

EEG-based neuroergonomics for 3D user interfaces: opportunities and challenges

Jérémy Frey, Martin Hachet, Fabien Lotte

► **To cite this version:**

Jérémy Frey, Martin Hachet, Fabien Lotte. EEG-based neuroergonomics for 3D user interfaces: opportunities and challenges. Le travail humain, Presses Universitaires de France, 2016. <hal-01394254>

HAL Id: hal-01394254

<https://hal.inria.fr/hal-01394254>

Submitted on 9 Nov 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

EEG-BASED NEUROERGONOMICS FOR 3D USER INTERFACES: OPPORTUNITIES AND CHALLENGES

JÉRÉMY FREY^{1,2}, MARTIN HACHET², FABIEN LOTTE²

1-UNIVERSITY OF BORDEAUX, 2-INRIA BORDEAUX SUD-OUEST

3D user interfaces (3DUI) are increasingly used in a number of applications, spanning from entertainment to industrial design. However, 3D interaction tasks are generally more complex for users since interacting with a 3D environment is more cognitively demanding than perceiving and interacting with a 2D one. As such, it is essential that we could evaluate finely user experience, in order to propose seamless interfaces. To do so, a promising research direction is to measure users' inner-state based on brain signals acquired during interaction, by following a neuroergonomics approach. Combined with existing methods, such tool can be used to strengthen the understanding of user experience. In this paper, we review the work being undergone in this area; what has been achieved and the new challenges that arise. We describe how a mobile brain imaging technique such as electroencephalography (EEG) brings continuous and non-disruptive measures. EEG-based evaluation of users can give insights about multiple dimensions of the user experience, with realistic interaction tasks or novel interfaces. We investigate four constructs: workload, attention, error recognition and visual comfort. Ultimately, these metrics could help to alleviate users when they interact with computers.

INTRODUCTION

Interactive systems that rely on 3D User Interfaces (UI) are increasingly used in a number of applications including industrial design, education, art or entertainment (Bowman, Kruijff, LaViola Jr., & Poupyrev, 2005; Jankowski & 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, Gabbard, & Hix, 2002; Jankowski & 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., 2005; Frey, Mühl, Lotte, & Hachet, 2014; Jankowski & Hachet, 2013). This resulted in the design of more relevant, efficient and easy-to-use 3DUI.

Nevertheless, there is still a lot of room for improvements in the currently used evaluation methods. In particular, traditional evaluation methods could either be

ambiguous, lack real-time recordings, or disrupt the interaction (Frey et al., 2014). For instance, although behavioral studies are able to account in real-time for users' interactions, they can be hard to interpret since measures may not be specific, e.g. a high reaction time can be caused either by a low concentration level or a high mental workload (Hart & Staveland, 1988). Questionnaires and other inquiry-based methods such as "think aloud" and focus group all suffer from the same limitation: resulting measures are prone to be contaminated by ambiguities (Nisbett & Wilson, 1977), social pressure (Picard, 1995) or participants' memory limitations (Kivikangas et al., 2010). There is therefore a need for more objective (or more precisely "exocentric", as defined in Frey et al. (2014)) and continuous measures of the usability qualities of 3DUI that do not interrupt the user during interaction.

In order to obtain such measures of the user's inner-state during interaction, a recent promising research direction is to measure such states based on brain signals acquired from the user during interaction (Frey et al., 2014). In other words, 3DUI could be evaluated by following a neuroergonomics approach (Parasuraman & Wilson, 2008). So far, neuroergonomics have been applied mostly to study and assess 2DUI, see, e.g. Parasuraman & Wilson (2008); Anderson et al. (2011); Peck, Yuksel, Ottley, Jacob, & Chang (2013). 3D interaction tasks are more complex for the user since interacting with a 3D environment is more cognitively demanding than perceiving and interacting with a 2D one. Indeed, it requires the user to perform 3D mental rotation tasks to successfully manipulate 3D objects or to orientate him/herself in the 3D environment. As such it involves cognitive mechanisms not solicited, or barely so, in classical 2D interaction tasks, such as spatial cognition. Moreover, 3D environments being usually much richer and complex to interact with than desktop 2DUI, they also generally lead to more pressure on the user's cognitive resources. Finally, perceiving 3D environments is also different from 2D ones, due to their 3D content, this content being potentially displayed in stereoscopy, leading to different perceptual mechanisms.

Therefore, in this paper, we show how 3DUI can benefit from neuroergonomics, and how new neuroergonomics tools and research are needed for this purpose. We notably focus on electroencephalography (EEG)-based neuroergonomics, since EEG are both accessible, portable, non-invasive and provide a high temporal resolution. This makes them suitable to measure a number of cognitive phenomena and constructs that are relevant to assess 3DUI, as we will see later in this manuscript. Overall this paper aims at identifying 1) the opportunities in EEG-based neuroergonomics for 3DUI, i.e. how useful metrics to assess 3DUI can be derived from EEG signals and 2) the remaining challenges, i.e. what still needs to be done to reach usable and useful neuroergonomics for 3DUI.

This paper is organized as follows. Section [Opportunities](#) identifies the opportunities in EEG-based neuroergonomics for 3DUI, i.e. what can we learn from EEG signals about a given 3DUI. In particular, we will see that analyzing EEG signals during 3D interaction tasks can provides us with information about the ergonomics of both the input devices and interaction techniques (Subsection [Input devices](#)), and the output, e.g. the display of the 3D environment (Subsection [Output devices](#)). We will

illustrate these opportunities with some of our recent works on this topic. Then, Section [Perspectives and Challenges](#) identifies the scientific challenges that we will have to solve to further advance the field, at the hardware (Subsection [Hardware](#)), software (Subsection [Software](#)), neuroscientific (Subsection [Neuroscience](#)) and human-computer interaction (HCI, Subsection [Human-Computer Interaction](#)) levels

OPPORTUNITIES

As mentioned earlier in this paper, 3DUI requires the user to perform generally complex 3D interaction tasks (3D objects manipulation, navigation (Bowman et al., 2005)) as well as to perceive a generally complex 3D environment. This is a typical interaction loop, with the user sending commands to the application, here 3D interaction tasks, and the system providing feedback to the user, here displaying a 3D environment. Interestingly enough, EEG signals can be used to study both aspects, i.e. how suitable is the input device/technique provided to the user, and how suitable is the output, typically the way the 3D content is displayed. More precisely, we and others have shown that we can estimate from EEG signals, more or less accurately, whether a 3D input device or technique is cognitively difficult and intuitive to use. It is also possible to infer from EEG signals whether a 3D display is comfortable for the user. We illustrate these two points with examples taken from our recent works in the following sections.

