MAUU(D)5900
MASTER THESIS
in
Universal Design of ICT

May 2018

Exploring task design for Cognitive Commands and Prerequisites of Accessibility with a Visual Imagery task for Spontaneous EEG

Nina Bauge

Department of Computer Science

Faculty of Technology, Art and Design

OSLO AND AKERSHUS
UNIVERSITY COLLEGE
OF APPLIED SCIENCES

PREFACE

This master thesis is a continuation of a topic I have been engrossed with since my undergrad project, and I have been happy to get to explore the field again. I don't feel that I have exhausted the learning potential in this project. For every piece of understanding I gain about the field of theory, or even about performing research, I find that there is far more I haven't approached yet. Given the opportunity I would gladly research Brain Computer Interfaces and intrinsic motivation in HCI for years to come.

I am grateful to have enjoyed the support of an attentive and invested supervisor, Tulpesh Patel, and my two-year-old son who decorate my notes and books with bright crayon colours.

Nina Bauge

ABSTRACT

I have looked at skills required for end-users to operate a BCI communication system, attempted to define optimal cognitive commands for task design with Stimulus-Response compatibility, and discussed how hardware and system design may support accessibility. I have performed experiments with an Emotiv EEG headset and a game interface to find and the participants have reported perceived Workload with a NASA Tlx form. I have assessed the effectiveness and user experience of spontaneous EEG, by testing training data with different cognitive tasks and compared them with number of classifications per session, and the score achieved. There are benefits in user motivation and workload if we can present adjust tasks to skill level and customize stepwise skill acquisition to the individual user. I have hypothesised that a trade-off in dataset directionality may be outweighed by a more accessible workflow, and that different sets of training data are sufficiently equal in efficiency and accuracy. There is no significant evidence of a difference in tested efficiency or accuracy between using visual imagery and visual perception as training data with visual imagery gameplay task. There is no significant evidence of a difference in perceived workload between playing with different sets of training data.

CONTENTS

1 Preface ___ 3

2 Abstract ___ 5

4 Tables ___ 9

5 Figures ___ 10

6 Introduction ___ 11

6.1 Third Principle of Universal design ___ 12

7 Literature review ___ 13

7.1 Development of BCI Technology ___ 13

7.2 Classification and Machine Learning ___ 15

7.3 Signal stimulation ___ 16

7.4 Dependence on internal aspects of normal brain function ___ 17

7.5 Stepwise skill acquisition as a design approach ___ 21

8 Methods ___ 26

8.1 Participants ___ 26

8.2 Tools / Measures ___ 27

8.3 Experiment Procedure ___ 30

9 Results and findings ___ 32

9.1 RQ ___ 32

9.2 Utility H1 ___ 32

9.3 User experience H2 ___ 33

9.4 Relationship between Utility and User experience H3 ___ 33

10 Discussion ___ 36

10.1 Findings ___ 36

10.2 Limitations of the study ____ 43

10.3 Future iterations of my experiment ____ 44

11 Conclusion ____ 47

12 Reference list ____ 48

A. Nasa TLX ____ I

B. Data from gametask in all sessions ____ II

C. NSD ____ III

12.1 Performed testing across participants ____ III

D. Prototype code ____ V

E. Resultater ____ XXIII

F. Sample t-test ____ XXIV

TABLES

Table 8.1Summarized notes from testing 39
Table 8.2 Time spent with Participants in BCI studies 42

FIGURES

Figure 7.1 EEG signal frequencies 14

Figure 7.2 Example of signal resolution and aliasing ("EEG > introduction Biomedical Signals Acquisition," 2005) 14

Figure 7.3 General scheme of an EEG-based BCI(Bobrov et al., 2011) 16

Figure 8.1 10-20 system("EEG > introduction Biomedical Signals Acquisition," 2005) 27

Figure 8.2 FlappyBird screenshot("stikling/FlapPyBird: A Flappy Bird Clone using python-pygame," n.d.) 28

Figure 8.3: Y-axis position post command, 1 second timeline 29

Figure 9.1 Figure showing the means of Scores with both visual perception (Method 1) and visual imagery (Method 2) 32

Figure 9.2 R2 and p values for correlation between score and workload. 34

Figure 9.3 Linear regression of Workload and Score. 35

Figure 9.4 Boxplots of Workload for conditon 1 and 2, h2 35

Figure 10.1 Flow model (Csikszentmihalyi, 1997 40

Figure 10.2 Total time spent with tasks each session 43

Figure 12.1 Text outout of ttests for h1 and h2 XXIV

Figure 12.2 TwoSample t-test definition from MatLab XXIV

INTRODUCTION

20% of the general population is believed to be EEG “illiterate”. Some studies claim that everyone could master EEG if given the correct training and an accessible system (Lotte, Larrue, & Mühl, 2013). This study will focus on a potential user prerequisite for successfully using EEG signal as a BCI technology. There are theories about how to solve issues during training and recording that haven’t been tested. This study investigated training methods for a spontaneous EEG system based on theories about training and visual imagery (Ganis, Thompson, & Kosslyn, 2004.

Persons with Complex Communication Needs (CCN) and intact cognitive abilities may be candidates for Augmentative and Alternative Communication (AAC) systems with BCI. BCI systems are widely used for research purposes, and some systems are developed for communication purposes. BCI communication systems are currently limited and fail to support the extent of user's abilities. The long-term goal for HCI development of BCI AAC systems is to introduce BCI technology in a manner that ensure technology acceptance and encourage skill development. In this section we are going to represent some of the background for the research questions and present associated concepts.

Several studies suggest spontaneous EEG for BCI to bridge an ability gap for users with severe motor impairments (Beukelman, Fager, Ball, & Dietz, 2007; Bobrov et al., 2011; Jure, Carrere, Gentiletti, & Tabernig, 2016; Kübler et al., 2014). There are assistive technologies meant to alleviate the difficulties on a wide spectrum of motor impairments, such as text to speech and eye-tracking(Beukelman et al., 2007). For some patients with more acute impairments, bordering on Locked-n Syndrome or Total Locked-in Syndrome1, Spontaneous BCI might be a last stronghold before TLIS. Also, systems for spontaneous BCI are developed for gaming or interaction purposes for ordinary users as well.

The users’ task is to create a pattern of brain activity that they can recreate and repeat at will. End users are faced with the challenge to produce and recreate clear brain patterns. (Lotte et al., 2013). As people don’t have much insight into their brain activity we need a strategy. There are three main approaches for spontaneous EEG: motor imagery, visual imagery and auditory imagery. Either you imagine moving a limb, to see something or hear something. In this paper, we will focus mostly on visual imagery as a use-case for controlling a computer. In theory (Ganis et al., 2004), a brain will emit the same signals whether you look at a picture or imagine looking at it. This is referred to, respectively, as visual *perception* and visual imagery with EEG. If a user is good at imagining pictures at will, and the computer recognize those brain signals, the user can do this deliberately to perform classified functions (Bobrov et al., 2011). Such as 'move a cursor', or 'push a button'. A goal in the data-acquisition-step, or data entry, is to collect robust recordings, with clear defining features, there are easily recognized and discriminated. What we want is as much recording of brain activity that are relevant for each visualization, rather than the remaining activity; the noise. This is referred to as signal-to-noise-ratio. Visual imagery is related to several areas in the brain. The frontal lobe, parietal lobe, temporal lobe, occipital lobe and posterior cingulate.

Third Principle of Universal design

When discussing Load in relation to Universal design in relation to cognitive load, it is first and foremost the third principle that is relevant. This is the principle of Simple and Intuitive Use; "Use of the design is easy to understand, regardless of the user's experience, knowledge, language skills, or current concentration level." It comes with a suggested set of guidelines. "3a. Eliminate unnecessary complexity. 3b. Be consistent with user expectations and intuition. 3c. Accommodate a wide range of literacy and language skills. 3d. Arrange information consistent with its importance. 3e. Provide effective prompting and feedback during and after task completion." In this project I attempt to explore and apply these principles to accompanying fields of theory.

LITERATURE REVIEW

There are currently several research groups who work with questions related to EEG interactions and AAC. BCI development is now considered a multidisciplinary field for linguists, neurologists', informatics, Ergonomics and psychology (Ball, Beukelman, & Pattee, 2004; Curran & Stokes, 2003; Gramann, Fairclough, Zander, & Ayaz, 2017; A. Kübler & Birbaumer, 2008; Lotte, Larrue, & Mühl, 2013; McFarland & Wolpaw, 2011).

Wolpaw et al published an extensive review on BCI for communication and control, for the journal Clinical Neurophysiology (McFarland & Wolpaw, 2011). They outline important considerations for further BCI development. I will present them sorted by Development of Technology, End-User Accessibility and Research specific considerations.

Development of BCI Technology

General principles of EEG

EEG is biometric data, collected from electrical currents emitted by neural activity in the brain. The activity can be recorded and monitored with sensors on the scalp. ("EEG > introduction Biomedical Signals Acquisition," 2005). This is a short introduction to present some of the underlying premises necessary for successful EEG classification.

