psychology*, 3(OCT):1–8. (Cited on page 127)

[Delorme and Makeig, 2004] Delorme, A. and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of neuroscience methods*. (Cited on page 61)

[Deterding et al., 2011] Deterding, S., Dixon, D., Khaled, R., and Nacke, L. (2011). From game design elements to gamefulness. In *MindTrek '11*, page 9, New York, New York, USA. ACM Press. (Cited on page 74)

[Dirican and Göktürk, 2011] Dirican, A. C. and Göktürk, M. (2011). Psychophysiological measures of human cognitive states applied in human computer interaction. *Procedia Computer Science*, 3:1361–1367. (Cited on page 4)

[Dittner et al., 2004] Dittner, a. J., Wessely, S. C., and Brown, R. G. (2004). The assessment of fatigue: a practical guide for clinicians and researchers. *J Psychosom Res*, 56(2):157–70. (Cited on page 13)

[Dodgson, 2004] Dodgson, N. A. (2004). Variation and extrema of human interpupillary distance. *Proc. SPIE*, 5291(January):36–46. (Cited on page 39)

[Dourish and Bellotti, 1992] Dourish, P. and Bellotti, V. (1992). Awareness and Coordination in Shared Workspaces. In *CSCW '92*, pages 107–114. (Cited on page xx)

[Drachen et al., 2009] Drachen, A., Canossa, A., and Yannakakis, G. N. (2009). Player modeling using self-organization in Tomb Raider: Underworld. *CIG '09*, pages 1–8. (Cited on page 3)

[Drossis et al., 2013] Drossis, G., Grammenos, D., Adami, I., and Stephanidis, C. (2013). 3D Visualization and Multimodal Interaction with Temporal Information Using Timelines. In Kotzé, P., Marsden, G., Lindgaard, G., Wesson, J., and Winckler, M., editors, *INTERACT '13*, volume 8119 of *Lecture Notes in Computer Science*, Berlin, Heidelberg. Springer Berlin Heidelberg. (Cited on page 35)

[Etienne et al., 2014] Etienne, S., Guthrie, M., Goillandeau, M., Nguyen, T. H., Orignac, H., Gross, C., and Boraud, T. (2014). Easy Rider: Monkeys Learn to Drive a Wheelchair to Navigate through a Complex Maze. *PLoS ONE*, 9(5):e96275. (Cited on page 230)

[Evans and Wobbrock, 2012] Evans, A. and Wobbrock, J. (2012). Taming wild behavior. In *CHI '12*, page 1947, New York, New York, USA. ACM Press. (Cited on page 3)

[Fabiani et al., 2007] Fabiani, M., Gratton, G., and Federmeier, K. D. (2007). Event-related brain potentials: Methods, theory, and applications. In *Handbook of psychophysiology*, chapter 12, pages 85–119. Cambridge University Press. (Cited on page 78)

[Fairclough, 2001] Fairclough, S. H. (2001). *Mental effort and sustained perceptual-motor performance*. PhD thesis, Loughborough University. (Cited on page 11)

[Fairclough, 2009] Fairclough, S. H. (2009). Fundamentals of physiological computing. *Interacting with Computers*, 21(1-2):133–145. (ULTRA combos on pages xviii, 2, 4, 21, 23, and 145)

[Fairclough, 2014] Fairclough, S. H. (2014). Human Sensors - Perspectives on the Digital Self. (Cited on page 110)

[Fairclough and Houston, 2004] Fairclough, S. H. and Houston, K. (2004). A metabolic measure of mental effort. *Biological psychology*, 66(2):177–90. (Cited on pages xviii and 11)

[Falkenstein et al., 2000] Falkenstein, M., Hoormann, J., Christ, S., and Hohnsbein, J. (2000). ERP components on reaction errors and their functional significance: a tutorial. *Biological Psychology*, 51(2-3):87–107. (Cited on page 16)

[Farb et al., 2015] Farb, N., Daubenmier, J., Price, C. J., Gard, T., Kerr, C., Dunn, B. D., Klein, A. C., Paulus, M. P., and Mehling, W. E. (2015). Interoception, contemplative practice, and health. *Frontiers in Psychology*, 6(June):763. (Cited on page 175)

[Fatourechi et al., 2007] Fatourechi, M., Bashashati, A., Ward, R. K., and Birch, G. E. (2007). EMG and EOG artifacts in brain computer interface systems: A survey. *Clinical neurophysiology : official journal of the International Federation of Clinical Neurophysiology*, 118(3):480–94. (Cited on pages 42, 61, and 133)

[Fedota and Parasuraman, 2010] Fedota, J. R. and Parasuraman, R. (2010). Neuroergonomics and human error. *Theoretical Issues in Ergonomics Science*, 11(5):402–421. (Cited on page 16)

[Ferree et al., 2001] Ferree, T. C., Luu, P., Russell, G. S., and Tucker, D. M. (2001). Scalp electrode impedance, infection risk, and EEG data quality. *Clinical Neurophysiology*, 112(3):536–544. (Cited on page 202)

[Ferrez and Millan, 2008] Ferrez, P. W. and Millan, J. D. R. (2008). Error-related EEG potentials generated during simulated brain-computer interaction. *IEEE Trans. Biomed. Eng.*, 55(3):923–9. (Multiple combos on pages 15, 16, 78, 82, and 93)

[Figner and Murphy, 2011] Figner, B. and Murphy, R. O. (2011). Using skin conductance in judgment and decision making research. In *A handbook of process tracing methods for decision research*, pages 163–184. Psychology Press, New York, NY. (Cited on pages 62, 63, and 155)

[Fitts, 1954] Fitts, P. M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *Journal of experimental psychology. General*, 47(6):381–391. (Cited on page 3)

[Fitzpatrick, 1988] Fitzpatrick, T. B. (1988). The validity and practicality of sun-reactive skin types I through VI. *Archives of dermatology*, 124(6):869–871. (Cited on page 214)

[Flatla et al., 2011] Flatla, D. R., Gutwin, C., Nacke, L. E., Bateman, S., and Mandryk, R. L. (2011). Calibration games: making calibration tasks enjoyable by adding motivating game elements. *UIST '11*, pages 403–412. (Cited on pages 74 and 94)

[Fleck and Simon, 2013] Fleck, S. and Simon, G. (2013). An Augmented Reality Environment for Astronomy Learning in Elementary Grades: An Exploratory Study. In *IHM '13*, pages 14–22. (Cited on page 128)

[Fraden, 2010] Fraden, J. (2010). *Handbook of Modern Sensors: Physics, Designs, and Applications, 2nd ed.* Springer. (Cited on page 208)

[Frey, 2015] Frey, J. (2015). Heart Rate Monitoring as an Easy Way to Increase Engagement in Human-Agent Interaction. In *PhyCS - International Conference on Physiological Computing Systems*. (Cited on page 96)

[Frey, 2016a] Frey, J. (2016a). Comparison of a consumer grade EEG amplifier with medical grade equipment in BCI applications. In *International BCI meeting*. (Cited on page 182)

[Frey, 2016b] Frey, J. (2016b). Remote Heart Rate Sensing and Projection to Renew Traditional Board Games and Foster Social Interactions. In *Proceedings of the extended abstracts of the 34th annual ACM conference on Human factors in computing systems - CHI EA '16*. (Cited on page 111)

[Frey et al., 2015] Frey, J., Appriou, A., Lotte, F., and Hachet, M. (2015). Estimating Visual Comfort in Stereoscopic Displays Using Electroencephalography: A Proof-of-Concept. In *INTERACT '15*. (Cited on pages 21, 28, and 35)

[Frey et al., 2016a] Frey, J., Appriou, A., Lotte, F., and Hachet, M. (2016a). Classifying EEG Signals during Stereoscopic Visualization to Estimate Visual Comfort. *Computational Intelligence and Neuroscience*, 2016:1–11. (Cited on pages 28 and 35)

[Frey et al., 2016b] Frey, J., Daniel, M., Castet, J., Hachet, M., and Lotte, F. (2016b). Framework for electroencephalography-based evaluation of user experience. In *Proceedings of the 34th annual ACM conference on Human factors in computing systems - CHI '16*. (Cited on page 73)

[Frey et al., 2014a] Frey, J., Gervais, R., Fleck, S., Lotte, F., and Hachet, M. (2014a). Teegi: Tangible EEG Interface. In *UIST-ACM User Interface Software and Technology Symposium*. (Cited on pages 122, 124, and 141)

[Frey et al., 2014b] Frey, J., Mühl, C., Lotte, F., and Hachet, M. (2014b). Review of the use of electroencephalography as an evaluation method for human-computer interaction. In *PhyCS - International Conference on Physiological Computing Systems*. (Cited on pages 1 and 145)

[Frey et al., 2014c] Frey, J., Pommereau, L., Lotte, F., and Hachet, M. (2014c). Assessing the zone of comfort in stereoscopic displays using EEG. In *Proceedings of the extended abstracts of the 32nd annual ACM conference on Human factors in computing systems - CHI EA '14*, pages 2041–2046, New York, New York, USA. ACM Press. (Multiple combos on pages 21, 28, 35, and 37)

[Friedman and Thayer, 1991] Friedman, B. H. and Thayer, J. F. (1991). Facial muscle activity and EEG recordings: redundancy analysis. *Electroencephalography and Clinical Neurophysiology*, 79(5):358–360. (Cited on page 50)

[Friedman, 1997] Friedman, J. (1997). On bias, variance, 0/1—loss, and the curse-of-dimensionality. *Data mining and knowledge discovery*, 77:55–77. (Cited on pages 26 and 50)