II.1 INPUT DEVICES

Recently we have been studying 3DUI through 2 separate studies. We started to investigate the use of physiological sensors to assess mental workload over the course of a 3D manipulation task in Wobrock et al. (2015). Then, in Frey, Daniel, Hachet, Castet, & Lotte (2016), we used specifically EEG in order to extend those results with more detailed measures and by encompassing more constructs, i.e. attention and error recognition. The realistic tasks that we have investigated demonstrate how EEG could be used to analyze complex 3DUI and give insights to designers while they develop new input modalities for HCI.

WORKLOAD DURING 3D MANIPULATION TASKS

We proposed in Wobrock et al. (2015) to use EEG and physiological signals – electrocardiography (ECG), electrodermal activity (EDA) – to study mental workload during a 3D docking task with a novel input device that possesses multiple degrees of freedom.

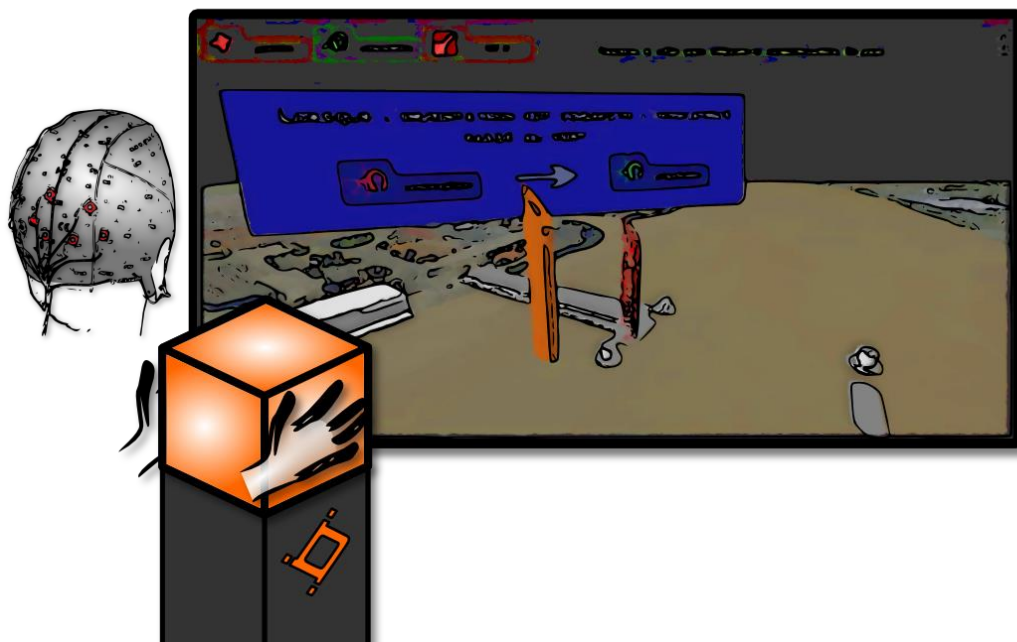


Figure 1. Schematic view of a user performing 3D manipulations tasks with the CubTile input device.

Participants manipulated a device known as the CubTile, developed by the Immersion company¹. The CubTile is made of 5 orthogonal touch surfaces (Figure 1). Thanks to those many planes, the CubTile enables users to manipulate 3D objects with more degrees of freedom than with traditional input devices; they can easily choose the axis onto which they perform scaling (with a “pinch” movement), rotation (when at least 2 fingers draw such motion) or translation (when fingers slide on one of the surfaces).

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. Participants had to complete a set of 7 docking tasks while their physiological signals were being monitored, by translating, rotating and scaling the bridge parts, in order to put them at the correct location. In particular, during the second task a time constraint was added and during the fifth task the controls were inverted on purpose, in order to test if we could measure a difference within physiological signals that would concur with such increases of difficulty.

With our approach, a calibration phase shall occur so that the system could learn which features are associated to a specific individual, during a task that is known to induce the studied construct. For workload estimation, we used a standard task known

1 <http://www.immersion.fr/>

as the N-back task (Kirchner, 1958), that here required the user to watch white letters on a black screen and to perform mouse-clicks. In the “0-back” (easy) condition participants simply have to discriminate between a target letter and distractors; in the “2-back” (difficult) condition, participants have to decide if the letter they see is the same as the one that appeared two times before – see Mühl, Jeunet, & Lotte (2014) for this implementation of the task. Hence, the N-back task plays on working memory, and it was used to induce two levels of workload – low and high. Once the calibration is completed, the machine could then use this knowledge to gain insights about a new context, here the 3D docking task (for comments about signal processing robust to context change between calibration and real task, see section [Software](#)).

Among the physiological signals that we investigated, EEG was the best able to discriminate workload levels reliably on the N-back task, with performances close to 90% on average for our 8 participants. This score is far beyond chance, which is 50% for a 2-class problem. In contrast, our analyses yielded poor performances with EDA (64.5% on average) and the classification accuracy of ECG signals was not significantly different from random. Hence, we focused our efforts on EEG while we pursued our evaluations.