Initially, all signals are sampled relative to a baseline signal. The baseline signal is the background activity unrelated to intended activity. Offset is fluctuations in the baseline signal, and the baseline signal may need to be adjusted from time to time.

Signal display properties such as rhythmic, arythmic or dysrythmic patterns. On top of this the signals hold morphologic attributes; the different shapes of the waveforms. Waves may be Transient, Monomorphic or Polymorphic EEG activity. While Monomorphic activity can be composed out of one activity, polymorphic activity is a complex waveform composed by multiple frequencies. While sinusoidal waves resemble sine waves, Transient patterns are distinctly separable as either Spikes or Sharp waves and are characteristic by their duration. Cognitive commands typically aim to classify Transient patterns. If the transient patterns occur with some intensity over several regions of the brain simultaneously, it is referred to as Hypersynchronous or Paroxysmal.

The signal intensity is measured in microvolts (µV). Signal wavelength is measured in signal frequency (Hz). Increase or decrease of signal intensity can be ascribed to stimulation or abnormal activity. Relevant terms cover abrupt gain in voltage; Paroxysmal gain, a general increase in voltage and rhythm; hypersynchrony, as well as decreased voltage referred to as Attenuation or blocking. These indicate some specific activity or stimulation. While transient spikes last between 20 and 70 msec, sharp waves last between 70 and 200 msec. To provide accurate representations of changes in signal intensity, samples must be collected more than twice as often as the fluctuations on Transient patterns.

Figure 0.1 EEG signal frequencies

	Frequency
Delta	< = 3 Hz
Theta	3.5 – 7.5 Hz
Alpha	7.5 – 13 Hz
Beta	> 14 Hz

Signal acquisition and processing

The first level of selection of signal types is positioning and wavelength range selections. The 10-20 system is a defined standard for sensor placement, defined in the guidelines SEPN. (Standard Electrode Position Nomenclature). The standard positions are reference for positions, relative to the nasal bones and the occipital bone at each side of the head, and from ear to ear. Their names are initials for the regions they are associated with. When designing an interaction, it is useful to choose sensors placements correlating with the desired signal types. The number of sensors will also contribute to the overall volume of data and resolution. The main frequencies are Delta, Theta, Alpha and Beta. The different frequencies are associated with different characteristics of activity in different mental states, ages and display of potential pathology.

A-D conversion is conversion of Analogue, “real” signals as they are detected, into numerical values fit for digital processing. Filtering parts of the bandwidth provide easier processing of desired frequencies. This function may be performed while collecting the analogue signal, by analogue filtering before digitalization, or with a digital filter. Filtering primarily remove noise, as well as restrict the total amount of data collected so that it takes less computing power and less time to converse and process.

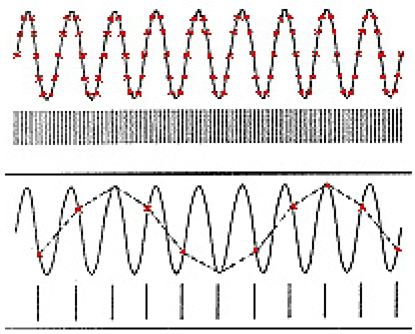

Figure 0.2 Example of signal resolution and aliasing (“EEG > introduction Biomedical Signals Acquisition,” 2005)

For instance, a standard band filter at 50-70 Hz may be applied during signal acquisition to remove background white noise, which occurs at about +- 60 Hz. Desired wavelengths may be amplified to increase voltage and signal attributes. Amplification may provide an added voltage, a higher signal intensity, to chosen wavelengths.

While less EEG data ease demand on computing power and -time, a minimum of bit data is required to sustain necessary resolution. Loss of data to less than required resolution may lead to amplitude saturation, and misclassification from aliasing. Amplitude saturation causes the signal to plateau from amplification, rather than provide a more intense signal. Aliasing means that the digital representations of wavelengths have a sample rate less than the signal rate, which means that the signal shape and rhythm may be misrepresented. Accurate resolution requires accessible sample representations of a voltage range twice as large as the digital range. The resolution requires Required sampling rate to be at least twice that of signal rate, as stated in Nyquist sampling rate. Sample interval is inverse to the rate. For sharp wave signals with durance from 70 to 200 msec, the sampling interval need be at least half that; this means sampling interval of 35 msec or less.

$$resolution = \frac{voltage\ range}{digital\ range\ =^{\#2bits}}$$

Noise

Noise is activity in the sampled signals that are not representative for the signal intent. Noise is categorized as environmental noise, system noise and user generated noise. Filtering handle a large part of bandwidth noise associated with the former, while the latter is made up from different signal types, artefacts, distractions and the overall difficulty producing stable, predictable brainwaves.

Classification and Machine Learning

When first forming a classification model with a support vector machine, the operator defines a bias, by assigning an appropriate class, then provide data with trait variation that is representative for the class. This is training data. The next step is to evaluate the classifier and tune parameters until predictions are sufficiently accurate for the purpose.

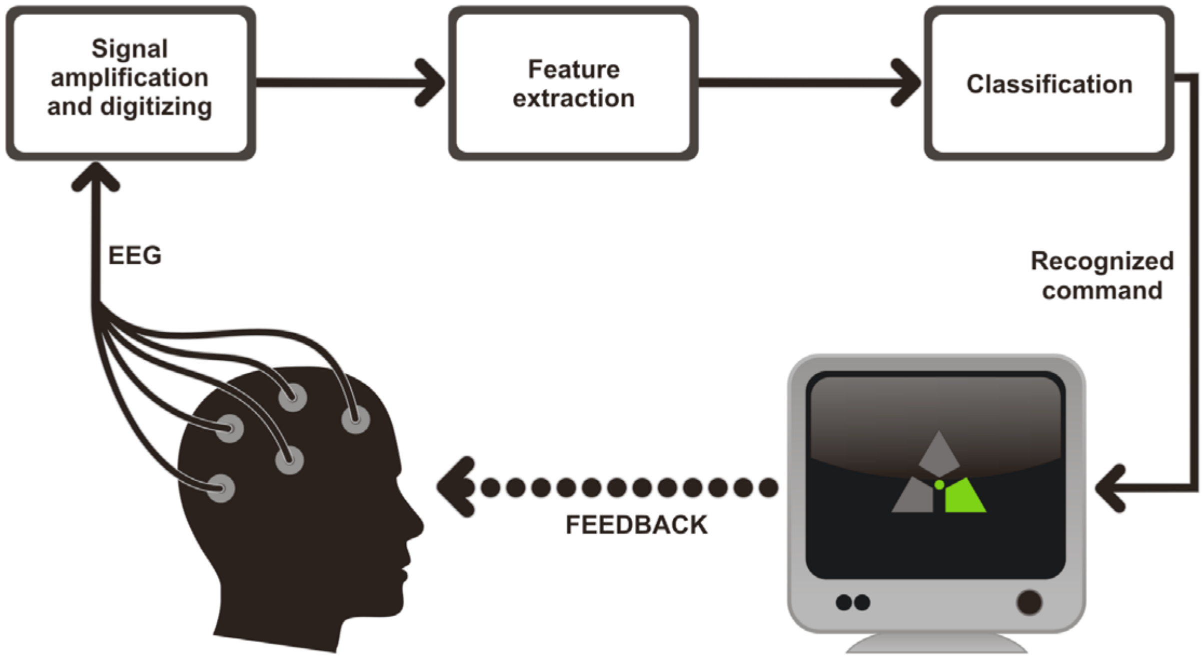

Figure 0.3 General scheme of an EEG-based BCI(Bobrov et al., 2011)

For accurate classification the EEG data stream must hold a sufficient similarity of traits with the trained datasets. The classification model makes predictions from transient patterns and synchronicity in the EEG signal data stream. If the classifier finds a pattern in the data stream has a high probability of matching a trained class, then the associated functions of the class are triggered.

Signal stimulation

Signal acquisition is another aspect of the three-way relationship between: (i)the mental states and processes that provide the content and motivation for the communication (ii)the brain activity, as recorded by EEG, being utilised to drive the device (iii)the voluntary manipulation and control of that brain activity(Curran & Stokes, 2003).

One of the first adaptations available during signal acquisition is signal stimulation as they are generated by the user. Signals can be divided into categories from Dependent and Independent BCI(Anderson et al., 2011; McFarland & Wolpaw, 2011). Methods utilizing Dependent BCI, such as Evoked Potentials, ‘depends’ on unaltered reactions and outputs. Neural artefacts directly related to perceptions and mental chronometry of motor operations fall into this category. Evoked potentials refer to the process of provoking, or “evoking” these artefacts, such as Visual Evoked Potentials (VEP). Independent BCIs, however, concern a capturing of the users’ intent and is the wilful mental command of forcing your brain to output specific patterns. This is also called Spontaneous BCI, as it is the users initiative that precede the occurrence of such signals. One utilization of Spontaneous, independent EEG is Imagery. Motor imagery is a common approach, frequently applied in BCI experiments (Bashashati, Ward, Birch, & Bashashati, 2015; Bashashati et al., 2015; Curran & Stokes, 2003; Friedrich, Scherer, & Neuper, 2013; Leeb et al., 2013; Zickler, Halder, Kleih, Herbert, & Kübler, 2013). In motor imagery, the user attempt to imagine and recreate patterns of imagined movement of limbs. For instance, a user imagine that he moves a finger to “push a button”. Another form of imagery is visual imagery.