[Frohlich et al., 1999] Frohlich, B., Barrass, S., Zehner, B., Plate, J., and Gobel, M. (1999). Exploring geo-scientific data in virtual environments. In *Visualization '99*, pages 169–173. IEEE. (Cited on page 35)

[Fukushima et al., 2011] Fukushima, H., Terasawa, Y., and Umeda, S. (2011). Association between interoception and empathy: evidence from heartbeat-evoked brain potential. *International journal of psychophysiology : official journal of the International Organization of Psychophysiology*, 79(2):259–65. (Cited on page 110)

[Gadia et al., 2014] Gadia, D., Garipoli, G., Bonanomi, C., Albani, L., and Rizzi, A. (2014). Assessing stereo blindness and stereo acuity on digital displays. *Displays*, 35(4):206–212. (Cited on page 41)

[Gaeblerlabel et al., 2014] Gaeblerlabel, M., Biessmannlabel, F., Lamke, J.-P., Müller, K.-R., Walter, H., and Hetzer, S. (2014). Stereoscopic depth increases intersubject correlations of brain networks. *NeuroImage*. (Cited on page 36)

[Gargiulo et al., 2011] Gargiulo, G., Bifulco, P., Calvo, R. A., Romano, M., Ruffo, M., and Shephard, R. (2011). Giga-ohm high impedance FET input amplifiers for dry electrode biosensor circuits and systems. In Iniewski, K., editor, *Integrated Microsystems: Electronics, Photonics, and Biotechnology*, chapter 8, pages 165–194. CRC press. (Cited on page 208)

[George and Lécuyer, 2010] George, L. and Lécuyer, A. (2010). An overview of research on'passive'brain-computer interfaces for implicit human-computer interaction. In *ICABB 2010*. (Multiple combos on pages xviii, 10, 23, and 30)

[George et al., 2011] George, L., Lotte, F., Abad, R. V., and Lécuyer, A. (2011). Using Scalp Electrical Biosignals to Control an Object by Concentration and Relaxation Tasks: Design and Evaluation. In *IEEE EMBS 2011*. (Cited on page 14)

[Gerjets et al., 2014] Gerjets, P., Walter, C., Rosenstiel, W., Bogdan, M., and Zander, T. O. (2014). Cognitive state monitoring and the design of adaptive instruction in digital environments: lessons learned from cognitive workload assessment using a passive brain-computer interface approach. *Frontiers in Neuroscience*, 8. (Cited on page 172)

[Gerlinghaus et al., 2012] Gerlinghaus, F., Pierce, B., Metzler, T., Jowers, I., Shea, K., and Cheng, G. (2012). Design and emotional expressiveness of Gertie (An open hardware robotic desk lamp). *IEEE RO-MAN '12*, pages 1129–1134. (Cited on page 109)

[Gervais et al., 2016] Gervais, R., Frey, J., Gay, A., Lotte, F., and Hachet, M. (2016). Tobe: Tangible Out-of-Body Experience. In *TEI '16*. (Cited on page 145)

[Gervais et al., 2015] Gervais, R., Frey, J., and Hachet, M. (2015). Pointing in Spatial Augmented Reality from 2D Pointing Devices. In *INTERACT '15*. (Cited on page 3)

[Ghaderi et al., 2013] Ghaderi, F., Kim, S. K., and Kirchner, E. A. (2013). Effects of eye artifact removal methods on single trial P300 detection, a comparative study. *Journal of neuroscience methods*, 221C:41–47. (Cited on page 42)

[Giacometti and Diamond, 2013] Giacometti, P. and Diamond, S. G. (2013). Compliant head probe for positioning electroencephalography electrodes and near-infrared spectroscopy optodes. *Journal of biomedical optics*, 18(2):27005. (Cited on page 208)

[Gilleade and Dix, 2004] Gilleade, K. M. and Dix, A. (2004). Using frustration in the design of adaptive videogames. *ACE '04*, pages 228–232. (Cited on pages 21 and 22)

[Golsteijn et al., 2012] Golsteijn, C., van den Hoven, E., Frohlich, D., and Sellen, A. (2012). Towards a more cherishable digital object. In *Proceedings of the Designing Interactive Systems Conference on - DIS '12*, page 655. (Cited on page 146)

[Goncharova et al., 2003] Goncharova, I., McFarland, D., Vaughan, T., and Wolpaw, J. (2003). EMG contamination of EEG: spectral and topographical characteristics. *Clinical Neurophysiology*, 114(9):1580–1593. (Cited on page 61)

[Grimes et al., 2008] Grimes, D., Tan, D., and Hudson, S. (2008). Feasibility and pragmatics of classifying working memory load with an electroencephalograph. *CHI '08*, page 835. (Multiple combos on pages 12, 58, 75, and 76)

[Grubbs, 1969] Grubbs, F. E. (1969). Procedures for Detecting Outlying Observations in Samples. *Technometrics*, 11(1):1–21. (Cited on page 87)

[Gunawardena, 1995] Gunawardena, C. (1995). Social presence theory and implications of interaction and collaborative learning in computer conferencing. *International Journal of Educational Telecommunications*, 1(2-3):147–166. (Cited on page xix)

[Guyon, 2003] Guyon, I. (2003). An Introduction to Variable and Feature Selection. *Journal of Machine Learning Research*, 3:1157–1182. (Cited on page 47)

[Haas, 2003] Haas, L. F. (2003). Hans Berger (1873-1941), Richard Caton (1842-1926), and electroencephalography. *Journal of neurology, neurosurgery, and psychiatry*, 74(1):9. (Cited on page 208)

[Hachet et al., 2011] Hachet, M., Bossavit, B., Cohé, A., and de la Rivière, J.-B. (2011). Toucheo: Multitouch and Stereo Combined in a Seamless Workspace. In *UIST '11*, page 587, New York, New York, USA. ACM Press. (Cited on page 35)

[Hamadicharef, 2010] Hamadicharef, B. (2010). Brain-Computer Interface (BCI) literature - a bibliometric study. In *ISSPA '10*, pages 626–629. IEEE. (Cited on page 5)

[Hancock et al., 2005] Hancock, P. A., Pepe, A. A., and Murphy, L. L. (2005). Hedonomics: The power of positive and pleasurable ergonomics. *Ergonomics in design*. (Cited on page 17)

[Harrison et al., 2012] Harrison, C., Horstman, J., Hsieh, G., and Hudson, S. (2012). Unlocking the expressivity of point lights. In *CHI '12*, page 1683, New York, New York, USA. ACM Press. (Cited on page 99)

[Hart and Staveland, 1988] Hart, S. G. and Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In *Human mental workload*, pages 139–183. Elsevier B.V. (ULTRA combos on pages xvii, 3, 11, 53, 74, 77, and 80)

[Hassenzahl and Sandweg, 2004] Hassenzahl, M. and Sandweg, N. (2004). From mental effort to perceived usability: transforming experiences into summary assessments. *CHI '04*, pages 1283–1286. (Cited on page 21)

[Hattie and Timperley, 2007] Hattie, J. and Timperley, H. (2007). The Power of Feedback. *Review of Educational Research*, 77(1):81–112. (Cited on page 223)

[He et al., 2009] He, W., Yamashita, T., Lu, H., and Lao, S. (2009). SURF tracking. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1586–1592. (Cited on page 115)

[Hedman et al., 2012] Hedman, E., Miller, L., Schoen, S., and Nielsen, D. (2012). Measuring Autonomic Arousal During Therapy. In *Proceedings of 8th International Design and Emotion Conference London*, pages 11–14. (Cited on page 163)

[Heingartner, 2009] Heingartner, D. (2009). Mental block. *IEEE Spectrum*, 46(1):42–43. (Cited on page 27)

[Herculano-Houzel, 2002] Herculano-Houzel, S. (2002). Do you know your brain? A survey on public neuroscience literacy at the closing of the decade of the brain. *The Neuroscientist: a review journal bringing neurobiology, neurology and psychiatry*, 8(2):98–110. (Cited on page 127)

[Hernandez and Picard, 2014] Hernandez, J. and Picard, R. W. (2014). SenseGlass: Using Google Glass to Sense Daily Emotions. In *Proceedings of the adjunct publication of the 27th annual ACM symposium on User interface software and technology - UIST'14 Adjunct*, pages 77–78. (Cited on page 163)

[Hewett et al., 1992] Hewett, T. T., Baecker, R., Card, S., Carey, T., Gasen, J., Mantei, M., Perlman, G., Strong, G., and Verplank, W. (1992). *ACM SIGCHI Curricula for Human-Computer Interaction*. ACM, New York. (Cited on page xvii)

[Hinckley et al., 1994] Hinckley, K., Pausch, R., Goble, J. C., and Kassell, N. F. (1994). Passive Real-world Interface Props for Neurosurgical Visualization. In *Proceedings of the SIGCHI conference on Human factors in computing systems celebrating interdependence - CHI '94*, volume 30, pages 452–458. (Cited on page 128)

[Hirshfield et al., 2009a] Hirshfield, L., Chauncey, K., and Gulotta, R. (2009a). Combining electroencephalograph and functional near infrared spectroscopy to explore users' mental workload. *FAC '09*. (Cited on page 12)

[Hirshfield et al., 2009b] Hirshfield, L. M., Solovey, E. T., Girouard, A., Kebinger, J., Jacob, R. J., Sassaroli, A., and Fantini, S. (2009b). Brain measurement for usability testing and adaptive interfaces. In *CHI '09*, page 2185, New York, New York, USA. ACM Press. (Cited on pages 12 and 21)