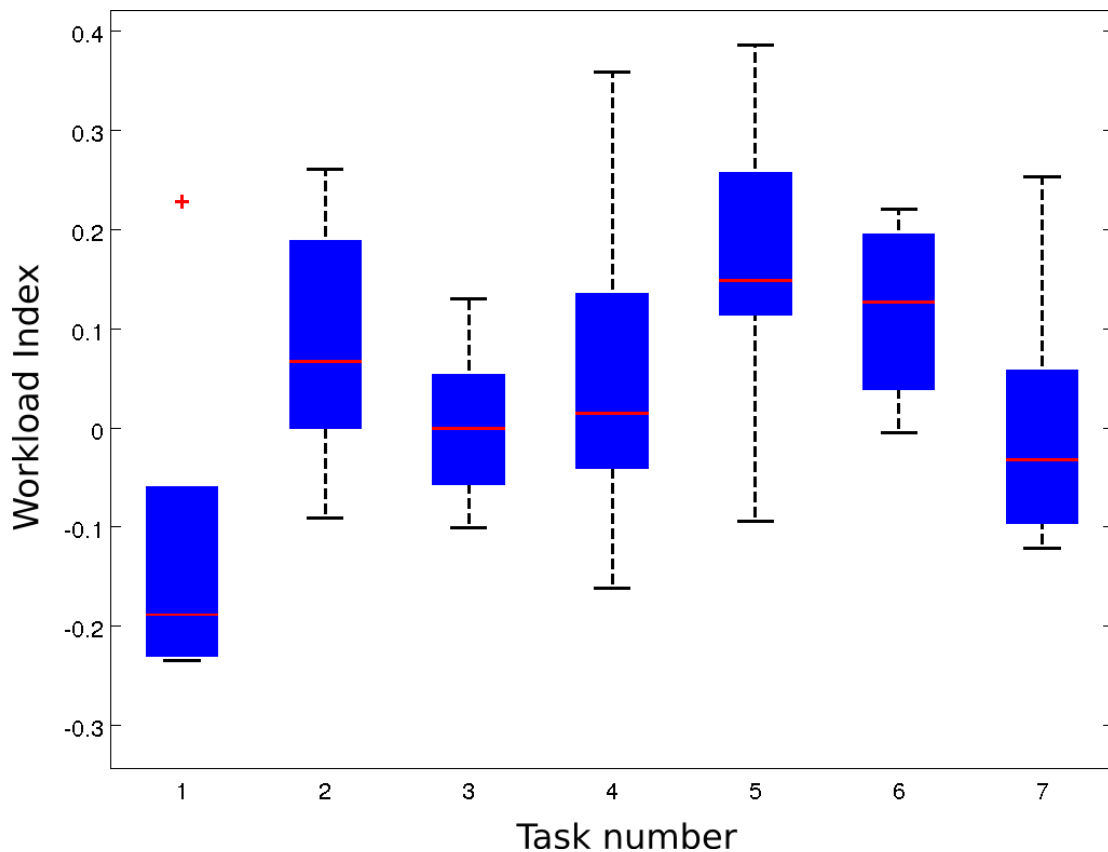


Figure 2. Workload index measured from EEG during various 3D docking tasks – the red cross pictures an outlier. A time constraint was added during task 2 and commands were inverted during task 5.

Once the baselines for the easy and difficult tasks were set, we processed the EEG activity recorded while participants were completing the 3D docking task. The results we obtained gave us unique insights into how much mental effort the participants were devoting to each task, and how this mental effort evolved over time (Figure 2).

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. The inverted commands during the fifth task 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, where the commands were back to normal – see Wobrock et al. (2015) for further discussions.

Overall, the mental workload that was measured with EEG along the course of the interaction matches the design of the tasks. Those results suggest that EEG may be a suitable tool for evaluating complex 3D interaction tasks, even when an interaction technique as novel as the CubTile is involved. We built upon this study to create an assessment of mental workload more precise and to integrate new dimensions of the user experience, such as the measure of attention.

FRAMEWORK FOR AN EEG-BASED EVALUATION OF USER EXPERIENCE

In Frey et al. (2016) we went 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, with various interaction techniques (Figure 3). We contribute a set of methods to estimate continuously the user’s mental workload, attention level and recognition of interaction errors 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, Tan, & Hudson (2008).



Figure 3. A keyboard and a touchscreen are compared using EEG measures during realistic interaction tasks.

The virtual environment takes the form of a maze where players have to learn and reproduce a path. A character displayed with a third person perspective moves by itself at a predefined speed inside orthogonal tunnels. Soon after the character enters a new tunnel, symbols appear on-screen, that users have to select timely in the right order. We crafted 4 levels of difficulty by tuning the speed or the number of items to remember. In order to make sure that we could elicit various mental states, we validated those levels in a pilot study using the NASA-TLX questionnaire (Hart & Staveland, 1988), a well established questionnaire that accounts for workload.

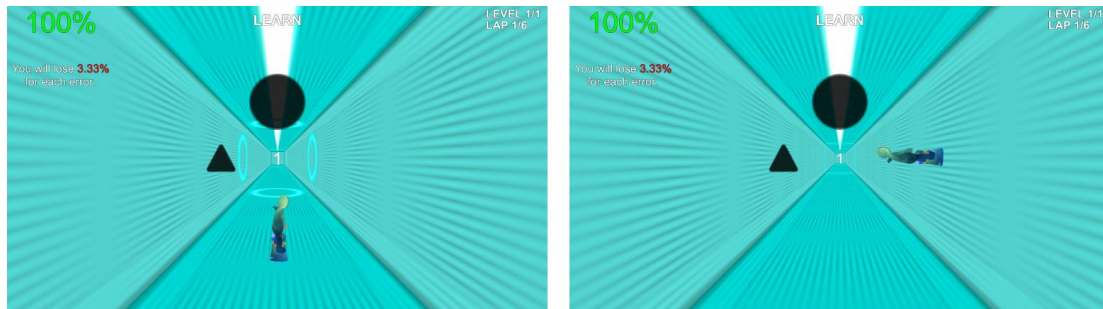


Figure 4. The virtual environment, where players control a character that moves by itself inside a 3D maze. Players have to select directions using the character as a frame of reference. Hence, when the character is on the right ceiling (right image), user have to press right in order to go up.

Beside learning a sequence, the principal challenge comes from *how* the directions are selected. The input device that users are controlling is mapped to the *character position* (Figure 4). Indeed, the character slides by itself from the bottom of the tunnel to one of the walls or to the ceiling from time to time. When the character is upside down for example, commands are inverted compared to what players are used to. This discrepancy between input and output is a reminder of the problematic often observed with 3DUI, where most users manipulate a device with 2 degrees of freedom (DOF), such as a mouse, to interact with a 6 DOF environment.

In particular, in this study we compared two interaction techniques: a keyboard and a touchscreen. 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 non-intuitive input commands, that were not co-localized with output directions.