BCI Tasks with visual imagery.

In a study with fMRI, Ganis et al monitored activation in different brain regions during diverse tasks, two of which was imagined and evoked visual signals. They remarked; “Imagery and perception activated frontal structures in remarkably similar ways; in all of the regions we examined, the spatial pattern of activation was identical”(Ganis, Thompson, & Kosslyn, 2004). Their finding could indicate that Imagery and Perception tasks emit activities so similar that they might be used interchangeably when measured in the frontal regions. This turned out to be, at least in part, transferrable to EEG: Dentico et al hypothesised and concluded that the neural flow changed directionality, depending on whether participants performed visual imagery or visual perception(Dentico et al., 2014). The active regions of visual stimulus are hypersynchronous transient patterns that may be both evoked and improvised. This have also been tested as a cognitive command with spontaneous EEG. Bobrov presented Visual Mental Imagery as a Spontaneous EEG Task strategy(Bobrov et al., 2011), where users emitted cognitive command by visualizing either a house or a car.

Dependence on internal aspects of normal brain function

“Dependence on normal brain function” is a partially defined cognitive prerequisite for using a BCI system, and enable a three-way relationship between: (i)the mental states and processes that provide the content and motivation for the communication, (ii)the brain activity, as recorded by EEG, being utilised to drive the device, (iii)the voluntary manipulation and control of that brain activity(Curran & Stokes, 2003). BCI ‘literacy’ is allegedly to about 80% in the general population, where the remaining 20% are considered BCI illiterate(Allison & Neuper, 2010). What does this literacy or illiteracy consist of?

Systemic ability gaps

Allison and Neuper emphasized that initial difficulties in using a BCI system are reinforced by limited BCI functionality. They compared BCI accessibility and ability to actual literacy; if a writer spells half the words in a paragraph correctly, he is likely to have communicated its intention and can be said to be 50 % literate. In contrast, a user that only express their intention correctly half of the time in a two-options AAC interface, experience 0% literacy. This is because the limitations of the AAC vocabulary don't provide linguistic context to make the text comprehensible. They argued that the different BCI systems cater to different abilities; A user might not be able to use all BCI systems, but it is likely that most of us can use at least one. This support an assumption that the context of a system, BCI task design and recognition of ability gaps, precedes a definition of required user ability.

Anatomical ability gaps

One size does not fit all. Anatomical differences may cause some users to have difficulties with a shelf ware headset. Ill fit with equipment may cause challenges with sensor placement. Structural differences in a robust nasal bone or occipital bone may block EEG signal acquisition with predefined, generic A-D and signal amplification settings.

Lack of skill or training

There is still no consensus in BCI research about the kind of skill or skills that must be acquired in order to successfully drive a BCI system(Curran & Stokes, 2003). Dependent signals are straightforward in the sense that they hardly require skill at all, while Independent signals must be provoked at will by the user, and as such the skill depends on the task design. Tasks with visual imagery incorporate spatial tasks and require spatial abilities. 'Encyclopaedia of the Sciences of Learning' defines Spatial Ability as "[…] the perception and/or mental manipulation of visual stimuli. It may also include mentally rehearsing a visual experience […]."(Zydney et al., 2012). Then the user need Spatial Ability to acquire Spatial skills.

Ability and Skill acquisition

Emotional regulation has been accepted as a requirement for Spatial visualization. Executive functions cover cognitive control, self-regulation, attention, working memory, fluid intelligence, inhibitory control, task switching, mental flexibility, creativity and reasoning (Diamond, 2013).Human factors cover research on awareness, attentional control and automaticity (Proctor & Vu, 2010). Complex cognitive challenges and skills are structured and usually handled by neural mechanisms in the Prefrontal Cortex. This is also referred to as top-down control and is associated with learning(Proctor & Vu, 2010).

The prefrontal cortex is vulnerable to stress. Diamond described the Prefrontal Cortex as "the canary in a coal mine", as it is the least robust brain region when faced with strain. The literature indicates that emotional and physical stress factors are likely to complicate learning of AAC and BCI systems in general, and BCI with spatial tasks. Mental states and processes that might alter a person's ability to attain and maintain voluntary control of EEG signals are (i)concentration/[lack of]focus, (ii)other thoughts/control of thoughts, (iii)frustration, (iv)other mental/emotional states (e.g., depression), (v)relaxation, (vi)fatigue, (vii) distractions/interruptions, (viii)motivation/desire, (ix)intentions(Curran & Stokes, 2003). Hence, a person experiencing relatively ordinary stress factors, however temporary, may also experience learning difficulties during this period, and difficulties performing complex, cognitive tasks.

However, as soon as skills are overlearned and automated, the user no longer depends on the top-down control of focus and discipline in the prefrontal cortex. Allegedly, seasoned BCI users have compared BCI use to "riding a bike", and the strategies applied to initial learning are made redundant by experience (Curran & Stokes, 2003). Overlearned skills rely on posterior cortices, and their activity is referred to as bottom-up activity; a phrase describing the directionality of signals moving from the lower astern regions up towards areas of cognition. (VanLehn, 1996). Hence, although the Executive functions are imperative for learning BCI skills, they are superseded by automation once skill processes are overlearned(Proctor & Vu, 2010). This may suggest that we must ensure accessible task design for initial learning processes to help users overcome the potential learning difficulties associated with strain and confusion, until they have succeeded in overlearning and automating the cognitive command skill.

Operation protocols and Task Design

As for BCI interaction, the user's task goal is to maximize the correlation between intent and signal to maximize the signal-to-noise ratio. (McFarland & Wolpaw, 2011). Operational competence is the competence to interact with an AAC interface, as described by Janice

Light(Light, 1989). This involves the physical ability to operate a device and contextual understanding of the cognitive efforts. (Cook & Polgar, 2015).

In Instruction theory, tasks can be divided into Skill-based and Problem-solving tasks (VanLehn, 1996). Skill-based tasks can be automated, and generally have a high level of performance. Skill task components include a) Means of work b) Input information c) Required actions d) Goal of the task. Given proper instructions, the task is skill-based. If the user is uncertain of how to solve a given task it is problem-based and need to be resolved before it can be trained. Bedny and Karwowski performed a functional analysis of attention in a review of SSAT (Systemic-Structural Activity Theory). They concluded that mental effort required to perform a task increase with difficulty. Then how do we alleviate the difficulty in such a complex content matter? They claim that users don't experience a tasks complexity, but its difficulty. We avert difficulty and task load if we provide a (i) Strategy for Task Performance, (ii) with Specific instructions, (iii) for Clearly defined Task Components.

In a review of Human Factors, Proctor & Vu presents Stimulus-Response compatibility (S-R compatibility) as a task strategy. S-R compatibility is a dimension of relevance between Stimulus (the objective), and its Task Response. Lack of S-R compatibility is referred to as the Simon Effect(Proctor & Vu, 2010), and requires more mental processing, a heavier cognitive load and a longer reaction time.

Task adaptation

S-R compatibility should be applicable as a strategy for task design to improve BCI interactions. One example is a Russian experiment with visuospatial tasks(Bobrov et al., 2011). The participants task was to visualize either an icon of a house or of a car (Response). Their que to perform the task was conditioned by a green triangle shifting position within a geometric figure. Was this Stimulus relevant to imagine a house or a car? For a visuospatial task to be compatible, the Imagined icon should have a significant relevance to the task objective. (Proctor & Vu, 2010). The task appears not to have a contextual objective, and the prompt appears to be unrelated to the icons. This could indicate S-R incompatibility and provoke a Simon Effect. The authors in Bobrovs report on visuospatial tasks addressed, in their Discussion, that they expect improved EEG Performance from an improved training procedure. Task strategy might be one of the improvements the authors had in mind.

Stepwise skill acquisition as a design approach

I have focused upon cognitive tasks for Spontaneous EEG with visual imagery for an initial learning phase. To minimize intrinsic load, it appears important to help user attain good first experiences with mastery and motivate them to the path of acquiring skills until tasks are automated. With a stepwise approach to skill acquisition we may introduce introduction level task with feedback and opportunities for self-assessment(Lotte et al., 2013). The issue in is that a novice practitioner may struggle to produce reliable training data, while having access to reliable training data is a critical step in the workflow. Poor quality in the training data may cause the provided feedback to appear arbitrary. Ideally, we could attain high-quality training data without expending the users' efforts and self-regulation pool.

One way to achieve this is to use bottom-up (dependent) signal for training a top-down independent signals for test data. The literature suggests that visual perception and visual imagery emit near identical signal patterns. If this is the case, and a signal acquired with one can be used interchangeably with a task emitting producing the other, we may be able to collect training data with dependent signals, perception, and utilize them in dependent signal imagery tasks with feedback during training.

Define a sufficient similarity between training data and test data; that we may collect training data with one method and test data with another.