[Hix et al., 1999] Hix, D., Swan, J., Gabbard, J., McGee, M., Durbin, J., and King, T. (1999). User-centered design and evaluation of a real-time battlefield visualization virtual environment. In *IEEE Virt. Real.*, pages 96–103. IEEE Comput. Soc. (Cited on page 20)

[Hochberg et al., 2006] Hochberg, L. R., Serruya, M. D., Friehs, G. M., Mukand, J. a., Saleh, M., Caplan, A. H., Branner, A., Chen, D., Penn, R. D., and Donoghue, J. P. (2006). Neuronal ensemble control of prosthetic devices by a human with tetraplegia. *Nature*, 442(7099):164–171. (Cited on page 29)

[Hoffman et al., 2008] Hoffman, D. M., Girshick, A. R., Akeley, K., and Banks, M. S. (2008). Vergence-accommodation conflicts hinder visual performance and cause visual fatigue. *Journal of vision*, 8(3):33.1–30. (Cited on page 36)

[Hoffmann et al., 2006] Hoffmann, U., Vesin, J., and Ebrahimi, T. (2006). Spatial filters for the classification of event-related potentials. In *ESANN '06*. (Cited on pages 44 and 86)

[Holroyd and Coles, 2002] Holroyd, C. B. and Coles, M. G. (2002). The neural basis of human error processing: Reinforcement learning, dopamine, and the error-related negativity. *Psychological Review*, 109(4):679–709. (Cited on page 16)

[Horn et al., 2009] Horn, M. S., Solovey, E. T., Crouser, R. J., and Jacob, R. J. K. (2009). Comparing the use of tangible and graphical programming languages for informal science education. In *CHI '09*, volume 32, page 975. (Cited on pages 128 and 148)

[Hornecker, 2011] Hornecker, E. (2011). The role of physicality in tangible and embodied interactions. *Interactions*, 18(2):19. (Cited on page 148)

[Huang et al., 2014] Huang, W., Liu, G., and Wen, W. (2014). MAPD: A Multi-subject Affective Physiological Database. In *2014 Seventh International Symposium on Computational Intelligence and Design*, pages 585–589. IEEE. (Cited on page 173)

[Huppi et al., 2003] Huppi, B. Q., Stringer, C. J., Bell, J., and Capener, C. J. (2003). United States Patent 6658577: Breathing status LED indicator. (Cited on page 99)

[Hurtienne et al., 2014] Hurtienne, J., Landeck, M., Ludwig, S., and Löffler, D. (2014). Beyond eye tracking analogies. In *CHI EA '14*, pages 1789–1794. (Cited on page 18)

[Ishii and Ullmer, 1997] Ishii, H. and Ullmer, B. (1997). Tangible Bits: Towards Seamless Interfaces between People, Bits and Atoms. In *CHI '97*, volume 39, pages 234–241, New York, New York, USA. ACM Press. (Cited on page 128)

[ITU, 2003] ITU (2003). P. 851, Subjective Quality Evaluation of Telephone Services Based on Spoken Dialogue Systems. *International Telecommunication Union, Geneva*. (Cited on page 107)

[Iturrate et al., 2010] Iturrate, I., Montesano, L., and Minguez, J. (2010). Single trial recognition of error-related potentials during observation of robot operation. *IEEE Eng. Med. Biol. Soc.*, 2010:4181–4. (Cited on page 16)

[Jankowski and Hachet, 2013] Jankowski, J. and Hachet, M. (2013). A Survey of Interaction Techniques for Interactive 3D Environments. In *Eurographics '13*. (Cited on page 53)

[Janssen et al., 2013] Janssen, J. H., Ijsselsteijn, W. a., Westerink, J. H., Tacken, P., and de Vries, G.-J. (2013). The Tell-Tale Heart: Perceived Emotional Intensity of Heartbeats. *International Journal of Synthetic Emotions*, 4(1):65–91. (Cited on page 148)

[Janssen et al., 2011] Janssen, J. H., Westerink, J. H., IJsselteijn, W. A., and van der Zwaag, M. D. (2011). The role of physiological computing in counteracting loneliness. In *CHI '11 Workshop: "Brain and Body Interfaces: Designing for Meaningful Interaction"*. (Cited on pages 113 and 114)

[Jennett, 2010] Jennett, C. (2010). *Is game immersion just another form of selective attention? An empirical investigation of real world dissociation in computer game immersion*. PhD thesis, UCL. (Cited on page 171)

[Jennett et al., 2008] Jennett, C., Cox, A. L., Cairns, P., Dhoparee, S., Epps, A., Tijs, T., and Walton, A. (2008). Measuring and defining the experience of immersion in games. *International Journal of Human Computer Studies*, 66(9):641–661. (Cited on page 22)

[Jin, 2012] Jin, S.-A. A. (2012). The virtual malleable self and the virtual identity discrepancy model: Investigative frameworks for virtual possible selves and others in avatar-based identity construction and social interaction. *Computers in Human Behavior*, 28(6):2160–2168. (Cited on page 165)

[Just et al., 2003] Just, M. A., Carpenter, P. a., and Miyake, A. (2003). Neuroindices of cognitive workload: Neuroimaging, pupillometric and event-related potential studies of brain work. *Theoretical Issues in Ergonomics Science*, 4(1-2):56–88. (Cited on pages 10 and 11)

[Karavidas et al., 2010] Karavidas, M. K., Lehrer, P. M., Lu, S.-E., Vaschillo, E., Vaschillo, B., and Cheng, A. (2010). The effects of workload on respiratory variables in simulated flight: a preliminary study. *Biological psychology*, 84(1):157–60. (Cited on page 4)

[Karlesky and Isbister, 2014] Karlesky, M. and Isbister, K. (2014). Designing for the Physical Margins of Digital Workspaces: Fidget Widgets in Support of Productivity and Creativity. In *TEI '14*. (Cited on page 98)

[Karnik et al., 2013] Karnik, A., Grossman, T., and Subramanian, S. (2013). Comparison of User Performance in Mixed 2D-3D Multi-Display Environments. In *INTERACT '13*, volume 8117 of *Lecture Notes in Computer Science*. (Cited on page 53)

[Khot et al., 2014] Khot, R. A., Hjorth, L., and Mueller, F. F. (2014). Understanding physical activity through 3D printed material artifacts. *CHI '14*, pages 3835–3844. (Cited on page 147)

[Kiesler et al., 1984] Kiesler, S., Siegel, J., and McGuire, T. W. (1984). Social psychological aspects of computer-mediated communication. *American Psychologist*, 39(10):1123–1134. (Cited on page xix)

[Kitamura et al., 2003] Kitamura, Y., Yamaguchi, Y., Hiroshi, I., Kishino, F., and Kawato, M. (2003). Things happening in the brain while humans learn to use new tools. In *CHI '03*, page 417, New York, New York, USA. ACM Press. (Cited on page xviii)

[Kivikangas et al., 2010] Kivikangas, J. M., Ekman, I., Chanel, G., Järvelä, S., Cowley, B., Henttonen, P., and Ravaja, N. (2010). Review on psychophysiological methods in game research. *Proc. of 1st Nordic DiGRA*. (ULTRA combos on pages xvii, xviii, 4, 12, 14, 17, 18, 19, and 121)

[Klimesch et al., 1998] Klimesch, W., Doppelmayr, M., Russegger, H., Pachinger, T., and Schwaiger, J. (1998). Induced alpha band power changes in the human EEG and attention. *Neuroscience letters*, 244(2):73–6. (Cited on page 14)

[Kohlmorgen et al., 2007] Kohlmorgen, J., Dornhege, G., Braun, M., Blankertz, B., Müller, K.-R., Curio, G., Hagemann, K., Bruns, A., Schrauf, M., and Kincses, W. (2007). Improving human performance in a real operating environment through real-time mental workload detection. In *Toward Brain-Computer Interfacing*. MIT Press. (Cited on page 2)

[Kranjec et al., 2014] Kranjec, J., Beguš, S., Geršak, G., and Drnovšek, J. (2014). Non-contact heart rate and heart rate variability measurements: A review. *Biomedical Signal Processing and Control*, 13:102–112. (Cited on pages 99, 152, and 212)

[Kuwabara et al., 2002] Kuwabara, K., Watanabe, T., Ohguro, T., Itoh, Y., and Maeda, Y. (2002). Connectedness oriented communication: fostering a sense of connectedness to augment social relationships. *IEEE SAINT '02 - Proceedings 2002 Symposium on Applications and the Internet*, pages 186–193. (Cited on page xx)

[Lachaux et al., 2000] Lachaux, J.-P., Rodriguez, E., Le Van Quyen, M., Lutz, A., Martinerie, J., and Varela, F. J. (2000). Studying single-trials of phase synchronous activity in the brain. *International Journal of Bifurcation and Chaos*, 10(10):2429–2439. (Cited on pages 133 and 154)

[Lambooij et al., 2009] Lambooij, M., IJsselsteijn, W., Fortuin, M., and Heynderickx, I. (2009). Visual Discomfort and Visual Fatigue of Stereoscopic Displays: A Review. *Journal of Imaging Science and Technology*, 53(3):030201. (Cited on pages 35 and 36)

[Latour et al., 2012] Latour, B., Jensen, P., Venturini, T., Grauwin, S., and Boullier, D. (2012). The whole is always smaller than its parts: a digital test of Gabriel Tardes' monads. *The British journal of sociology*, 63(4):590–615. (Cited on page 23)

[Laurent et al., 2013] Laurent, F., Valderrama, M., Besserve, M., Guillard, M., Lachaux, J.-P., Martinerie, J., and Florence, G. (2013). Multimodal information improves the rapid detection of mental fatigue. *Biomedical Signal Processing and Control*, pages 1–9. (Cited on page 14)