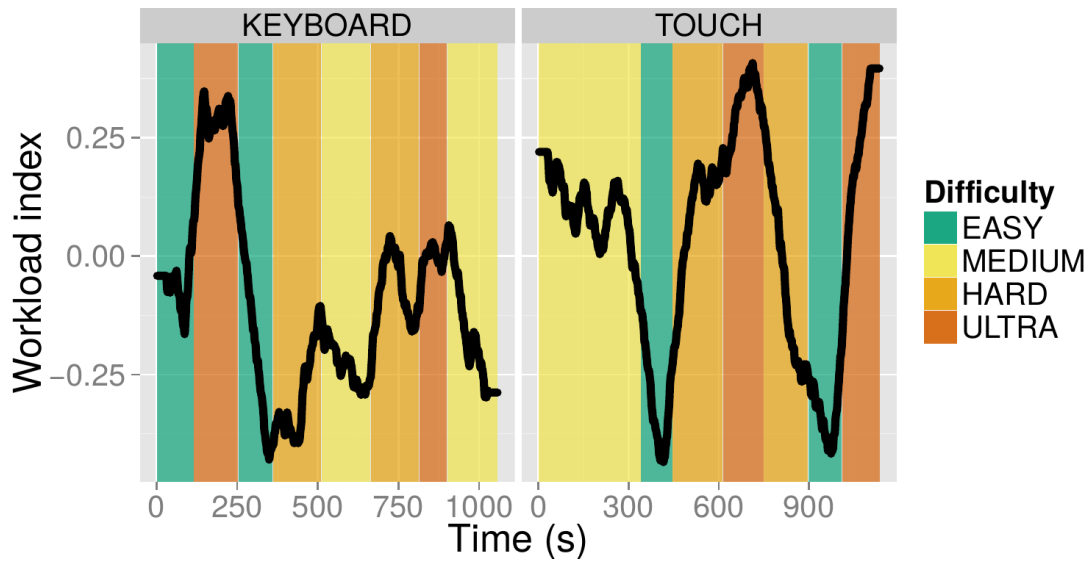


Figure 5. Workload index over time for participant 3. Left: “keyboard” condition, right: “touchscreen” condition. Background color represents the corresponding difficulty level.

Applying the same workflow as in the previous section, we managed to measure through EEG a workload index that varied depending both on the level of difficulty and on the interaction technique. The differences were significant and matched the intended design: workload increased with the difficulty and was overall more important when participants used the touchscreen, more complicated to handle on purpose. Notably, we were able to measure a near-continuous index of workload with EEG (2-second time windows). E.g. the best performer during workload calibration shows patterns that clearly meet the expectations concerning both difficulty levels and interactions, as pictured in Figure 5.

Moreover, we were able to measure two other constructs alongside workload during this study, attention and error recognition.

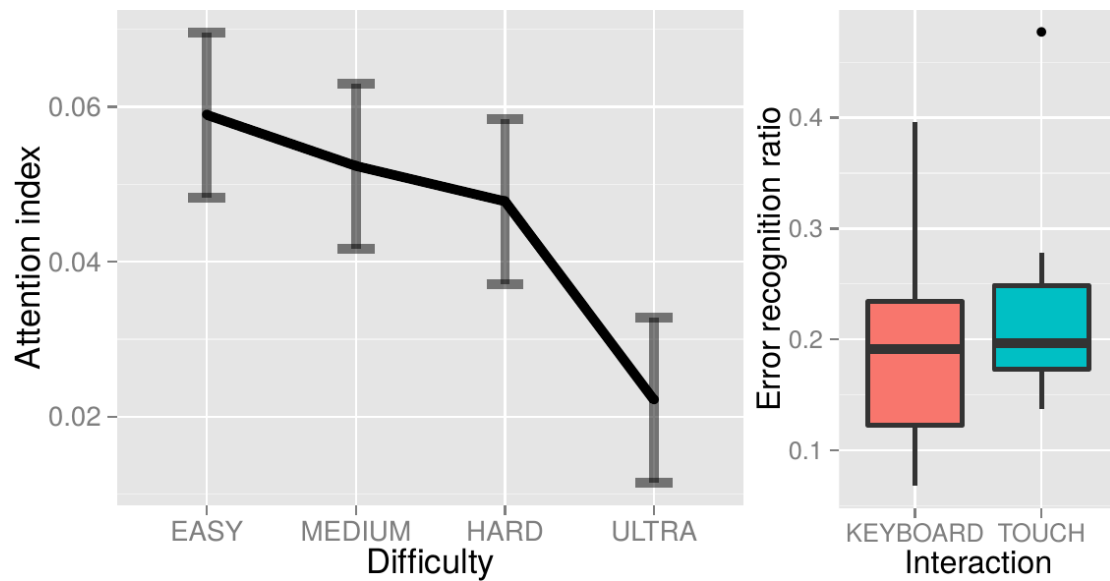


Figure 6. Left: The attention index significantly differs across difficulties. Right: The number of interaction errors differs by a tendency between interaction techniques.

Attention refers to the ability to focus cognitive resources on a particular stimulus (Kivikangas et al., 2010). In HCI, measuring the attention level could help to estimate how much information users perceive. In the present work the measure of attention relates to inattentive blindness; i.e. it concerns participants' capacity to process stimuli irrelevant to the task (Cartwright-Finch & Lavie, 2007). We calibrated and measured attention by playing “odd” sounds (high pitched bell) in the background, based on (Burns & Fairclough, 2015). Results showed that in the high difficulty condition the attention level of participants toward external stimuli was significantly lower – i.e. inattentive blindness increased (Figure 6, left).

Error recognition relates to the detection by users of an outcome different from what is expected (Nieuwenhuis, Ridderinkhof, Blom, Band, & Kok, 2001). We focused on *interaction errors* (Ferrez & Millan, 2008), which arise when a system reacts in an unexpected way, for example if a touch gesture is badly recognized. Interaction errors enable to assess how intuitive a UI is, and they are hardly measurable by another physiological signal than EEG. We calibrated error recognition using the task described in Ferrez & Millan (2008). Maybe because said task was too simplistic compared to our virtual environment – moving a 2D cursor on screen vs navigating in a 3D maze – , the number of interaction errors differed only by a tendency when we processed EEG signals, with the touchscreen that raised slightly more interaction errors than the keyboard (Figure 6, right). While tenuous, this result is coherent with our hypotheses.