Will a stored signal with bottom-up directionality (visual perception) classify correctly in an online session with a top-down (visual imagery) command? If not identical, then at least so similar that the user is not handed an intolerable disadvantage? We are looking at a situation where there is no discernible difference in classification between outcomes achieved with training data acquired by dissimilar methods, or a situation where outcomes are only similar when training data and test data are similar.

Research metrics

In this section I will present research strategies, methods for experiment procedures and analysis from relevant papers, and present research question and hypothesises. In general, this will revolve around methods for how to measure the amount of BCI activity, and how to analyse its correlation to users' intent.

UX Methods with Cognitive Load

Cognitive Load is a framework developed for Instructional Design Theory, describing the different forms of stress, as well as contexts for measurement. Cognitive Load can be divided into Intrinsic cognitive load, germane cognitive load and extraneous cognitive load, representing the internal load brought by the user, the inherent stress expended by a task and the added load expended by insufficient instruction or clumsy design, respectively. It may be measured as biological responses or reported as a subjective experience by users.

In a review of UX evaluation metrics, Kübler et al divided metrics into Efficiency, Effectiveness and Satisfaction. (Andrea Kübler et al., 2014).They suggested to measure User Experience Efficiency with NASA Workload. Efficiency is a counterintuitive term in this context. NASA TLX measure Workload, as an experienced negative effect.

NASA TLX

NASA developed a system of workload measurement called NASA TLX, for use in pilot training. The NASA TLX form is a survey that collect information about the participants experience of tasks. A total of six factors, with ordinal values between 1 and 20, giving a total value Workload between 21 and 120. The factors are; Mental load, Physical load, Temporal load, Effort, Performance and Frustration.

Those studies of BCI technologies that formally incorporate user feedback with BCI, that are reviewed in this paper all applied NASA TLX to gather user experience data. The method is based on self-evaluation of tasks. The form have since been applied to studies in HCI and BCI (Anderson et al., 2011; Chavarriaga, Fried-Oken, Kleih, Lotte, & Scherer, 2017; McKinley, McIntire, Schmidt, Repperger, & Caldwell, 2011) Initially the user was asked to weight each option against each other, however new research areas have applied the total Workload in quantitative studies, with a simple ANOVA analysis instead. This unweighted version is referred to as Raw NASA TLX (Moroney, Biers, Eggemeier, & Mitchell, 1992).

Some work with so called Raw NASA TLX, where the factors are summed up, not weighted. More recent studies suggest measuring and comparing each factor independently, rather than as the sum. These factors have also been applied to analyse causality between the workload factors and Cognitive Load theory.(Galy, Paxion, & Berthelon, 2018). Looking at Classical Test Theory, this approach appears equally valid if the dimensions compared are parallel. In line with CTT, we assume that the total Score of NASA Factors = a hypothetical true Score with noise. However, the factors are not equally relevant, and in cases where it is difficult to obtain significant test results of interest in a hypothesis we may consider performing a more extensive factor analysis. We can remove factors that are similar across test groups. If we remove the same factor dimensions and are clear about what information we then obtain from the net total, we may have obtained significant test result with available data.

Utility metrics

As mentioned, Kübler et al divided metrics into Efficiency, Effectiveness and Satisfaction. (Andrea Kübler et al., 2014).They suggested to measure utility Efficiency with ITR, and Effectiveness with percentage correct responses. Number of classes is the complexity added to the interaction. Accuracy, or effectiveness, is a performance metric that evaluate whether the submitted bitrate was an expression of user control. Accuracy decline is negatively correlated with number of available classes (trained commands). Target detection time is the amount of time it takes to issue a command or make a choice between options. In a review of BCI interfaces, Ramadan & Vasilakos listed utilitarian metrics as Information Transfer Rate with bitrate and (I-R); Target detection Accuracy, Number of classes and Target detection time.(Ramadan & Vasilakos, 2017). The definition of bitrate varies.

Bitrate definitions and applications

According to Ramadan and Vasilikos, bitrate is the most common measure for BCI systems (Ramadan & Vasilakos, 2017). According to them, bitrate is measured as the number of times a system classify commands per attempt/trial/minute. The details of what that entails vary. A Genevan research group(Julian Kronegg, Voloshynovskiy, & Pun, 2005; Julian Kronegg et al., 2005) have published reviews on bitrate and Information Rate (I-R) models from established BCI research institutions in Graz, Wadsworth and Albany(Curran & Stokes, 2003; McFarland & Wolpaw, 2011; Nicolas-Alonso & Gomez-Gil, 2012). Bitrate $_{\text{Farwell \& Donchin}}$ is defined as a perfect classifier without errors; the amount of classified successes is then equal to the amount of emitted commands. Bitrate$_{\text{Wolpaw}}$ defines bitrate in relation to Information Rate; as any classified command that is distinguished from baseline signal with rejection inherent in I-R Accuracy. The Genevans argue for a third option, bitrate$_{\text{Nykopp}}$, which only ever consider accuracy.

The definitions have different areas of application. Bitrate $_{\text{Farwell \& Donchin}}$ appear to be a theoretical idol of complete overlap between classifier and commands that is, in my estimation, not necessarily intended for practical use. Accuracy is a separate term from bitrate in the Wolpaw Information Rate Model. Kronegg et al presents arguments that bitrate$_{\text{Wolpaw}}$ loses accuracy and validity for higher number of classes. Still, there is a discrepancy when Kronegg emphasize this condition to advice against use of bitrate$_{\text{Wolpaw}}$ in circumstances where Wolpaws reference to I-R already takes accuracy into account. Bitrate$_{\text{Nykopp}}$ include a model for SNR optimization for offline analysis and is hence considered more accurate.

In a more recent paper, the Genevan group differentiate between Single-trial protocols and repeated attempts (Julien Kronegg, Alecu, & Pun, 2003). They propose the term Average Trial Protocol to describe that the user may emit the same mental command repeatedly over time. The term Average Trial Protocol may be appropriate as an assumption for an analysis protocol of online classifications. Rather than assume bitrate$_{\text{Wolpaw}}$ to be a continuous value variable, they argue that the number of registered classifications in bitrate$_{\text{Wolpaw}}$ is limited to a finite amount dictated by an optimal bitrate, where the bitrate is a Probability P of Classification rate V and leaves a possibility for alternative classifications in a context with multiclass options.

Research question

Can spontaneous EEG recordings, associated with either perceived or imagined stimuli, be used interchangeably without significant dissimilarity in results?

Hypothesis 1

$H1_{0:}$: Spontaneous EEG commands with visual imagery perform as Efficiently and Accurately, whether they are tested with recordings from perceived visual stimuli or conducted with recordings with imagined visual stimuli.

$H1_1$: Spontaneous EEG commands with visual imagery perform more Efficiently and Accurately when they are tested with recordings from visual imagery than if conducted with recordings from visual perception.

Hypothesis 2

$H2_{0:}$: Spontaneous EEG commands performed with recordings from perceived visual stimuli and imagined visual stimuli, demand the same amount of cognitive load; measured as perceived workload .

$H2_{1:}$: Spontaneous EEG commands performed with recordings from perceived visual stimuli demand a higher cognitive load, measured as perceived workload, than if they were conducted with recordings from imagined visual stimuli.

Hypothesis 3

Is there a correlation between experienced cognitive load during training, and signal Efficiency & Accuracy?

$H3A_0$: There is no correlation between performance variables Score & $bitrate_{Wolpaw}$, and cognitive load measured as workload.

$H3A_1$: There is a positive correlation between cognitive load measured as workload and high performance, measured as $bitrate_{Wolpaw}$ and Score.

$H3B_0$: Data training with visual imagery and visual perception are equally demanding in terms of workload.

$H3B_1$: Data training with visual imagery is more demanding to do than visual perception in terms of workload.

METHODS

Experimental study adapted procedure comparing training data, across the conditions; imagined and perceived visual imagery, with assumed high S&R compatibility. Measured workload with workload factors, bitrate, accuracy as a score and correlation between inputs and objective. The following section presents the various methods, their fit to the research question and hypothesis. It will cover what is being examined, how and with what tools. Then follows a presentation of Analysis performed and how the data will be stored for future use. The last paragraph present Ethical considerations for this project.

Participants

In this subsection I will present the population and the population sample. 13 participants were recruited into the project. 5 female, 5 male and 3 abstained from reporting gender. Aged between 22 and 51. The participants were recruited based on convenience sampling. We demanded only that the users reported to possess normal eyesight ability. One prospective volunteer was dissuaded from applying to the project due to legal blindness. Two prospective volunteers were dissuaded as they did not have access to glasses or lenses.

Ethical considerations

The project was reported to, and received permission from, NSD to conduct experiments and information handling in the way the project was finally conducted. Contact information and names were stored in two places; UiO servers and a folder placed in a locked closet. Notes were kept in a separate folder along with NASA tlx forms. Other collected information was stored in a server on a personal laptop with backups on cloud servers belonging to OsloMet. Participants were informed of the experiments content, the projects purpose and their right to withdraw consent ahead of signing up. Experiments were designed as an HCI experiment, with the least necessary amount of impact on participants, and least time consuming as practicable. This has also been mentioned in Tasks were designed to not inflict stress or task load on users. The perceived task load for each participant can be found in the chapter for Results and Findings. Prospective participants would initially sign up for the project using Nettskjema. Nettskjema is a cloud survey tool hosted by The University of Oslo, and apply to the University standards for data storage and security. Later, when recruitment was more direct, participants also filles a paper formed in the beginning of their first session.