[Laviole and Hachet, 2012] Laviole, J. and Hachet, M. (2012). PapARt: Interactive 3D graphics and multi-touch augmented paper for artistic creation. In *3DUI '12*, pages 3–6. (Cited on page 115)

[Le Tallec et al., 2011] Le Tallec, M., Antoine, J.-Y., Villaneau, J., and Duhaut, D. (2011). Affective interaction with a companion robot for hospitalized children: a linguistically based model for emotion detection. In *5th Language and Technology Conference (LTC'2011)*. (Cited on page 105)

[Ledoit and Wolf, 2004] Ledoit, O. and Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, 88(2):365–411. (Cited on pages 44 and 87)

[Lee and Nass, 2003] Lee, K. M. and Nass, C. (2003). Designing social presence of social actors in human computer interaction. In *Proceedings of the conference on Human factors in computing systems - CHI '03*, page 289, New York, New York, USA. ACM Press. (Cited on pages 98, 99, and 100)

[Lee et al., 2014] Lee, M., Kim, K., Rho, H., and Kim, S. J. (2014). Empa Talk: A Physiological Data Incorporated Human-Computer Interactions. In

CHI EA '14, pages 1897–1902, New York, New York, USA. ACM Press. (Cited on pages 109, 112, and 164)

[Lehmann et al., 2012] Lehmann, D., Faber, P. L., Tei, S., Pascual-Marqui, R. D., Milz, P., and Kochi, K. (2012). Reduced functional connectivity between cortical sources in five meditation traditions detected with lagged coherence using EEG tomography. *NeuroImage*, 60(2):1574–86. (Cited on pages 134 and 154)

[Lemm et al., 2005] Lemm, S., Blankertz, B., Curio, G., and Müller, K.-R. (2005). Spatio-spectral filters for improving the classification of single trial EEG. *IEEE transactions on bio-medical engineering*, 52(9):1541–8. (Cited on page 46)

[Leuthardt et al., 2006] Leuthardt, E., Miller, K., Schalk, G., Rao, R., and Ojemann, J. (2006). Electrocorticography-Based Brain Computer Interface—The Seattle Experience. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 14(2):194–198. (Cited on page 29)

[Lewis et al., 2013] Lewis, J. R., Utesch, B. S., and Maher, D. E. (2013). UMUX-LITE: when there's no time for the SUS. In *CHI '13*, page 2099, New York, New York, USA. ACM Press. (Cited on page 20)

[Li et al., 2008] Li, H.-C. O., Seo, J., Kham, K., and Lee, S. (2008). Measurement of 3D Visual Fatigue Using Event-Related Potential (ERP): 3D Oddball Paradigm. *3DTV-CON*, pages 213–216. (Cited on page 37)

[Limare et al., 2011] Limare, N., Lisani, J.-L., Morel, J.-M., Petro, A. B., and Sbert, C. (2011). Simplest Color Balance. *Image Processing On Line*. (Cited on page 213)

[Lisetti and Nasoz, 2004] Lisetti, C. L. and Nasoz, F. (2004). Using Non-invasive Wearable Computers to Recognize Human Emotions from Physiological Signals. *EURASIP J ADV SIG PR*, 2004(11):1672–1687. (Cited on pages 19 and 98)

[Liu et al., 2011] Liu, Y., Sourina, O., and Nguyen, M. (2011). Real-time EEG-based emotion recognition and its applications. In *Transactions on computational science XII*, pages 256–277. Springer Berlin Heidelberg. (Cited on page 19)

[Loggia et al., 2011] Loggia, M. L., Juneau, M., and Bushnell, M. C. (2011). Autonomic responses to heat pain: Heart rate, skin conductance, and their relation to verbal ratings and stimulus intensity. *Pain*, 152(3):592–8. (Cited on page 8)

[Lorist et al., 2000] Lorist, M. M., Klein, M., Nieuwenhuis, S., De Jong, R., Mulder, G., and Meijman, T. F. (2000). Mental fatigue and task control: planning and preparation. *Psychophysiology*, 37(5):614–25. (Cited on pages 13 and 14)

[Lotte, 2014] Lotte, F. (2014). A Tutorial on EEG Signal Processing Techniques for Mental State Recognition in Brain-Computer Interfaces. In *Brain-Computer Music Interfacing*. Springer. (Cited on pages 61 and 63)

[Lotte and Guan, 2011] Lotte, F. and Guan, C. (2011). Regularizing common spatial patterns to improve BCI designs: unified theory and new algorithms. *IEEE transactions on bio-medical engineering*, 58(2):355–62. (Cited on page 61)

[Lotte et al., 2008] Lotte, F., Renard, Y., and Lécuyer, A. (2008). Self-Paced Brain-Computer Interaction with Virtual Worlds: A Quantitative and Qualitative Study "Out of the Lab". In *4th international Brain Computer Interface Workshop and Training Course*. (Cited on page 30)

[Lotte et al., 2010] Lotte, F., Van Langhenhove, A., Lamarche, F., Ernest, T., Renard, Y., Arnaldi, B., and Lécuyer, A. (2010). Exploring Large Virtual Environments by Thoughts using a Brain-Computer Interface based on Motor Imagery and High-Level Commands. *Presence: Teleoperators and Virtual Environments*, 19(1):54–70. (Cited on page 30)

[Maan et al., 2010] Maan, S., Merkus, B., Ham, J., and Midden, C. (2010). Making it not too obvious: the effect of ambient light feedback on space heating energy consumption. *Energy Efficiency*, 4(2):175–183. (Cited on page 223)

[MacDorman, 2005] MacDorman, K. (2005). Androids as an experimental apparatus: Why is there an uncanny valley and can we exploit it. *CogSci-2005 workshop: toward social mechanisms of android science*, 3. (Cited on page 99)

[Mahmoud, 2008] Mahmoud, T. M. (2008). A New Fast Skin Color Detection Technique. *Engineering and Technology*, 2(7):501–505. (Cited on page 213)

[Mandryk et al., 2006] Mandryk, R., Inkpen, K., and Calvert, T. (2006). Using psychophysiological techniques to measure user experience with entertainment technologies. *Behaviour & Information Technology*. (Cited on pages 23, 24, and 98)

[Mandryk and Atkins, 2007] Mandryk, R. L. and Atkins, M. S. (2007). A fuzzy physiological approach for continuously modeling emotion during interaction with play technologies. *Int. J. Hum-Comput St.*, 65(4):329–347. (Cited on page 19)

[Mathan et al., 2007] Mathan, S., Whitlow, S., and Feyereisen, T. (2007). WorkSense: Exploring the Feasibility of Human Factors Assessment Using. Electrophysiological Sensors. In *4th Conf of the Augmented Cognition Society*. (Cited on pages 11, 12, and 25)

[Matsuzaka et al., 2012] Matsuzaka, Y., Ichihara, T., Abe, T., and Mushiake, H. (2012). Bio-amplifier with Driven Shield Inputs to Reduce Electrical Noise and its Application to Laboratory Teaching of Electrophysiology. *J Undergrad Neurosci Educ.*, 10(2):118–124. (Cited on page 208)

[Matthews et al., 2002] Matthews, G., Campbell, S. E., Falconer, S., Joyner, L. a., Huggins, J., Gilliland, K., Grier, R., and Warm, J. S. (2002). Fundamental dimensions of subjective state in performance settings: Task engagement, distress, and worry. *Emotion*, 2(4):315–340. (Cited on pages 21, 22, and 98)

[Maule and Edland, 1997] Maule, A. and Edland, A. C. (1997). The effects of time pressure on human judgment and decision making. In *Decision making: Cognitive models and Explanations*, pages 128–144. Psychology Press. (Cited on page 77)

[McCandless, 2010] McCandless, D. (2010). *Information is Beautiful*. HarperCollins. (Cited on page 145)

[McCraty et al., 2009] McCraty, R., Atkinson, M., Tomasino, D., and Bradley, R. T. (2009). The coherent heart: Heart-brain interactions, psychophysiological coherence, and the emergence of system-wide order. *Integral Review*, 5(2):10–115. (Cited on page 160)

[Mehler et al., 2011] Mehler, B., Reimer, B., and Wang, Y. (2011). A comparison of heart rate and heart rate variability indices in distinguishing single task driving and driving under secondary cognitive workload. *... on Human Factors ...*, pages 590–597. (Cited on pages 62 and 65)

[Mehta and Parasuraman, 2013] Mehta, R. K. and Parasuraman, R. (2013). Neuroergonomics: a review of applications to physical and cognitive work. *Frontiers in human neuroscience*, 7(December):889. (Cited on pages xviii and 5)

[Mercier-ganady et al., 2014] Mercier-ganady, J., Lotte, F., Loup-escande, E., Marchal, M., and Lécuyer, A. (2014). The Mind-Mirror: See Your Brain in Action in Your Head Using EEG and Augmented Reality. In *VR*. (Cited on pages 127 and 129)

[Mikkola et al., 2010] Mikkola, M., Boev, A., and Gotchev, A. (2010). Relative importance of depth cues on portable autostereoscopic display. In *MoViD '10*, page 63. (Cited on page 38)

[Milekovic et al., 2013] Milekovic, T., Ball, T., Schulze-Bonhage, A., Aertsen, A., and Mehring, C. (2013). Detection of error related neuronal responses recorded by electrocorticography in humans during continuous movements. *PloS one*, 8(2):e55235. (Cited on page 17)