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 this study was threefold. 1) The hardware we used – active electrodes with Ag/AgCl coating – is robust to cable movements, see

e.g. J. Wilson, Guger, & Schalk (2012). 2) The classifiers were trained on features not related to motion artifacts or motor cortex activation. 3) The position of the touch screen minimized participants' motion, and gestures occurred mostly before the time window used for detecting interaction errors. 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. Should artifacts bias our classifiers, differences would have appeared between the keyboard and touch screen 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.

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, or for people who design new interaction techniques.

II.2 OUTPUT DEVICES

EEG is not only an opportunity to frame more precisely the quality of input devices, it can assess dimensions of the user experience that relate to output devices as well. For instance, in Frey, Appriou, Lotte, & Hachet (2015); Frey, Appriou, Lotte, & Hachet (2016) we explored how EEG can be used to monitor visual comfort during the display of stereoscopic images. We present the first system that discriminates comfortable conditions from uncomfortable ones during stereoscopic vision using EEG.

Indeed, with stereoscopic displays a sensation of depth that is too strong could impede visual comfort and may result in fatigue or pain – whenever devices use shutter or polarized glasses (respectively “active” and “passive” displays), parallax barrier (“autoscopic” displays, e.g. Nintendo 3DS) or head-mounted displays (as with the Oculus Rift) to produce pairs of images. Some viewers could even feel pain (Lambooi, IJsselsteijn, Fortuin, & Heynderickx, 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. However, as opposed to real-life, where eyeballs rotate and eyes focus accordingly to the distance of objects, with stereoscopy the eyes remain focus on a fixed plane (the display) while the eyeballs rotate depending on the virtual depth. This discrepancy

between vergence (eyeball rotation) and accommodation (crystalline lenses that focus light beams) is known as the “vergence-accommodation conflict” (VAC).

VAC 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. VAC is one of the major causes of the symptoms associated to visual fatigue in stereoscopic displays (Hoffman, Girshick, Akeley, & Banks, 2008; Lambooij et al., 2009).

Shibata et al. (Shibata, Kim, Hoffman, & Banks, 2011) established a “zone of comfort” using questionnaires, a zone within which the apparent depth of objects should remain to avoid discomfort for viewers. Unfortunately, because of individual differences (Lambooij et al., 2009), it is 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. Finding a way to measure continuously such conflict would help to prevent the appearance of symptoms or to adapt on-the-fly displays to reduce stress when it is needed.

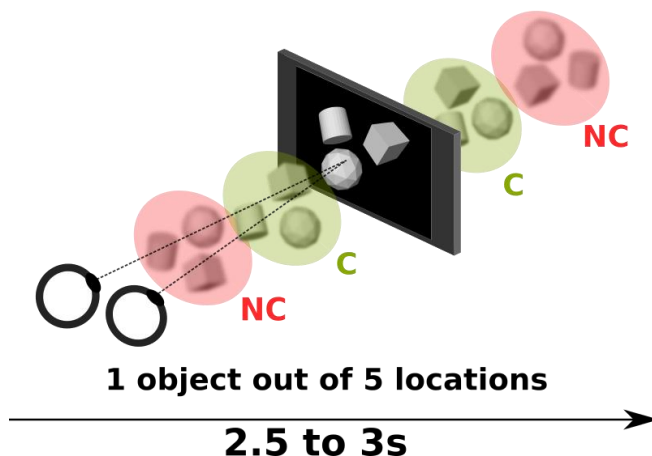


Figure 7. In order to investigate visual comfort with stereoscopic displays, still objects were presented at various virtual depths for a few seconds. C: comfortable condition; NC: non-comfortable condition.

In our experiment, we studied the appearance of virtual objects, still primitives such as cubes, cylinders or icospheres. We studied two conditions: objects appeared at a comfortable position (“C” condition, slightly before or behind the screen) or at an uncomfortable position (“NC” condition, very far away from the screen or very close to viewers) – Figure 7. Those conditions were validated with a questionnaire analogous to that of Shibata et al. (2011).

By studying the EEG activity that followed images’ appearance, we were able to train a classifier to discriminate between both conditions with an average accuracy of

63.3% in single-trial (1-second time window), a percentage significantly different from chance, that went above 70% for one third of our participants. These scores, 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 & Subramanian, 2012). The reliability of the system improved by a fair margin when we simulated the aggregation of several trials – up to 90% with 5 trials – suggesting that in a real-life scenarios the performance would increase if the system waits for several measures before making an assumption.

Using short time windows (features were extracted over 1s), we set the basics of a tool capable of monitoring user experience with stereoscopic displays in near real time. A passive stereoscopic comfort detector could potentially be useful for multiple applications, as a tool to compare with exocentric measures (possibly offline) different stereoscopic displays; or to dynamically adapt discrepancy to avoid discomfort (e.g. during 3D movies, should the adaptation be smooth and not disturb viewing), among many others.

It is interesting to note that such measure of visual comfort could well be combined with the investigation of others constructs, for example the measure of workload or attention level as described in the previous section. Encompassing several dimensions at once will help to maximize further the overall user experience, let it be beforehand during HCI evaluation or over the course of an interaction with neuroadaptive technologies (Zander & Kothe, 2011).

To summarize, EEG holds great promises to study the ergonomics pros and cons of both input and output devices and techniques for 3DUI. It can indeed provide continuous and exocentric measures of the usability of an input device, technique or task, by estimating mental states such as mental workload, inattention blindness or error recognition, even during rich and complex 3D interaction tasks. It can also provide measures of the level of comfort or quality the user perceives with the output devices and software, e.g. 3D displays, for instance by estimating how comfortable stereoscopic displays are. Naturally, these are recent works and examples, and not yet full-fledged solutions. Much still needs to be researched and developed to make EEG a fully mature evaluation tool for 3DUI. It nonetheless shows that EEG is a relevant and useful technology to do so, that needs to be further investigated, as discussed in the next sections.