Tools / Measures

EEG equipment

Consumer grade EEG system Emotiv Epoc, a headset with 16 sensors. The sensor placements are distributed over the frontal, side and back regions of the skull; AF3, AF4, F7, F3, F4, F8, FC5, FC6, T7, T8, P3(ref), P4(ref), P7, P8, O1 and O2. Frequency range from 0.16 to 43 Hz. Saline soaked pads, no gel required. Comes with proprietary Bluetooth connection and software. Emotiv Epoc EEG system have previously been assessed for research purposes and applied to experiments with spontaneous EEG, it is considered reasonably independent from noise produced by muscular artefacts. (Bobrov et al., 2011; McMahan, Parberry, & Parsons, 2015; Taylor & Schmidt, 2012).

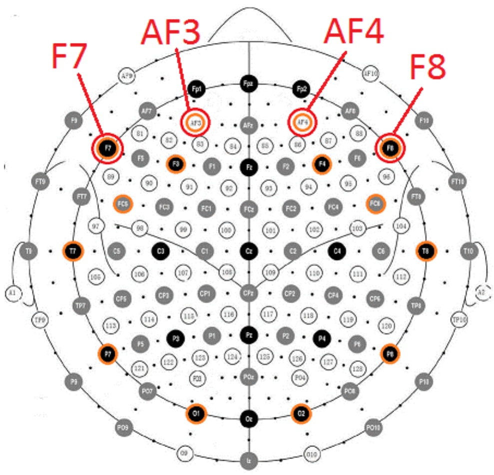

Figure 0.1 10-20 system("EEG > introduction Biomedical Signals Acquisition," 2005)

Software

Emotiv Software: Emotiv Xavier Control Panel

The software converts the data from analogue to digital signals, filters, classifies and generates an classifcation algorithm based on collected samples of training data. It also contains a GUI for sampling and classification.. Once the algorithms are generated, the software monitors online signal for trained actions.

Emotiv Software: Emotiv Xavier Emokey

EmoKey is also developed by Emotiv, for use with the Emotiv Control Panel. This software output a keypress in the event of detected, trained actions. In this project it is configured to relay the action Push, limited to when the action is detected for more than 0.2 microseconds. EmoKey emit Keypress Spacebar to the application in focus.

Flappybird: Experiment interface

The interface chosen is a customized smartphone game, FlappyBird, a side scrolling game with a single input, a Spacebar Key press event. The objective of the game is to avoid obstacles, and to achieve this the Participant both activate and relieve a single command, to navigate above and below objects in their avatars path. Each time the avatar dodges an obstacle the player gains an additional point.

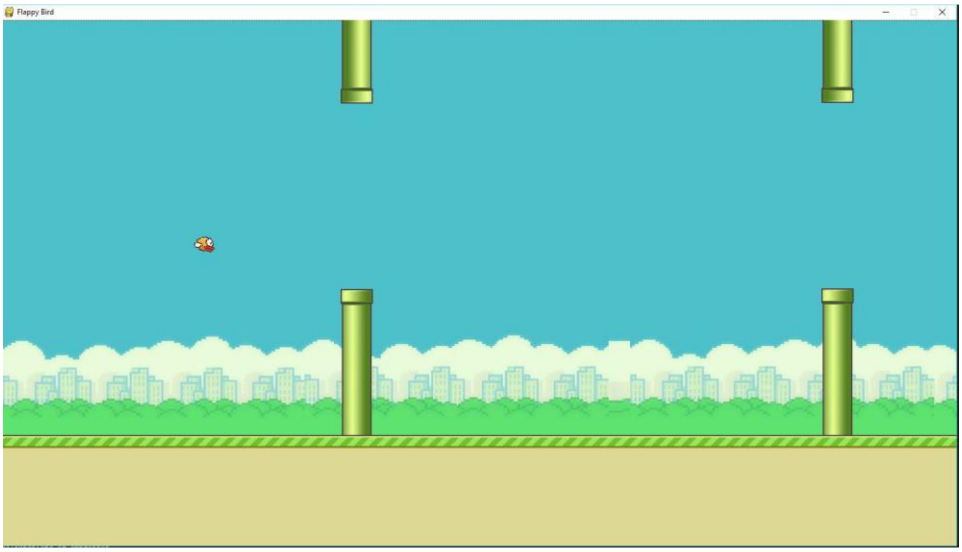

Figure 0.2 FlappyBird screenshot("stikling/FlapPyBird: A Flappy Bird Clone using python-pygame," n.d.)

All commands in a gameplay is counted and divided by game time, giving a continuous variable Bitrate/min. Accuracy is calculated from the achievement of intent, or game objective. The objective is to navigate the sprite between obstacles, and the success/fail is analysed from control limits. Lower control limit is 220 and upper control limit is 550 pixels. Each input adjusts the position of a sprite 33 pixels upwards. In the absence of input, the sprite adjusts back to its position, then decline 30 pixels per blit (for about 30 FPS / 30 blit per second) within the confines of the frame, see Figure 0.3 for the movement relative to timeline of 0-1 second. The first leg, before the first obstacle, lasts for 11 seconds. Each leg after lasts for 6 seconds. Gameplay ends after 21 obstacles, where 20 legs are analysed for performance. FlappyBird generate and save lines of text to a .txt file for offline performance analysis. The line of text includes the Sprites y-axis position, a timestamp and a participant identifier. I have adapted the prototype for HCI testing. The customizations include even spaces between obstacles, and a set duration of each gameplay, and also removed the function that usually aborts gameplay. The applied version is stored in a separate fork repository on GitHub. (“stikling/FlapPyBird: A Flappy Bird Clone using python-pygame,” n.d.)

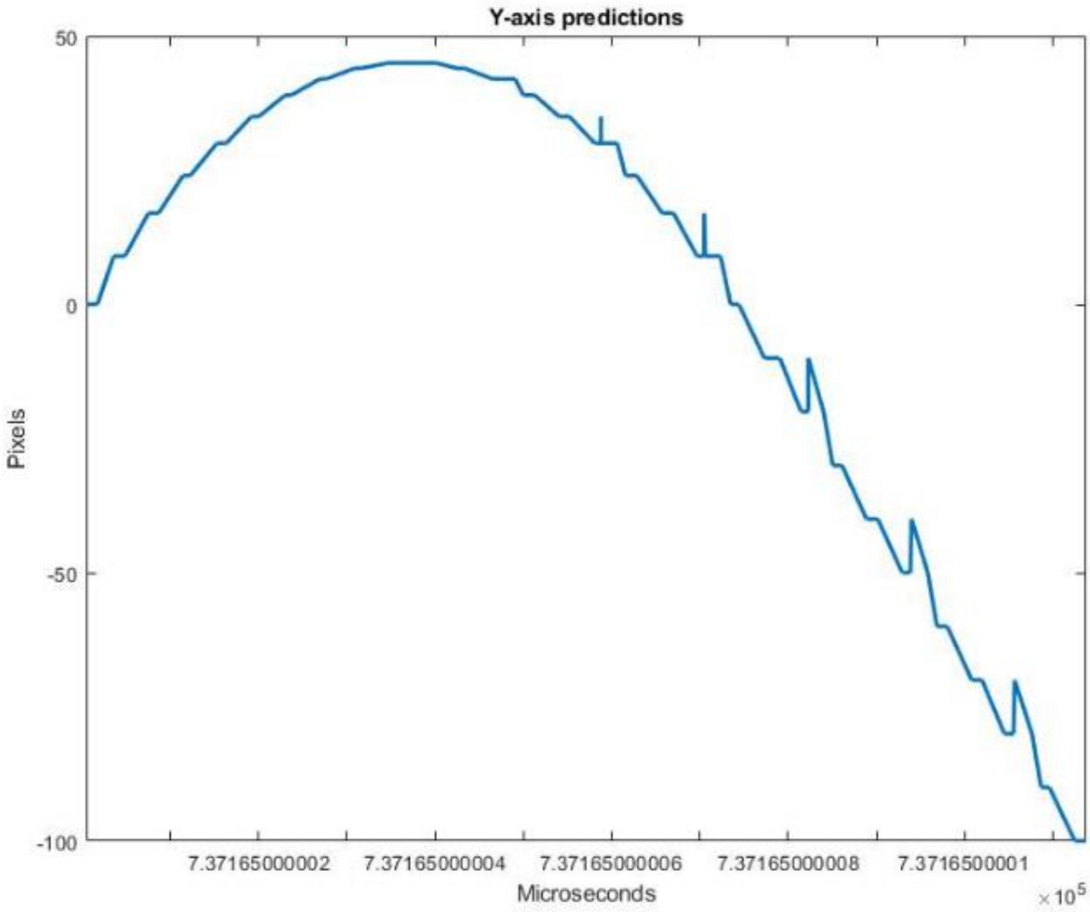

Figure 0.3: Y-axis position post command, 1 second timeline

Location

The test facility is a selection of similar rooms, booked special for each session. It is a rectangular facility with white plaster walls, a window facing a quiet street on one end, and a door in a semi-transparent wall on the other end. The window has white curtains. Dark shutters on the facade with an available control unit. Curtains and shutters are drawn shut during the experiment. The facility has an even temperature and air quality. The facility is not soundproofed. Noise level vary. The setup includes a rectangular table pushed up against one side of the room. Facing the wall, an ergonomic mid-range office chair for the participant. On the table, an open laptop facing the researcher, a BenQ FP222WH 22inch

LCD monitor facing the participants seating. For the participant there is also an additional keyboard, an Emotiv headset and a folder. On the opposite wall, directly behind the participant, an A3 poster of an upward pointing arrow.