[Mirza-Babaei et al., 2013] Mirza-Babaei, P., Nacke, L., and Gregory, J. (2013). How Does It Play Better? Exploring User Testing and Biometric Storyboards in Games User Research. In *CHI '13*, pages 1499–1508. (Cited on page 16)

[Mognon et al., 2010] Mognon, A., Jovicich, J., Bruzzone, L., and Buiatti, M. (2010). ADJUST: An automatic EEG artifact detector based on the joint use of spatial and temporal features. *Psychophysiology*, 48:229–240. (Cited on page 42)

[Molina et al., 2009] Molina, G. G., Tsoneva, T., and Nijholt, A. (2009). Emotional brain-computer interfaces. *International Journal of Autonomous and Adaptive Communications Systems*, 6(1):9. (Cited on pages 19 and 154)

[Möller et al., 2007] Möller, S., Smeele, P., Boland, H., and Krebber, J. (2007). Evaluating spoken dialogue systems according to de-facto standards: A case study. *Computer Speech & Language*, 21(1):26–53. (Cited on page 107)

[Momeni-Moghadam et al., 2012] Momeni-Moghadam, H., Kundart, J., Ehsani, M., and Gholami, K. (2012). Stereopsis with TNO and Titmus Tests in Symptomatic and Asymptomatic University Students. *Journal of Behavioral Optometry*, 23(2):35–39. (Cited on page 41)

[Motti and Caine, 2015] Motti, V. and Caine, K. (2015). Users' Privacy Concerns About Wearables: impact of form factor, sensors and type of data collected. *workshop co-located with Financial Crypto 2015*. (Cited on page 174)

[Mühl et al., 2011] Mühl, C., Brouwer, A., van Wouwe, N., van den Broek, E. L., Nijboer, F., and Heylen, D. (2011). Modality-specific Affective Responses and their Implications for Affective BCI. In *Proceedings of the Fifth International Brain-Computer Interface Conference 2011*, pages 120–123. Verlag der Technischen Universität. (Cited on page 19)

[Mühl et al., 2014] Mühl, C., Jeunet, C., and Lotte, F. (2014). EEG-based workload estimation across affective contexts. *Frontiers in Neuroscience*, 8(8 JUN):1–15. (Cited on pages 54, 58, and 61)

[Mullen et al., 2013] Mullen, T., Kothe, C., Chi, Y. M., Ojeda, A., Kerth, T., Makeig, S., Cauwenberghs, G., and Tzyy-Ping Jung (2013). Real-time modeling and 3D visualization of source dynamics and connectivity using wearable EEG. In *EMBC '13*, pages 2184–2187. IEEE. (Cited on page 127)

[Müller-Putz et al., 2008] Müller-Putz, G., Scherer, R., Brunner, C., Leeb, R., and Pfurtscheller, G. (2008). Better than random? A closer look on BCI results. *International Journal of Bioelectromagnetism*, 10(1):52–55. (Cited on pages 48, 64, and 207)

[Mustafa et al., 2012] Mustafa, M., Lindemann, L., and Magnor, M. (2012). EEG analysis of implicit human visual perception. *CHI '12*, page 513. (Cited on pages 14 and 25)

[Nacke et al., 2009] Nacke, L., Ambinder, M., Canossa, A., Mandryk, R., and Stach, T. (2009). Game Metrics and Biometrics: The Future of Player Experience Research. *Panel at Future Play 2009*. (Cited on page 4)

[Nacke and Lindley, 2009] Nacke, L. E. and Lindley, C. A. (2009). Affective ludology, flow and immersion in a first-person shooter: Measurement of player experience. *The Journal of the Canadian Game Studies Association*, 3(5). (Multiple combos on pages 21, 22, 23, and 24)

[Nacke et al., 2010] Nacke, L. E., Stellmach, S., and Lindley, C. A. (2010). Electroencephalographic Assessment of Player Experience: A Pilot Study in Affective Ludology. *Simulation & Gaming*, 42(5):632–655. (Cited on pages 22 and 23)

[Nathan and Jafari, 2014] Nathan, V. and Jafari, R. (2014). Characterizing Contact Impedance, Signal Quality and Robustness as a Function of the Cardinality and Arrangement of Fingers on Dry Contact EEG Electrodes. In *EMBC '14*. (Cited on page 203)

[Niedermeyer and da Silva, 2005] Niedermeyer, E. and da Silva, F. L. (2005). *Electroencephalography: basic principles, clinical applications, and related fields*. Lippincott Williams & Wilkins. (Cited on pages 125, 126, and 133)

[Nieuwenhuis et al., 2001] Nieuwenhuis, S., Ridderinkhof, K. R., Blom, J., Band, G. P., and Kok, A. (2001). Error-related brain potentials are differentially related to awareness of response errors: evidence from an antisaccade task. *Psychophysiology*, 38(5):752–60. (Cited on pages 15 and 16)

[Nijboer et al., 2011] Nijboer, F., Clausen, J., Allison, B. Z., and Haselager, P. (2011). Researchers' opinions about ethically sound dissemination of BCI research to the public media. *International Journal of Bioelectromagnetism*, 13(3):108–109. (Cited on page 141)

[Nijboer et al., 2009] Nijboer, F., Morin, F. O., Carmien, S. P., Koene, R. a., Leon, E., and Hoffmann, U. (2009). Affective brain-computer interfaces: Psychophysiological markers of emotion in healthy persons and in persons with amyotrophic lateral sclerosis. *ACII '09*, pages 1–11. (Cited on page 18)

[Nijboer et al., 2015] Nijboer, F., van de Laar, B., Gerritsen, S., Nijholt, A., and Poel, M. (2015). Usability of Three Electroencephalogram Headsets for Brain–Computer Interfaces: A Within Subject Comparison.

Interacting with Computers, 27(5):500–511. (Cited on pages 170, 183, and 209)

[Nijholt et al., 2009] Nijholt, A., Bos, D. P.-O., and Reuderink, B. (2009). Turning shortcomings into challenges: Brain–computer interfaces for games. *Entertainment Computing*, 1(2):85–94. (Cited on pages 22, 31, and 112)

[Nisbett and Wilson, 1977] Nisbett, R. E. and Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, 84(3):231–260. (Multiple combos on pages xvii, 4, 147, and 150)

[Noble, 2009] Noble, W. S. (2009). How does multiple testing correction work? *Nature biotechnology*, 27(12):1135–1137. (Cited on pages 81 and 118)

[Norman et al., 2009] Norman, J. F., Swindle, J. M., Jennings, L. R., Mullins, E. M., and Beers, A. M. (2009). Stereoscopic shape discrimination is well preserved across changes in object size. *Acta psychologica*, 131(2):129–35. (Cited on page 38)

[Norooz et al., 2015] Norooz, L., Mauriello, M. L., Jorgensen, A., Mcnally, B., and Froehlich, J. E. (2015). BodyVis: A New Approach to Body Learning Through Wearable Sensing and Visualization. In *CHI '15*. (Cited on page 146)

[Norton et al., 2015] Norton, J. J. S., Lee, D. S., Lee, J. W., Lee, W., Kwon, O., Won, P., Jung, S.-Y., Cheng, H., Jeong, J.-W., Akce, A., Umunna, S., Na, I., Kwon, Y. H., Wang, X.-Q., Liu, Z., Paik, U., Huang, Y., Bretl, T., Yeo, W.-H., and Rogers, J. a. (2015). Soft, curved electrode systems capable of integration on the auricle as a persistent brain–computer interface. *Proceedings of the National Academy of Sciences*, page 201424875. (Cited on page 202)

[Norton et al., 2012] Norton, M. I., Mochon, D., and Ariely, D. (2012). The IKEA effect: When labor leads to love. *Journal of Consumer Psychology*, 22(3):453–460. (Cited on page 149)

[Nourbakhsh et al., 2013] Nourbakhsh, N., Wang, Y., and Chen, F. (2013). GSR and Blink Features for Cognitive Load Classification. In *INTERACT '13*. (Cited on page 54)

[Nygårds and Sörnmo, 1983] Nygårds, M. E. and Sörnmo, L. (1983). Delineation of the QRS complex using the envelope of the e.c.g. *Medical & Biological Engineering & Computing*, 21(5):538–547. (Cited on pages 62 and 70)

[O'Brien et al., 1986] O'Brien, I. a., O'Hare, P., and Corrall, R. J. (1986). Heart rate variability in healthy subjects: effect of age and the derivation of normal ranges for tests of autonomic function. *British heart journal*, 55(4):348–354. (Cited on page 217)

[Ogolla, 2011] Ogolla, J. A. (2011). *Usability Evaluation: Tasks Susceptible to Concurrent Think-Aloud Protocol*. Master thesis, Linköping University. (Cited on page 4)

[Oken et al., 2006] Oken, B. S., Salinsky, M. C., and Elsas, S. M. (2006). Vigilance, alertness, or sustained attention: physiological basis and measurement. *Clin Neurophysiol*, 117(9):1885–901. (Cited on pages 12, 13, and 154)

[Owen et al., 2005] Owen, A. M., McMillan, K. M., Laird, A. R., and Bullmore, E. (2005). N-back working memory paradigm: A meta-analysis of normative functional neuroimaging studies. *Human Brain Mapping*, 25(1):46–59. (Cited on pages 58 and 68)

[Paiva et al., 2005] Paiva, A., Dias, J., Sobral, D., Aylett, R., Woods, S., Hall, L., and Zoll, C. (2005). Learning by Feeling: Evoking Empathy with Synthetic Characters. *Applied Artificial Intelligence*, 19(3-4):235–266. (Cited on page 140)