PERSPECTIVES AND CHALLENGES

These different works and opportunities clearly showed the promising potential of a neuroergonomics approach based on EEG to evaluate 3DUI. However, research works in this direction are relatively recent, and there are still a number of scientific challenges that remain to be solved to be able to apply this neuroergonomics approach in practice to the design of 3DUI. In particular, there are challenges at the hardware

level – notably regarding the EEG sensors used –; at the software level – regarding EEG signal processing algorithms and the accessibility of software to use them –; at the neuroscience level – to understand further existing cognitive mechanisms and identify new ones relevant for 3DUI evaluation –; and finally at the HCI level – to use in practice neuroergonomics methods during the design of 3DUI. These various challenges are described in more details in the following sections.

III.1 HARDWARE

As with all physiological measures, one limitation of EEG comes from the hardware, mandatory to make the measures and yet not necessarily practical to use in the laboratory or in the field. In this section we summarize issues that arise from the use of EEG and we list the characteristics of a forthcoming “ideal” EEG device – see also Frey (2016) for deeper insights.

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); J. Wilson et al. (2012). Hence, some companies, while trying to transform EEG into a mass-product, bring less reliable technology to the market. Those devices often possess 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 prevents users from freely choosing their workflow, e.g. one may have to switch to another operating system because of poor support from the manufacturer.

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 where we must comply with the choice of the manufacturer – that sometimes differ from standard 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, whatever the platform or the operating system one chooses to use at a specific moment.

- 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. 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 say 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 for the technology to reach the general public – we have seen how EEG could be effective for neuroadaptive technologies.
- 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.
- Wireless, as no cables will ensure that users can move freely while interacting.
- No or little preparation. The setup time is one tedious step both for participants and for the experimenter: fill electrode one by one with gel or other conducting solution, wait for good measures, sometimes check the impedance and correct the position accordingly, etc. Such operation could last for dozens of minutes if there are many electrodes. As such, dry electrodes seem the obvious choice.

With all these observations in mind, the appearance of a community-driven initiative such as OpenBCI² is a step toward the dissemination of EEG devices. It is open-hardware, the firmware and the proposed software suite are open-source, it is nearly a hundred times cheaper than the traditional equipment found in the laboratory, and both the form-factor design and the proposed 3D printable headset oversaw many possible uses.

Such a solution has downsides, of course. First, this particular device possesses only 16 electrodes, which limits its use – most of the work we presented in this paper relied on a 32 channels amplifier. The more channels we have the better, since a better

2 <http://openbci.com/>

scalp coverage enables more accurate measures. However, as mentioned at the beginning of this section, when the EEG markers are known a good placement can overcome this difficulty. More importantly, the practicality of the device should not let one forget about the reliability of the measures. While the first comparisons we made are encouraging and suggest very close recordings between the OpenBCI board and a medical grade equipment (Frey, 2016), more thorough examinations are required before we could exactly know the advantages and the limitations that more users-driven solutions could offer to neuroergonomics.

III.2 SOFTWARE

At the software level, the main challenges are related to EEG signal processing, and in particular about how to extract reliable estimates of cognitive processes and constructs from EEG signals, despite their noisy and non-stationary nature. These two issues with EEG are particularly prominent when using EEG to study 3DUI. Indeed, EEG signals are rather sensitive to noise and artifacts, see, e.g. Goncharova, McFarland, Vaughan, & Wolpaw (2003), such as motion artifacts (due to users' movements), muscle tension artifacts (e.g. from face muscles) or electromagnetic interferences from nearby electronic equipment. Contrary to a lab 2DUI experiment in which the user is quietly sitting in front of a standard screen, all these types of noise are likely to occur with 3DUI. Indeed, with touch devices to a minor extent or, to a larger one, with full virtual reality (VR) equipment, the user will be moving more, or even standing or walking while using the 3D device, hence resulting in potential motion artifacts. Then, the use of heavy 3D equipment such as VR caves, 3D shutter glasses or unconventional interaction devices such as haptic input devices or trackers, may lead to stronger electromagnetic interference. Finally, the richness and complexity of the 3D environment may lead to stronger emotions in the user, e.g. frustration or joy, which in turns could lead to stronger facial expressions and thus facial muscle artifacts. Although, to the best of our knowledge, there have been no report of strong EEG signal degradation due to the first two types of noise (motion artifacts and electromagnetic interference) during 3D interaction, developing EEG signal processing tools to clean such artifacts or to estimate cognitive states while being robust to them, should improve the reliability of the whole system.

Another reliability issue with EEG signals is their non-stationarity. This can be due to the electrodes' gel drying, or to changes in the fatigue or attention level of the user over the course of the experiment (Roy, Bonnet, Charbonnier, & Campagne, 2013). As far as 3DUI is concerned, there is another source of non-stationarity that is prominent and a major problem: non-stationarity due to changes in context. Indeed, designing a cognitive state estimator, e.g. a workload level estimator, requires calibrating such estimator on examples of EEG signals from the user. To do so, classical cognitive state induction tasks are used, such as the N-back task for workload (Mühl et al., 2014; Wobrock et al., 2015). Indeed, these validated tasks ensure a proper calibration of the system as they ensure the EEG signals collected really correspond to the cognitive state we want to estimate in the EEG signals. However, these tasks are

very different from the context in which the cognitive state estimator will be used. For instance, whereas the N-back task only involves letters' appearance, during a 3D interaction task, the user will be much more active, and confronted to a much richer environment and stimuli, which will change her/his EEG signals. As such, directly using an estimator calibrated on an N-back task to estimate workload in a complex 3D interaction task may simply fail. To avoid this problem, we designed a specific EEG signal processing algorithm to calibrate spatial filters robust to such context changes for mental state estimation based on EEG oscillations (Frey et al., 2016; Wobrock et al., 2015). In the future, it would be necessary to design similar algorithms to estimate mental states based on event related potentials (ERP), as well as to calibrate not only spatial filters that are robust to context changes, but also temporal filters and classifiers.

The calibration of the cognitive state estimators is also an issue in itself. Indeed, this calibration must be performed for each cognitive state we want to study and for each user. This is very time consuming and limits the number of states that can be studied simultaneously (e.g. in Frey et al. (2016) we studied 3 states simultaneously, which was already more than most works which study a single one), and may exhaust the user due to their length. It is therefore necessary to design EEG signal processing tools with reduced calibration time, or, ideally, no calibration at all - a so-called subject-independent estimator. Recent works in that direction in the field of EEG-based Brain-Computer Interfaces should therefore be pursued (Lotte, 2015).