Experiment Procedure

This subsection concerns the treatment applied to the research objects, in the form of introduction to an EEG exercise, rehearsal and test of performance. In this study we experiment with adaption of training data, by testing the utilitarian performance of training data with minor changes in the task protocol. The main task is to perform cognitive commands with visuospatial imagery. The protocol includes A) learning and memorizing a visuospatial cognitive command B) training the command with visual imagery without feedback, C) Test the visual imagery skill with a BCI interface and feedback. The BCi interface is a single class interface, with only one input using a single cognitive command, that they must both activate and relieve to achieve the task objective.

Testing procedure during a Session

Each session is appointed with which test condition will apply, and which training data will be collected during the session. It is either Method 1; Visual Perception. Or Method 2; visual imagery. The first activity is to watch and memorize an icon depicting an upwards-pointing arrow. During memorization the participant sits approximately 1 meter from a wall, facing a A3 hard copy of the icon for 60 seconds. The second activity is to visualize the memorized icon. The Participant notify the operator when she is ready to begin the cognitive task. The operator abort the visualizaton each 8 seconds so that the Participant get to pause a moment to refocus. This visualization task is repeated 5 times. The procedure with Memorization training and Visualization training is performed three times. Depending on the predetermined IV for the given session the operator ccollect command data, that the system uses to classify input signals. We record 5*3 sets of training data á 8 seconds, and a total of 2 minutes. These recordings are deleted immediately after the session. The Third Task is to play FlappyBird one time using visuospatial imagery a bird. Once the participant has completed the game they often share spontaneous feedback. I distribute a NASA tlx form, and the participants assess each task separately.

Statistical analysis

Measured workload with workload factors, bitrate, accuracy as a score and correlation between inputs and objective.

H1

In this hypothesis we compare Methods 1, Visual Perception and Method 2 Visual Imagery using Efficiency as $\text{bitrate}_{\text{Wolpaw}}$, and Accuracy as Score.

For $\text{bitrate}_{\text{Wolpaw}}$ we compare number of classified commands as continuous variables with a paired t-test. For Score I compare the Bernoulli value Successes per Session and perform a paired t-test using Score as a continuous variable.

H2

This is a test of similarity of reported Workload across conditions. As subscales in NASA TLX are parallel and unidimensional(DeVellis, 2006), I count score for each item and add them up as a total workload, then test with a paired t-test of means across conditions for the gameplay Task. I also perform a paired t-test of Workload reported in the activities related to acquisition of training data for each method to compare the workload associated with Visual Perception and workload associated with visual imagery.

H3

A test of correlation between Effort and Effect. It is a linear regression analysis of the correlation between Score and Workload, where the Score of the gameplay task is tested against the workload of the associated data acquisition task for the given Method.

RESULTS AND FINDINGS

In this section I will present descriptive and inferential statistics for each hypothesis, with the values collected in Experiments, as described in Methods.

RQ

Can spontaneous EEG recordings, associated with either perceived or imagined visual stimuli, be used interchangeably without significant dissimilarity in results?

Utility H1

$H1_0$: Spontaneous EEG commands with visual imagery perform as Efficiently and Accurately, whether they are tested with recordings from perceived visual stimuli or conducted with recordings with imagined visual stimuli.

$H1_1$: Spontaneous EEG commands with visual imagery perform more Efficiently and Accurately when they are tested with recordings from visual imagery than if conducted with recordings from visual perception.

The performance was measured as Efficacy and Accuracy, represented by $\text{bitrate}_{\text{Wolpaw}}$ and Score. Mean overall $\text{bitrate}_{\text{Wolpaw}}$ per attempt is 146.562 (SD = 125.117). With visual perception-data the user achieves an average of 88.625 classifications per session, and with

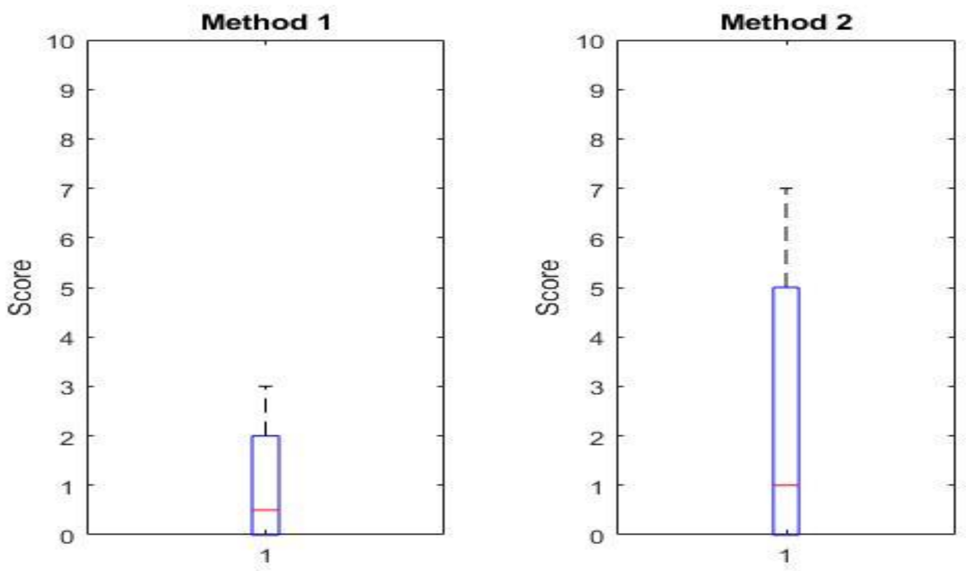

Figure 0.1 Figure showing the means of Scores with both visual perception (Method 1) and visual imagery (Method 2)

visual imagery-data an average of 204.50 bit. Total mean Accuracy, as Score (n=8) averaged 1.6 (SD = 2.2). Participants scored an average 1.6 successes of 20 attempts using visual perception data, which amounts to 5 percent of their opportunities. Using visual imagery data, participants scored a mean of 2.3 successes per session, which amounts to 11.8 percent of 20 attempts.

With the current variations in $\text{bitrate}_{\text{Wolpaw}}$ (± 125.117) of the sample data (n = 8), there was no significant difference detected between the achieved $\text{bitrate}_{\text{Wolpaw}}$ with visual perception data (88.625) and $\text{bitrate}_{\text{Wolpaw}}$ with visual imagery data(204.50) in a two sample t-test (p = 0.085). The difference in the metric Score, there was no statistically significant difference in Accuracy between visual perception (1 out of 20) and visual imagery (2.37 out of 20) with n = 8 (two sample t-test, p = 0.2322).

User experience H2

In this hypothesis we tested the overall cognitive demand of performing the tasks associated with $H2_{0:}$: Spontaneous EEG commands performed with recordings from perceived visual stimuli and imagined visual stimuli, demand the same amount of cognitive load; measured as perceived workload. $H2_{1:}$: Spontaneous EEG commands performed with recordings from perceived visual stimuli demand a higher cognitive load, measured as perceived workload, than if they were conducted with recordings from imagined visual stimuli.

For each test session the participants answered a NASA TLX form for each task of the experiment. The total sum of factors for each form in the NASA TLX is Workload. The workloads of the associated signal acquisition task were compared across methods with a paired two-tailed t-test. The participants reported average perceived Load of 33.12. The variation of reported Load was sd = 31.93 (n = 12), with no discernible difference between training data for Method 1 and training data for Method 2 (two sample t-test, p = 0,5). There is no significant difference in each user's experience of workload based on the performed tests, even if the experience between users varied.

Relationship between Utility and User experience H3

This parameter is a test of a possible correlation between the measured utility and the self reported user experience. Is there a correlation between experienced cognitive load during training, and signal Efficiency & Accuracy?

$H3A_0$: There is no correlation between performance variables Score & $bitrate_{Wolpaw}$, and cognitive load measured as workload.

$H3A_1$: There is a positive correlation between cognitive load measured as workload and high performance, measured as $bitrate_{Wolpaw}$ and Score.

$H3B_0$: Data training with visual imagery and visual perception are equally demanding in terms of workload.

$H3B_1$: Data training with visual imagery is more demanding to do than visual perception in terms of workload.