[Parasuraman, 2013] Parasuraman, R. (2013). Neuroergonomics: Brain-Inspired Cognitive engineering. In *The Oxford Handbook Of Cognitive Engineering*, page 672. Oxford University Press, USA. (Cited on pages 2, 10, and 14)

[Parra et al., 2003] Parra, L. C., Spence, C. D., Gerson, A. D., and Sajda, P. (2003). Response error correction–a demonstration of improved human-machine performance using real-time EEG monitoring. *IEEE Eng. Med. Biol. Soc.*, 11(2):173–7. (Cited on page 17)

[Partala and Surakka, 2003] Partala, T. and Surakka, V. (2003). Pupil size variation as an indication of affective processing. *International Journal of Human-Computer Studies*, 59(1-2):185–198. (Cited on page 4)

[Pascual-Marqui, 2002] Pascual-Marqui, R. (2002). Standardized low-resolution brain electromagnetic tomography (sLORETA): technical details. *Methods Find Exp Clin Pharmacol*, pages 5–12. (Cited on page 132)

[Peck et al., 2013] Peck, E. M. M., Yuksel, B. F., Ottley, A., Jacob, R. J., and Chang, R. (2013). Using fNIRS brain sensing to evaluate information visualization interfaces. *CHI '13*, page 473. (Cited on pages 12 and 54)

[Pfurtscheller and Lopes da Silva, 1999] Pfurtscheller, G. and Lopes da Silva, F. H. (1999). Event-related EEG/MEG synchronization and desynchronization: basic principles. *Clinical neurophysiology: official journal of the International Federation of Clinical Neurophysiology*, 110(11):1842–57. (Cited on page 133)

[Picard et al., 2001] Picard, R., Vyzas, E., and Healey, J. (2001). Toward machine emotional intelligence: analysis of affective physiological state. *IEEE Trans. Pattern Anal. Mach. Intell.*, 23(10):1175–1191. (Cited on page 19)

[Picard, 1995] Picard, R. W. (1995). Affective computing. Technical Report 321, MIT Media Laboratory. (ULTRA combos on pages xvii, xviii, 4, 17, 98, 100, 112, and 145)

[Picard, 2009] Picard, R. W. (2009). Future affective technology for autism and emotion communication. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 364(1535):3575–3584. (Cited on page 163)

[Pike et al., 2012] Pike, M., Wilson, M., Divoli, A., and Medelyan, A. (2012). CUES: Cognitive Usability Evaluation System. *EuroHCIR '12*, pages 1–4. (Cited on pages xviii, 20, and 21)

[Piper et al., 2002] Piper, B., Ratti, C., and Ishii, H. (2002). Illuminating Clay : A 3-D Tangible Interface for Landscape Analysis. In *CHI '02*, pages 355–362. (Cited on page 128)

[Placido da Silva et al., 2014] Placido da Silva, H., Fred, A., and Martins, R. (2014). Biosignals for Everyone. *IEEE Pervasive Computing*, 13(4):64–71. (Cited on page 57)

[Poh et al., 2010] Poh, M.-Z., Swenson, N., and Picard, R. W. (2010). A wearable sensor for unobtrusive, long-term assesment of electrodermal activity. *IEEE Transactions on Biomedical Engineering*, 57(5):1243–1252. (Cited on page 153)

[Poor et al., 2013] Poor, G. M., Zimmerman, G., Klopfer, D. S., Jaffee, S. D., Leventhal, L. M., and Barnes, J. (2013). Mobility Matters: Identifying Cognitive Demands That Are Sensitive to Orientation. In *INTERACT '13*, pages 193–210. (Cited on page 77)

[Posner et al., 2005] Posner, J., Russell, J. a., and Peterson, B. S. (2005). The circumplex model of affect: an integrative approach to affective neuroscience, cognitive development, and psychopathology. *Dev. Psychopathol.*, 17(3):715–34. (Cited on pages 17 and 18)

[Prasad, 2013] Prasad, E. (2013). *Wearable Electrodermal Activity Sensing-a Validation Study*. PhD thesis, University of Houston. (Cited on page 153)

[Prendinger et al., 2004] Prendinger, H., Dohi, H., Wang, H., Mayer, S., and Ishizuka, M. (2004). Empathic Embodied Interfaces: Addressing Users' Affective State. In *Affective Dialogue Systems*, pages 53–64. Springer Berlin Heidelberg. (Cited on pages 98 and 99)

[Raca and Dillenbourg, 2013] Raca, M. and Dillenbourg, P. (2013). System for Assessing Classroom Attention. In *LAK '13*, pages 1–5. (Cited on page 163)

[Ramoser et al., 2000] Ramoser, H., Müller-Gerking, J., and Pfurtscheller, G. (2000). Optimal spatial filtering of single trial EEG during imagined hand movement. *IEEE Trans Rehabil Eng*, 8(4):441–446. (Cited on page 61)

[Raskar et al., 2001] Raskar, R., Welch, G., Low, K.-L., and Bandyopadhyay, D. (2001). Shader lamps: Animating real objects with image-based illumination. In *Proceedings of the 12th Eurographics Workshop on Rendering Techniques*. (Cited on pages 113, 128, and 155)

[Ravaja, 2009] Ravaja, N. (2009). FUGA: The Fun of Gaming: Measuring the Human Experience of Media Enjoyment. Final Activity Report. Technical report, Helsinki Institute for Information Technology. (Multiple combos on pages xviii, 4, 22, 23, and 24)

[Ravaja et al., 2006] Ravaja, N., Saari, T., Salminen, M., Laarni, J., and Kallinen, K. (2006). Phasic Emotional Reactions to Video Game Events: A Psychophysiological Investigation. *Media Psychology*, 8(4):343–367. (Cited on page 19)

[Reidsma et al., 2010] Reidsma, D., Nijholt, A., Tschacher, W., and Ramseyer, F. (2010). Measuring Multimodal Synchrony for Human-Computer Interaction. In *2010 International Conference on Cyberworlds*, pages 67–71. IEEE. (Cited on page 98)

[Reis et al., 2014] Reis, P. M. R., Hebenstreit, F., Gabsteiger, F., von Tscharner, V., and Lochmann, M. (2014). Methodological aspects of EEG and body dynamics measurements during motion. *Frontiers in human neuroscience*, 8(March):156. (Cited on page 209)

[Renard et al., 2010] Renard, Y., Lotte, F., Gibert, G., Congedo, M., Maby, E., Delannoy, V., Bertrand, O., and Lécuyer, A. (2010). OpenViBE: An Open-Source Software Platform to Design, Test, and Use Brain-Computer Interfaces in Real and Virtual Environments. *Presence: Teleoperators and Virtual Environments*, 19(1):35–53. (Cited on pages 104, 126, and 134)

[Rennert and Karapanos, 2013] Rennert, K. and Karapanos, E. (2013). Faceit: supporting reflection upon social anxiety events with lifelogging. *CHI'13 Extended Abstracts on Human Factors ...*, pages 457–462. (Cited on page 175)

[Rettie, 2003] Rettie, R. (2003). Connectedness, awareness and social presence. *Proceedings of PRESENCE 2003*, 8:474–84. (Cited on page xx)

[Rich, 1983] Rich, A. (1983). Shielding and guarding, how to exclude interference-type noise, what to do and why to do: a rationale approach. *Analog Devices*, 17(1). (Cited on page 208)

[Rose, 2014] Rose, D. (2014). *Enchanted Objects: Design, Human Desire, and the Internet of Things*. Simon and Schuster. (Cited on page 159)

[Roussel, 2002] Roussel, N. (2002). *Support informatique à une communication médiatisée*. PhD thesis, Paris-Sud. (Cited on page xvii)

[Roussel, 2007] Roussel, N. (2007). *Nouvelles formes de communication et nouvelles interactions homme-machine pour enrichir et simplifier le quotidien*. PhD thesis, Université Paris-Sud. (Cited on pages xix and xx)

[Roussel, 2014] Roussel, N. (2014). Looking back: a very brief history of HCI. (Cited on page xvii)

[Saavedra and Bougrain, 2012] Saavedra, C. and Bougrain, L. (2012). Processing Stages of Visual Stimuli and Event-Related Potentials. *The NeuroComp/KEOpS'12 workshop*, 2:1–5. (Cited on page 14)

[Sajda et al., 2010] Sajda, P., Pohlmeyer, E., Parra, L., Christoforou, C., Dmochowski, J., Hanna, B., Bahlmann, C., and Singh, M. (2010). In a Blink of an Eye and a Switch of a Transistor: Cortically Coupled Computer Vision. *Proceedings of the IEEE*, 98(3):462–478. (Cited on page 30)

[Sallnäs et al., 2000] Sallnäs, E.-L., Rassmus-Gröhn, K., and Sjöström, C. (2000). Supporting Presence in Collaborative Environments by Haptic Force Feedback. *ACM Transactions on Computer-Human Interaction*, 7(4):461–476. (Cited on pages xix and 22)

[Salzmann et al., 2009] Salzmann, H., Jacobs, J., and Froehlich, B. (2009). Collaborative interaction in co-located two-user scenarios. In *Joint Virtual Reality Conference of EGVE - ICAT - EuroVR*. (Cited on page 35)

[Samek et al., 2013] Samek, W., Meinecke, F. C., and Muller, K.-R. (2013). Transferring Subspaces Between Subjects in Brain–Computer Interfacing. *IEEE Transactions on Biomedical Engineering*, 60(8):2289–2298. (Multiple combos on pages 61, 62, 70, and 71)

[Sauro and Lewis, 2009] Sauro, J. and Lewis, J. R. (2009). Correlations among prototypical usability metrics. In *CHI '09*, page 1609, New York, New York, USA. ACM Press. (Cited on page 20)