Finally, to make these neuroergonomics technologies practical, it is necessary to make them easily accessible and usable. In other words, 3DUI designers should be able to use these technologies without deep understanding of EEG or signal processing. Ideally, they should be usable with just a few clicks. Research and Development works should therefore be conducted to create easily usable software to perform EEG-based neuroergonomics evaluations. Solutions such as the open-source, real-time neuroscience OpenViBE software, are a first step in that direction (Renard et al., 2010).

III.3 NEUROSCIENCE

At the neuroscience level, there is also necessary research to be conducted, in particular to know more about cognitive states that could be relevant to study 3DUI, and how they are reflected in the EEG signals. In particular, there are a number of cognitive states that are relevant to assess interaction, and that could be estimated to some extent in average EEG signals, at least during lab experiments (Frey et al., 2014). It remains to be shown whether they could be also estimated continuously, in single-trial, during complex, real-life interaction tasks.

Notably, there are some evidences that the extent of error recognition - i.e. how much the interaction outcome differs from the expected one, is reflected in scalp EEG signals (Frey et al., 2014). It remains to be shown whether it can be estimated continuously for each interaction task, for real applications. If that is so, this could be a very interesting measure of how intuitive is a 3DUI (or any UI by the way). Similarly, being able to measure the user's emotion in EEG signals, could provide valuable

insights into the user experience. Although, EEG-based emotion recognition is still a challenging task so far (Mühl, Allison, Nijholt, & Chanel, 2014), this could potentially inform us about how enjoyable or frustrating a UI is.

III.4 HUMAN-COMPUTER INTERACTION

In the field of Human-Computer Interaction, evaluation of interfaces and systems play a central role. It allows validating new concepts, and it guides designers in their general conception choices. To achieve robust assessments that are scientifically valid, evaluation protocols need to be designed with great care. This notably concerns the tested population, the definition of the experimental conditions, the sequence in which different parameters are tested, and so on. It also concerns the data that is stored, and the way it is analyzed. Even if each experiment is different, there are general rules and practices that are widely shared among the community.

In the future, with the emergence of new physiological assessments as the ones we described above, the standard evaluation protocols will evolve. First experiments that are done in this direction are highly customized. A big challenge will be to standardized the protocols based on EEG ergonomics for tomorrow evaluations. A set of guidelines and standard procedures should be defined to help researchers in their evaluation work.

Compared to standard evaluation protocols, EEG-based protocols have several specificities that need to be taken into account. First, as stated in Subsection [Software](#), the recorded data is strongly subject dependent, meaning that at the moment it is hardly possible to transfer a calibration from one participant to another – or even from one session to another – without great loss of accuracy. Even when gamification is used to make calibration tasks less tiresome (Flatla, Gutwin, Nacke, Bateman, & Mandryk, 2011) or when calibration tasks are alternated with the main interaction (Frey et al., 2016), this requirement is tedious for participants.

Second, the need for an accurate synchronization between, from the one hand, EEG recordings and, from the other hand, the interactions performed by participants may also necessitate to tune the software that is being evaluated. For instance, each event of interest – e.g. target selection, beginning of a new task – should be associated to timestamps for offline analyses, or should “ping” the program that is recording EEG for online analyses. The need for such precise logging information is nowadays easier to fulfill thanks to protocol alike LSL³, that posses simple bindings in many programming languages, but it should still be considered beforehand when experiments are designed.

EEG-based evaluations bring new constraints. They also open new opportunities for HCI. In particular, they will allow fine assessments of the user interaction. Hence, UI designers will be able to inject the results of the experiments in the developed

3 <https://github.com/sccn/labstreaminglayer>

technique at an early stage of the creation process. We may assume that, with such kind of EEG-based evaluation – if they are powerful enough –, several cycles of standard evaluations, which would require a large number of users, could be saved. These new practices that redefine standard evaluation and design approaches should progressively appear in the research community, in conjunction with more traditional approaches.

CONCLUSION

We demonstrated how brain signals – recorded by means of electroencephalography – could be put into practice in order to gather insights about multiple dimensions of the user experience, whether input or output devices are considered, with realistic interaction tasks or novel interfaces. As a supplementary evaluation method, EEG brings continuous and exocentric measures that does not disrupt tasks performed by the users. So far we have investigated workload, attention and visual comfort, and we got encouraging results with interaction errors, a metric that relates to how intuitive an interface is.

While challenges lie ahead and need to be overcome for the methodology to truly disseminate – e.g. features robust to context changes or faster setup times – the increasing availability of EEG devices opens new possibilities. Thanks to a deeper understanding of users, EEG-based neuroergonomics will lead to better HCI in general, and, since 3D interactions are more demanding, to better 3DUI in particular.

REFERENCES

- Anderson, E. W., Potter, K. C., Matzen, L. E., Shepherd, J. F., Preston, G. A., & Silva, C. T. (2011). A user study of visualization effectiveness using EEG and cognitive load. In *Eurographics* (pp. 791–800).
- Bangor, A. W. (2001). *Display technology and ambient illumination influences on visual fatigue at VDT workstations* (thèse de doctorat). Virginia Tech.
- Blankertz, B., Tangermann, M., Vidaurre, C., Fazli, S., Sannelli, C., Haufe, S., ... Müller, K.-R. (2010). The Berlin Brain-Computer Interface: Non-Medical Uses of BCI Technology. *Front Neurosci*, 4(December), 198. <http://doi.org/10.3389/fnins.2010.00198>
- Bowman, D. A., Gabbard, J. J., & Hix, D. (2002). A Survey of Usability Evaluation in Virtual Environments: Classification and Comparison of Methods. *Presence*, 11(4), 404–424.
- Bowman, D. A., Kruijff, E., LaViola Jr., J. J., & Poupyrev, I. (2005). *3D User Interfaces: Theory and Practice*. Addison-Wesley/Pearson Education.