$H3A_0$ was tested with Score and workload with linear regression analysis with Method 1 and Method 2 separately. R squared for Method 1 was 0.167, while R squared for Method 2 was 0.305.

	R^2	p
Method 1	0.167	0.315
Method 2	0.305	0.156

Figure 0.2 R2 and p values for correlation between score and workload.

R squared for Method 1 was 0.167 (p = 0.315), while R squared for Method 2 was 0.305 (p = 0.156). Higher performance with Spontaneous EEG is not significantly correlated to task load, represented by Workload. There is a difference between the two, where Method 2 have a somewhat higher correlation. Method 2 might indicate a negative correlation, while there is no such correlation with Method 1.

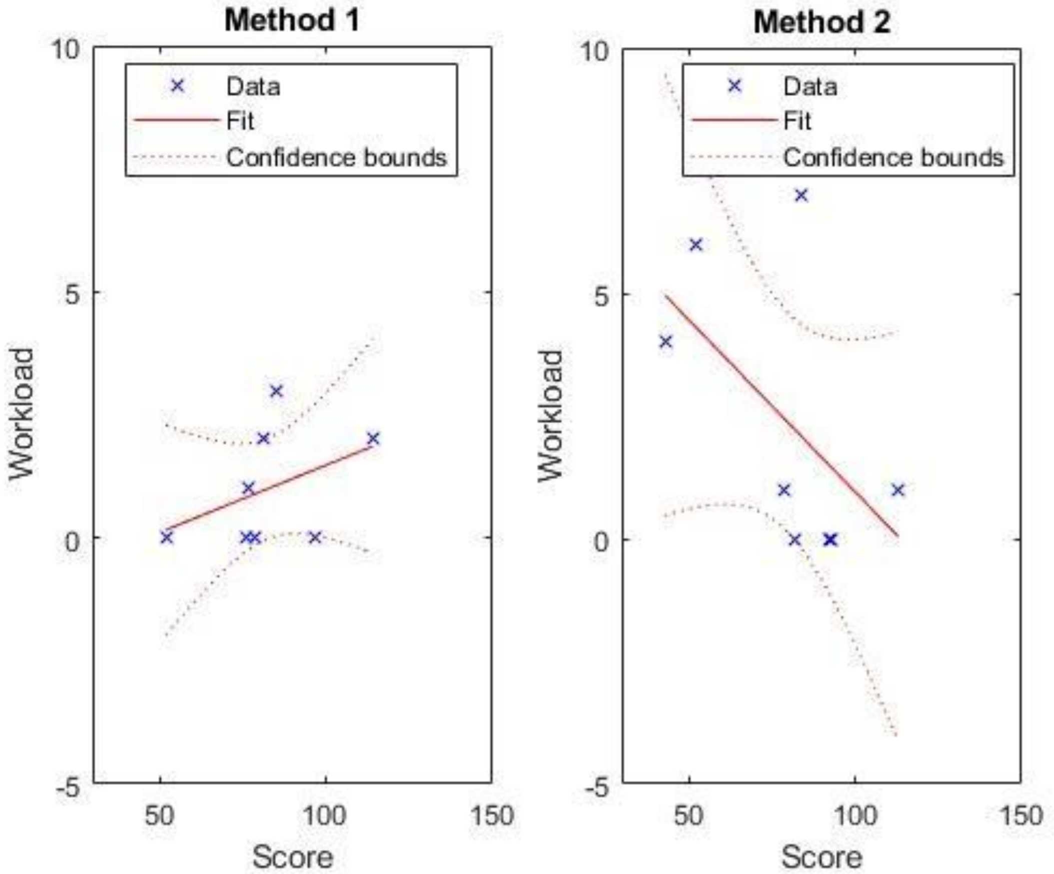

Figure 0.3 Linear regression of Workload and Score.

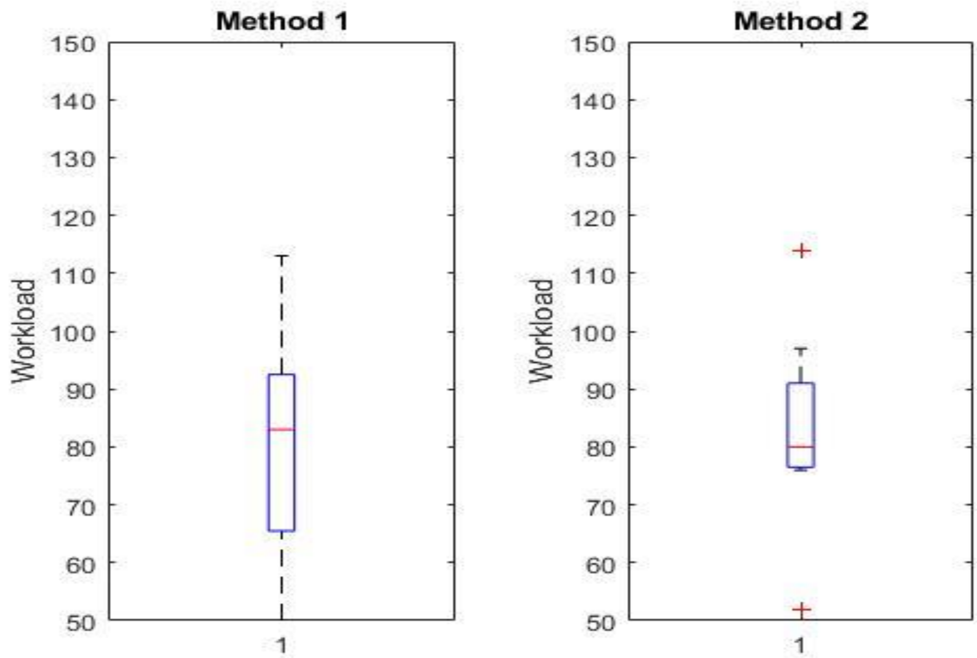

Figure 0.4 Boxplots of Workload for conditon 1 and 2, h2

DISCUSSION

In this study I have assessed the effectiveness and user experience of spontaneous EEG, by testing training data with different cognitive tasks. I have looked at skills required for end-users to operate a BCI communication system, attempted to define optimal cognitive commands for task design with Stimulus-Response compatibility and discussed how hardware and system design can support accessibility. In the Discussion we will look more closely at the relationship between the performed study and the results, relating these results to previous literature. I will present potential for identifying new challenges for accessibility as well as suggest improvements to the experiment procedure and its execution.

Findings

We failed to find significant evidence for a higher task performance with perceived versus visualised cognitive command inputs. The results are inconclusive, and we accept the null hypothesis. The population was too small to reach a conclusive result. There were notable individual differences in performance that were far more significant than the variance seen between input methods. How may we address these individual differences? However, this experiment was performed with standard equipment, and the generic Emotiv EPOC algorithm for identifying cognitive commands was not adapted specifically to the challenge; specific adaptations of the algorithm to recognize signal patterns with an opposite directionality may contribute further to eliminate differences in task performance.

We failed to find significant evidence that the perceived workload is affected by the type of its training method (perceived or visualised), and we accept the null hypothesis. The population was too small compared to the between-user variance. A test of power revealed it would require 900 participants to achieve a sufficient power. However, the variance may also be explained by a confounding factor in the workload subscales. Overall workload may be thwarted by inclusion of irrelevant factors that may are more subject-dependent than task-dependent. One of the subscales, Effort, was explained to be the one most closely correlated and indicative with Workload at any given time(Hart & Staveland, 1988).

Testing with Effort. Two tailed t-test with Effort performing the task with both methods, n = 8, sd = 3.52, p = 0.4893. Conclusively similar Effort exercised for both methods during the game task. According to the literature there should be a greater difference between these tasks in term of WL. In fact, when looking at Effort, there is a significant difference. Effort is, allegedly, representative for Intrinsic, extraneous loads, as well as the mental resources expended by the task (Galy et al., 2018) and is the single factor most correlated to overall workload according to NASA TLX developers (Hart & Staveland, 1988).

There was no significant relationship between perceived workload and task performance, for either the perceived or visualised conditions. Small population. As in H2 we may not have used the most relevant factors from the NASA TLX. Temporal Demand appear relatively unrelated to Workload, while Difficulty is a relevant factor when comparing tasks and protocols. We may use factor analysis to compare with Effort instead. As the workload was inconclusive, it is reasonable to extend the suggested alternative analysis to this hypothesis H3. Therefore I have performed a few more tests of our data to explore how the Efficiency and Accuracy of the game task induced positive and negative emotion. In that regard I ran linear regression on bitrate' affect on Frustration and Perceived Performance, the score' effect on Frustration and Perceived Performance

Frustration vs bitrate: Number of observations; 16, n = 8, R^2 = 0.209, p = 0.0748, Frustration vs. score: Number of observations; 16, n = 8, R^2 = 0.0834, p = 0.2,

Perceived Performance and Score: Number of observations; 16, n = 8, R^2 = 0.38, p = 0.011,Perceived Performance and Bitrate: Number of observations; 16, n = 8, R^2 = 0.204 p = 0.0787.