[Schalk et al., 2000] Schalk, G., Wolpaw, J. R., McFarland, D. J., and Pfurtscheller, G. (2000). EEG-based communication: presence of an error potential. *Clin. Neurophysiol.*, 111(12):2138–44. (Cited on page 16)

[Scheirer et al., 2002] Scheirer, J., Fernandez, R., Klein, J., and Picard, R. W. (2002). Frustrating the user on purpose: a step toward building an affective computer. *Interacting with computers*, 14(2):93–118. (Cited on page 22)

[Schlögl et al., 2007] Schlögl, A., Keinrath, C., Zimmermann, D., Scherer, R., Leeb, R., and Pfurtscheller, G. (2007). A fully automated correction method of EOG artifacts in EEG recordings. *Clinical neurophysiology : official journal of the International Federation of Clinical Neurophysiology*, 118(1):98–104. (Cited on page 50)

[Schmidt et al., 2012] Schmidt, N. M., Blankertz, B., and Treder, M. S. (2012). Online detection of error-related potentials boosts the performance of mental typewriters. *BMC neuroscience*, 13(1):19. (Cited on pages 17 and 31)

[Schmitz, 2010] Schmitz, M. (2010). *Tangible Interaction with Anthropomorphic Smart Objects in Instrumented Environments*. PhD thesis, Saarland University. (Cited on page 148)

[Schober et al., 1995] Schober, F., Schellenberg, R., and Dimpfel, W. (1995). Reflection of mental exercise in the dynamic quantitative topographical EEG. *Neuropsychobiology*, 31(2):98–112. (Cited on page 154)

[Scholler et al., 2012] Scholler, S., Bosse, S., Treder, M. S., Blankertz, B., Curio, G., Müller, K.-R., and Wiegand, T. (2012). Toward a direct measure of video quality perception using EEG. *IEEE Trans. Image Process.*, 21(5):2619–29. (Cited on page 14)

[Schwartz et al., 1993] Schwartz, J. E., Jandorf, L., and Krupp, L. B. (1993). The measurement of fatigue: A new instrument. *J Psychosom Res*, 37(7):753–762. (Cited on page 13)

[Shaer, 2009] Shaer, O. (2009). Tangible User Interfaces: Past, Present, and Future Directions. *Foundations and Trends® in Human-Computer Interaction*, 3(1-2):1–137. (Cited on page 128)

[Shaw, 2003] Shaw, J. C. (2003). *The brain's alpha rhythms and the mind*. Elsevier. (Cited on page 14)

[Sheridan, 1992] Sheridan, T. (1992). Musings on telepresence and virtual presence. *Presence Teleoperators and Virtual Environments*, 1:120–126. (Cited on page xx)

[Shibata et al., 2011] Shibata, T., Kim, J., Hoffman, D. M., and Banks, M. S. (2011). The zone of comfort: Predicting visual discomfort with stereo displays. *Journal of Vision*, 11:1–29. (Multiple combos on pages 36, 37, 38, and 41)

[Short et al., 1976] Short, J., Christie, B., and Williams, E. (1976). *The Social Psychology of Telecommunications*. Wiley, John & Sons, Incorporated. (Cited on page xix)

[Short, 1974] Short, J. a. (1974). Effects of Medium of Communication on Experimental Negotiation. *Human Relations*, 27(3):225–234. (Cited on page xix)

[Siegel et al., 1986] Siegel, J., Dubrovsky, V., Kiesler, S., and McGuire, T. W. (1986). Group Processes in Computer-mediated Communication. *Organizational Behavior and Human Decision Processes*, 37(2):157–187. (Cited on page xix)

[Simons and Chabris, 2011] Simons, D. J. and Chabris, C. F. (2011). What people believe about how memory works: A representative survey of the U.S. population. *PLoS ONE*, 6(8). (Cited on page 127)

[Sjölie et al., 2010] Sjölie, D., Bodin, K., Elgh, E., Eriksson, J., Janlert, L.-E., and Nyberg, L. (2010). Effects of interactivity and 3D-motion on mental rotation brain activity in an immersive virtual environment. In *CHI '10*, page 869, New York, New York, USA. ACM Press. (Cited on page xviii)

[Slater et al., 2009] Slater, M., Lotto, B., Arnold, M. M., and Sanchez-Vives, M. V. (2009). How we experience immersive virtual environments: the concept of presence and its measurement. *Anuario de psicología*, 40(2773):193–210. (Cited on pages xix, 21, and 22)

[Slovák et al., 2012] Slovák, P., Janssen, J. H., and Fitzpatrick, G. (2012). Understanding heart rate sharing: towards unpacking physiosocial space. *CHI '12*, pages 859–868. (Multiple combos on pages xx, 99, 109, 112, and 113)

[Sobolewski et al., 2013] Sobolewski, A., Chavarriaga, R., and Millán, J. (2013). Error Processing of Self-paced Movements. In *TOBI Workshop IV*, pages 137–138. (Cited on page 17)

[Soleymani et al., 2009] Soleymani, M., Chanel, G., Kierkels, J. J. M., and Pun, T. (2009). Affective Characterization of Movie Scenes Based on Content Analysis and Physiological Changes. *International Journal of Semantic Computing*, 03(02):235–254. (Cited on pages 18 and 19)

[Stanney et al., 2003] Stanney, K. M., Mollaghasemi, M., Reeves, L., Breaux, R., and Graeber, D. a. (2003). Usability engineering of virtual environments (VEs): identifying multiple criteria that drive effective VE system design. *Int. J. Hum-Comput St.*, 58(4):447–481. (Cited on page 20)

[Sternberg, 1966] Sternberg, S. (1966). High-speed scanning in human memory. *Science (New York, N.Y.)*, 153(736):652–654. (Cited on page 76)

[Stopczynski et al., 2014] Stopczynski, A., Stahlhut, C., Larsen, J. E., Petersen, M. K., and Hansen, L. K. (2014). The smartphone brain scanner: A portable real-time neuroimaging system. *PLoS ONE*, 9(2). (Cited on page 127)

[Sumioka et al., 2014] Sumioka, H., Nishio, S., Minato, T., Yamazaki, R., and Ishiguro, H. (2014). Minimal Human Design Approach for sonzai-kan Media: Investigation of a Feeling of Human Presence. *Cognitive Computation*, pages 1–3. (Cited on page 150)

[Sun et al., 2014] Sun, D., Paredes, P., and Canny, J. (2014). MouStress. In *CHI '14*, pages 61–70, New York, New York, USA. ACM Press. (Cited on page 18)

[Tadel et al., 2011] Tadel, F., Baillet, S., Mosher, J. C., Pantazis, D., and Leahy, R. M. (2011). Brainstorm: a user-friendly application for MEG/EEG analysis. *Computational intelligence and neuroscience*, 2011:879716. (Cited on page 132)

[Tallgren et al., 2005] Tallgren, P., Vanhatalo, S., Kaila, K., and Voipio, J. (2005). Evaluation of commercially available electrodes and gels for recording of slow EEG potentials. *Clinical neurophysiology : official journal of the International Federation of Clinical Neurophysiology*, 116(4):799–806. (Cited on page 202)

[Tan and Nijholt, 2010] Tan, D. S. and Nijholt, A. (2010). *Brain-Computer Interfaces: Applying Our Minds to Human-Computer Interaction*. Springer. (Cited on page 29)

[Tanaka et al., 2001] Tanaka, H., Monahan, K. D., and Seals, D. R. (2001). Age-predicted maximal heart rate revisited. *Journal of the American College of Cardiology*, 37(1):153–156. (Cited on page 217)

[Tautan et al., 2013] Tautan, A.-M., Serdijn, W., Mihajlovic, V., Grundlehner, B., and Penders, J. (2013). Framework for evaluating EEG signal quality of dry electrode recordings. *2013 IEEE Biomedical Circuits and Systems Conference (BioCAS)*, pages 186–189. (Cited on pages 183 and 204)

[Tholander and Nylander, 2015] Tholander, J. and Nylander, S. (2015). Snot, Sweat, Pain, Mud, and Snow: Performance and Experience in the Use of Sports Watches. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15*, pages 2913–2922. (Cited on page 112)

[Trachel et al., 2013] Trachel, R., Brochier, T., and Clerc, M. (2013). Enhancing visuospatial attention performance with brain-computer interfaces. *CHI '13*, page 1245. (Cited on page 14)

[Tuček et al., 2012] Tuček, D., Mount, W., and Abbass, H. (2012). Neural and Speech Indicators of Cognitive Load for Sudoku Game Interfaces. *Neural Information Processing*, pages 210–217. (Cited on page 11)

[Underkoffler and Ishii, 1999] Underkoffler, J. and Ishii, H. (1999). Urp: A luminous-tangible workbench for urban planning and design. *CHI '99*, pages 386–393. (Cited on page 128)

[Usakli, 2010] Usakli, A. B. (2010). Improvement of EEG signal acquisition: an electrical aspect for state of the art of front end. *Computational intelligence and neuroscience*, 2010:630649. (Cited on pages 179 and 208)

[Uğur, 2013] Uğur, S. (2013). *Wearing Embodied Emotions: A Practice Based Design Research on Wearable Technology*. Springer Milan. (Cited on page 146)

[van de Laar et al., 2013] van de Laar, B., Gürkök, H., Bos, D. P.-O., Nijboer, F., and Nijholt, A. (2013). Brain–Computer Interfaces and User Experience Evaluation. In Allison, B. Z., Dunne, S., Leeb, R., Del R. Millán, J., and Nijholt, A., editors, *Towards Practical Brain-Computer Interfaces*, Biological and Medical Physics, Biomedical Engineering, pages 223–237. Springer Berlin Heidelberg, Berlin, Heidelberg. (Cited on pages 7 and 24)