- Burns, C. G., & 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.
- Cartwright-Finch, U., & Lavie, N. (2007). The role of perceptual load in inattentive blindness. *Cognition*, 102(3), 321–340.
- David Hairston, W., Whitaker, K. W., Ries, A. J., Vettel, J. M., Courtney Bradford, J., Kerick, S. E., & McDowell, K. (2014). Usability of four commercially-oriented EEG systems. *Journal of Neural Engineering*, 11(4), 046018. <http://doi.org/10.1088/1741-2560/11/4/046018>
- Ferrez, P. W., & Millan, J. D. R. (2008). Error-related EEG potentials generated during simulated brain-computer interaction. *IEEE Trans. Biomed. Eng.*, 55(3), 923–9.
- Flatla, D. R., Gutwin, C., Nacke, L. E., Bateman, S., & Mandryk, R. L. (2011). Calibration games: making calibration tasks enjoyable by adding motivating game elements. *UIST '11*, 403–412.
- Frey, J. (2016). *Leveraging human-computer interactions and social presence with physiological computing* (thèse de doctorat). University of Bordeaux, France.
- Frey, J., Appriou, A., Lotte, F., & Hachet, M. (2015). Estimating Visual Comfort in Stereoscopic Displays Using Electroencephalography: A Proof-of-Concept.
- Frey, J., Appriou, A., Lotte, F., & Hachet, M. (2016). Classifying EEG Signals during Stereoscopic Visualization to Estimate Visual Comfort. *Computational Intelligence & Neuroscience*.
- Frey, J., Daniel, M., Hachet, M., Castet, J., & Lotte, F. (2016). A framework for electroencephalography-based evaluation of user experience.
- Frey, J., Mühl, C., Lotte, F., & Hachet, M. (2014). Review of the use of electroencephalography as an evaluation method for human-computer interaction. In *Proceedings of the International Conference on Physiological Computing Systems* (pp. 214–223).
- Goncharova, I., McFarland, D., Vaughan, T., & Wolpaw, J. (2003). EMG contamination of EEG: spectral and topographical characteristics. *Clinical Neurophysiology*, 114(9), 1580–1593. [http://doi.org/10.1016/S1388-2457\(03\)00093-2](http://doi.org/10.1016/S1388-2457(03)00093-2)
- Grimes, D., Tan, D., & Hudson, S. (2008). Feasibility and pragmatics of classifying working memory load with an electroencephalograph. *CHI '08*, 835.
- Hart, S., & Staveland, L. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Human mental workload*.

- Hoffman, D. M., Girshick, A. R., Akeley, K., & Banks, M. S. (2008). Vergence-accommodation conflicts hinder visual performance and cause visual fatigue. *Journal of Vision*, 8(3), 33.1–30. <http://doi.org/10.1167/8.3.33>
- Jankowski, J., & Hachet, M. (2013). A survey of interaction techniques for interactive 3d environments. In *Eurographics 2013-STAR*.
- Kirchner, W. K. (1958). Age differences in short-term retention of rapidly changing information. *Journal of Experimental Psychology*, 55(4), 352–358. <http://doi.org/10.1037/h0043688>
- Kivikangas, J. M., Ekman, I., Chanel, G., Järvelä, S., Cowley, B., Henttonen, P., & Ravaja, N. (2010). Review on psychophysiological methods in game research. *Proc. of 1st Nordic DiGRA*.
- Lambooi, M., IJsselsteijn, W., Fortuin, M., & Heynderickx, I. (2009). Visual Discomfort and Visual Fatigue of Stereoscopic Displays: A Review. *Journal of Imaging Science and Technology*, 53(3), 030201. <http://doi.org/10.2352/J.ImagingSci.Technol.2009.53.3.030201>
- Lotte, F. (2015). Signal Processing Approaches to Minimize or Suppress Calibration Time in Oscillatory Activity-Based Brain–Computer Interfaces. *Proceedings of the IEEE*, 103(6), 871–890.
- Mühl, C., Allison, B., Nijholt, A., & Chanel, G. (2014). A survey of affective brain computer interfaces: principles, state-of-the-art, and challenges. *Brain-Computer Interfaces*, 1–19.
- Mühl, C., Jeunet, C., & Lotte, F. (2014). EEG-based Workload Estimation Across Affective Contexts. *Frontiers in Neuroscience*, 8, 114.
- Nieuwenhuis, S., Ridderinkhof, K. R., Blom, J., Band, G. P., & 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.
- Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psych. Rev.*, 84(3), 231–260.
- Parasuraman, R., & Wilson, G. F. (2008). Putting the brain to work: Neuroergonomics Past, Present, and Future. *Human Factors*, 50(3), 468–474.
- Peck, E. M. M., Yuksel, B. F., Ottley, A., Jacob, R. J., & Chang, R. (2013). Using fNIRS brain sensing to evaluate information visualization interfaces. In *Proc. CHI* (pp. 473–482).
- Picard, R. W. (1995). *Affective computing* (No. 321). MIT Media Laboratory.

- Renard, Y., Lotte, F., Gibert, G., Congedo, M., Maby, E., Delannoy, V., ... 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. <http://doi.org/10.1162/pres.19.1.35>
- Roy, R. N., Bonnet, S., Charbonnier, S., & Campagne, A. (2013). Mental fatigue and working memory load estimation: Interaction and implications for EEG-based passive BCI. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, 6607–6610. <http://doi.org/10.1109/EMBC.2013.6611070>
- Shibata, T., Kim, J., Hoffman, D. M., & Banks, M. S. (2011). The zone of comfort: Predicting visual discomfort with stereo displays. *J. Vis.*, 11(8), 11–11. <http://doi.org/10.1167/11.8.11.Introduction>
- 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. <http://doi.org/10.1155/2010/630649>
- Vi, C., & Subramanian, S. (2012). Detecting error-related negativity for interaction design. *CHI '12*, 493. <http://doi.org/10.1145/2207676.2207744>
- Wilson, J., Guger, C., & Schalk, G. (2012). BCI hardware and software. In *Brain-computer interfaces: principles and practice* (pp. 165–188).
- Wobrock, D., Frey, J., Graef, D., Rivière, J.-B. de la, Castet, J., & Lotte, F. (2015). Continuous Mental Effort Evaluation during 3D Object Manipulation Tasks based on Brain and Physiological Signals.
- Zander, T. O., & 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. <http://doi.org/10.1088/1741-2560/8/2/025005>