[van de Velde et al., 1998] van de Velde, M., van Erp, G., and Cluitmans, P. J. (1998). Detection of muscle artefact in the normal human awake EEG. *Electroencephalography and Clinical Neurophysiology*, 107(2):149–158. (Cited on page 159)

[van der Linden et al., 2003] van der Linden, D., Frese, M., and Meijman, T. F. (2003). Mental fatigue and the control of cognitive processes: effects on perseveration and planning. *Acta Psychologica*, 113(1):45–65. (Cited on page 13)

[van Drunen et al., 2009] van Drunen, A., van den Broek, E. L., Spink, A. J., and Heffelaar, T. (2009). Exploring workload and attention measurements with uLog mouse data. *Behavior research methods*, 41(3):868–75. (Cited on page 11)

[van Erp et al., 2012] van Erp, J., Lotte, F., and Tangermann, M. (2012). Brain-Computer Interfaces: Beyond Medical Applications. *Computer*, 45(4):26–34. (Cited on page 61)

[van Erp et al., 2010] van Erp, J. B. F., Veltman, H., and Grootjen, M. (2010). Brain-Based Indices for User System Symbiosis. In Tan, D. S. and Nijholt, A., editors, *Brain-Computer Interfaces*, Human-Computer Interaction Series, pages 201–219. Springer London, London. (Cited on pages 12, 13, and 24)

[Vecchiato et al., 2011] Vecchiato, G., Astolfi, L., De Vico Fallani, F., Toppi, J., Aloise, F., Bez, F., Wei, D., Kong, W., Dai, J., Cincotti, F., Mattia, D., and Babiloni, F. (2011). On the use of EEG or MEG brain imaging tools in neuromarketing research. *Comput Intell Neurosci*, 2011:643489. (Cited on page 19)

[Vi and Subramanian, 2012] Vi, C. and Subramanian, S. (2012). Detecting error-related negativity for interaction design. *CHI '12*, page 493. (Cited on pages 17 and 49)

[Vidaurre et al., 2011] Vidaurre, C., Sander, T. H., and Schlögl, A. (2011). BioSig: the free and open source software library for biomedical signal processing. *Computational intelligence and neuroscience*, 2011. (Cited on page 62)

[Villon and Lisetti, 2006] Villon, O. and Lisetti, C. (2006). A user-modeling approach to build user's psycho-physiological maps of emotions using bio-sensors. In *RO-MAN '06*, pages 269–276. (Cited on page xviii)

[Viola and Jones, 2001] Viola, P. and Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. In *CVPR '01*, volume 1. (Cited on page 213)

[Vosniadou, 1992] Vosniadou, S. (1992). Knowledge acquisition and conceptual change. *Applied Psychology: An International Review*, 41(4):347–357. (Cited on pages 127 and 139)

[Vosniadou et al., 2001] Vosniadou, S., Ioannides, C., Dimitrakopoulou, A., and Papademetriou, E. (2001). Designing learning environments to promote conceptual change in science. *Learning and Instruction*, 11(4-5):381–419. (Cited on pages 126 and 128)

[Vrzakova, 2013] Vrzakova, H. (2013). Inferential methods in interaction, usability and user experience. *CHI'13*, pages 1965–1968. (Cited on page 21)

[Wadhwa et al., 2014] Wadhwa, N., Rubinstein, M., Durand, F., and Freeman, W. T. (2014). Riesz pyramids for fast phase-based video magnification. *2014 IEEE International Conference on Computational Photography, ICCP 2014*. (Cited on page 212)

[Wagner and Schmalstieg, 2007] Wagner, D. and Schmalstieg, D. (2007). ARToolKitPlus for Pose Tracking on Mobile Devices. *Proceedings of 12th Computer Vision Winter Workshop CVWW07*, pages 139–146. (Cited on page 134)

[Walmink et al., 2014] Walmink, W., Wilde, D., and Mueller, F. F. (2014). Displaying Heart Rate Data on a Bicycle Helmet to Support Social Exertion Experiences. In *TEI '14*. (Cited on pages 109, 112, and 163)

[Wang et al., 2014] Wang, C., Geelhoed, E. N., Stenton, P. P., and Cesar, P. (2014). Sensing a live audience. In *CHI '14*, pages 1909–1912. (Cited on page 163)

[Ward, 2006] Ward, J. (2006). *The Student's Guide to Cognitive Neuroscience*. Hove : Psychology Press. (Cited on page 5)

[Weber, 2007] Weber, J. (2007). Think Aloud Best Practices Study. (Cited on page 3)

[Weiser, 1993] Weiser, M. (1993). Some computer science issues in ubiquitous computing. (Cited on page 156)

[Wellington, 1990] Wellington, J. (1990). Formal and informal learning in science: the role of the interactive science centres. *Physics Education*, 25(5):247–252. (Cited on page 128)

[Williams et al., 2015] Williams, M. A., Roseway, A., O'Dowd, C., Czerwinski, M., and Morris, M. R. (2015). SWARM: An Actuated Wearable for Mediating Affect. In *TEI '15*, pages 293–300, New York, New York, USA. ACM Press. (Cited on pages 112 and 146)

[Wilson et al., 2012a] Wilson, A., Benko, H., Izadi, S., and Hilliges, O. (2012a). Steerable augmented reality with the beamatron. In *UIST '12*, page 413, New York, New York, USA. ACM Press. (Cited on page 128)

[Wilson et al., 2012b] Wilson, J., Guger, C., and Schalk, G. (2012b). BCI hardware and software. In *Brain-computer interfaces: principles and practice*, pages 165–188. Oxford University Press. (Cited on page 94)

[Winton et al., 1984] Winton, W. M., Putnam, L. E., and Krauss, R. M. (1984). Facial and autonomic manifestations of the dimensional structure of emotion. *Journal of Experimental Social Psychology*, 20(3):195–216. (Cited on page 100)

[Witmer and Singer, 1998] Witmer, B. G. and Singer, M. J. (1998). Measuring Presence in Virtual Environments: A Presence Questionnaire. *Presence: Teleoper. Virtual Environ.*, 7(3):225–240. (Cited on page xix)

[Wobrock et al., 2015] Wobrock, D., Frey, J., Graef, D., de la Rivière, J.-B., Castet, J., and Lotte, F. (2015). Continuous Mental Effort Evaluation during 3D Object Manipulation Tasks based on Brain and Physiological Signals. In *INTERACT '15*. (Cited on pages 28 and 52)

[Wolpaw et al., 2002] Wolpaw, J. R., Birbaumer, N., McFarland, D. J., Pfurtscheller, G., and Vaughan, T. M. (2002). Brain-computer interfaces for communication and control. *Clin. Neurophysiol.*, 113(6):767–91. (Cited on pages xviii and 29)

[Wolpaw and Wolpaw, 2012] Wolpaw, J. R. and Wolpaw, E. W. (2012). *BCI: Principles and Practice*. Oxford University Press. (Cited on page 142)

[Wright and McCarthy, 2008] Wright, P. and McCarthy, J. (2008). Empathy and experience in HCI. *CHI '08*. (Cited on page 105)

[Wu et al., 2012] Wu, H.-Y., Rubinstein, M., Shih, E., Guttag, J., Durand, F., and Freeman, W. (2012). Eulerian video magnification for revealing subtle changes in the world. *ACM Transactions on Graphics*, 31(4):1–8. (Cited on pages 211 and 212)

[Yamabe et al., 2010] Yamabe, T., Kosunen, I., Ekman, I., Liikkanen, L. A., Kuikkaniemi, K., and Nakajima, T. (2010). Biofeedback Training with EmoPoker: Controlling Emotional Arousal for Better Poker Play. In *BioS-Play '10*. (Cited on page 113)

[Ye et al., 2014] Ye, X., He, D., and Zheng, P. H. (2014). ESE 498 Hypnolarm. Technical Report May, Washington University. (Cited on page 204)

[Zander, 2011] Zander, T. (2011). *Utilizing Brain-Computer Interfaces for Human-Machine Systems*. PhD thesis, Technischen Universität Berlin. (Cited on page 32)

[Zander and Kothe, 2011] Zander, T. O. and Kothe, C. (2011). Towards passive brain-computer interfaces: applying brain-computer interface technology to human-machine systems in general. *J. Neural. Eng*, 8(2):025005. (Multiple combos on pages xviii, 30, 31, and 145)

[Zander et al., 2011] Zander, T. O., Lehne, M., Ihme, K., Jatzev, S., Correia, J., Kothe, C., Picht, B., and Nijboer, F. (2011). A Dry EEG-System for Scientific Research and Brain–Computer Interfaces. *Frontiers in Neuroscience*, 5(May):1–10. (Cited on pages 187 and 208)

[Zao et al., 2014] Zao, J. K., Gan, T.-T., You, C.-K., Chung, C.-E., Wang, Y.-T., Rodríguez Méndez, S. J., Mullen, T., Yu, C., Kothe, C., Hsiao, C.-T., Chu, S.-L., Shieh, C.-K., and Jung, T.-P. (2014). Pervasive brain monitoring and data sharing based on multi-tier distributed computing and linked data technology. *Frontiers in human neuroscience*, 8(June):370. (Cited on page 173)

[Zhai and Milgram, 1998] Zhai, S. and Milgram, P. (1998). Quantifying coordination in multiple DOF movement and its application to evaluating 6 DOF input devices. In *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '98*, pages 320–327, New York, New York, USA. ACM Press. (Cited on page 92)