I have also tested the correlation between Effort and Frustration, to see if the Expendedefforts themselves generated negative emotion. I tested this correlation across all sessions and all tasks where we collected NASA tlx data. Total number of observations; 76, R^2 = 0.608, p = 1.05e-16.

From the developers of the NASA TLX system, Frustration was explained as; "in a relatively less ambiguous way, relates task requirements, exerted effort, and success or failure."(Hart & Staveland, 1988). Correlation between frustration and bitrate; Frustration is negatively correlated with bitrate, more so than a failure to score points. Performance is indicative of situation awareness and represents the users' self-assessment. Sadly, none of the users appear to take much credit for their achievements. Only the user with the very highest bitrate reported a high performance, and the participant with the highest score reported only a moderate performance. Only the second runner up on the high score reported a high degree of performance. The users appear to experience a high degree of performance when they produce classifications and achieve a high bitrate.

A higher workload doesn't automatically lead to a higher alpha wave, which is usually indicative of stress (Insert reference). This result is very useful, as we may not dismiss poor results based on a biologically conditioned performance boundary in Alpha Waves related to stress.

Complementary observations

Participant activity 4, Relaxation excercise	3 participants expressed discomfort with this exercise. As a response to the description of the task and its purpose the participants, all but one, spontaneously shared details about their day, and stress factors in their lives.
Participant activity 1, memorization excercie	One participant expressed annoyance, two participants also expressed that they spent the full minute in each repetition memorizing detail such as angles and corners. One of them expected a task where *he should describe or draw the arrow, or in some other way show off their memory verbally.
Participant activity 2, Visualization exercise	Some expressed that this was a task it was difficult to conceptualize. Some took longer between each attempt before they were ready to clear their mind and focus on the visualizations. Every participant seemed to search their mind for the internal image before they gave all-clear to begin the recording process. One expressed *he doesn't know how to visualize anything, never has visualized anything, and would rather use other coping mechanisms.

Participant activity 3, Gameplay	Participants appear to use some attempts while they are playing, rather than think of each epoch as a goal that they have ambitions to cross successfully. They would make attempts, shake off the strain of the attempt, wait for the next epoch they were decidedly ready for and make their next attempt. Some struggled to make commands at all (floor effect), while yet others struggled to turn the signal off once they got started (ceiling effect) Have only seen ceiling effect with Condition 2.
Technical Recording process	Noticed calculated skill level in Emotiv control panel did not have a linear development for each recording. Also noticed that some participants had active feedback while preparing imagination task, but upon informing me of being ready, their focus dropped and feedback did not necessarily recover within the recording time. I quickly made a habit of delaying timed intervals for recording a couple of seconds, and in many instances their focus picked back up nicely within the timeframe. I did not wait for active feedback, but recorded training data at timed intervals.

Table 0.1Summarized notes from testing

Positive Psychology

“Adapting the task difficulty to users skill improved one dimension of flow state, cognitive control. People who faced a challenge better suited to their skill felt more in control.”(Mladenović, Frey, Bonnet-Save, Mattout, & Lotte, 2017) It is interesting to pull a wide relation to Eudemonics. “Eudemonic view equates happiness with the human ability to pursue complex goals which are meaningful to the individual and society”. In this sense, difficulty is not necessarily a negative experience, but it depends on an intrinsic motivation. Flow Experience Theory have defined 7 principles; Knowing what to do, Knowing how to do it, Knowing how well you are doing, Knowing where to go, High perceived challenges, High perceived skills and Freedom from distractions. The possible rewards from fulfilling these 7 principles lies in the definition of Flow Experience; “the mental state of operation in which a person performing an activity is fully immersed in a feeling of energized focus, full involvement, and enjoyment in the process of the activity”.

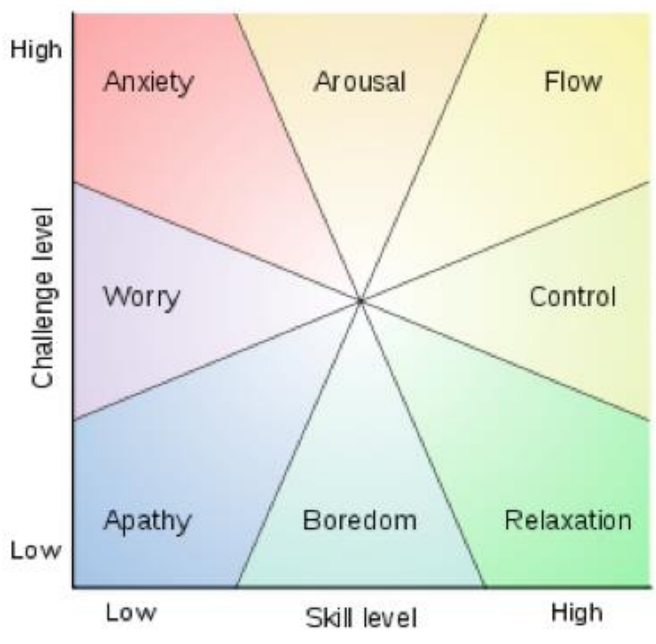

Figure 0.1 Flow model (Csikszentmihalyi, 1997

)Performance Achieving bitrate indicates that the task itself is understood. A failure to produce classifiable signals is an indication that you have failed to understand the task and indicate high difficulty with low skills, which is positively anxiety inducing, according to the Flow model. As we can see from , Mikhel Csikszentmihalyi have presented a model for the interaction between Skill level and Challenge level. Where Lack of skill produce negative emotion, and lack of challenge produce indifference in the user. A perceived high skill level always produces positive emotion regardless of the challenge level, the challenge level is experienced quite differently based on the skills. A high challenge level produce both

Flow and a sense of Arousal with high levels of relevant skill but is only anxiety inducing if the user lack the reasonable skills to meet the challenge. (Csikszentmihalyi, 1997)

Skill Acquisition

The Difficulty of a task is, as mentioned earlier, a product of Task Components and the specificity of given Instructions. (Bedny & Karwowski, 2011) This can be understood as whether the instructions are specific enough to cover the underlying principles of components. It can also be understood as a warning of overexplaining a task, so that the student miss opportunities to self-explain and internalize learning. Lotte et al performed heuristic evaluations of BCI research procedures, based on Instructional theory (Lotte et al., 2013). Lotte stated that training procedures for BCI were inconsistent and lacked conformity with best practise principles. For instance, he explains that the users need feedback to correct their performance. This is supported by vanLehn (VanLehn, 1996). However, Lotte don't suggest how the cognitive process may be monitored upon or mentored, which also makes it difficult to imagine how students may self-monitor or self-explain their process.

One of the challenges of introducing users to Spontaneous EEG with independent signals, is the lack of feedback related to the users' performance. The opportunity to use training data from visual perception and visual imagery interchangeably, is the opportunity to gather dependent signals which require low effort for training data. It would grant a system or procedure designer relatively wide opportunities for Skill training programme with immediate feedback at the first challenging task with visual imagery.

Training procedures

While some central voices in BCI research emphasize Training procedures as one of the central fields of BCI development, other researchers have expressed a goal to abolish all need for instruction and training for users(Blankertz et al., 2006; Krauledat, Tangermann, Blankertz, & Müller, 2008). Krauledat, Tangermann, Blankerts og Müller stated that efficient detection algorithms and sophisticated hardware should provide sufficient capability for both system and for users. AAC stakeholders, however, have expressed a demand for instructions and training. Light and McNaughton encouraged researchers and rehabilitators to take a user-oriented perspective, rather than push their enthusiasm for potentials in specific technologies *"[…] there is a danger that intervention will be limited to the provision of a device, without providing appropriate training and supports to maximize communicative* *competence."* (Light & McNaughton, 2013) Beukelman et al requested better instructions for AAC equipment *"The need to provide targeted instruction and support for these individuals as well as those who rely on AAC remains an important future goal for the AAC field. [sic]"* (Beukelman, Fager, Ball, & Dietz, 2007).

Comparable BCI projects have invested far more time and resources on their participants. I have added up total time spend in different projects in Table 0.2. While I have invested up to 2 hours with each of the 16 participants, Bobrov et al invested approximately 30 hours per participant. This is the most appropriate comparison, as they also researched Spontaneous EEG with visual imagery, with the same type of EEG equipment. The actual interaction in my experiment lasted mere minutes altogether. How would a more generous training programme have affected the results? If we had spend hours adding training data, how well could the prediction model have performed? What classification rate would that have granted us? In contrast, see Figure 10.2, where netto time spent directly with the experiment take less than 10 minutes.

Table 0.2 Time spent with Participants in BCI studies

Participants	**Experiment time per participant**	**Approximate total experiment time (hours)**	**Number of sensors**	**Reference**
7	4 days (approx. 30 hours)	210	16 sensors on day 1-3/ 30 sensors on day 4	(Bobrov et al., 2011)
20	2 hours	40	fMRI -little to no assembly	(Ganis et al., 2004)
20	2 hours	40	19 sensors	(Shourie, Firoozabadi, & Badie, 2014)
12	5	60	121 sensors	(Winkler, Haufe, & Tangermann, 2011)

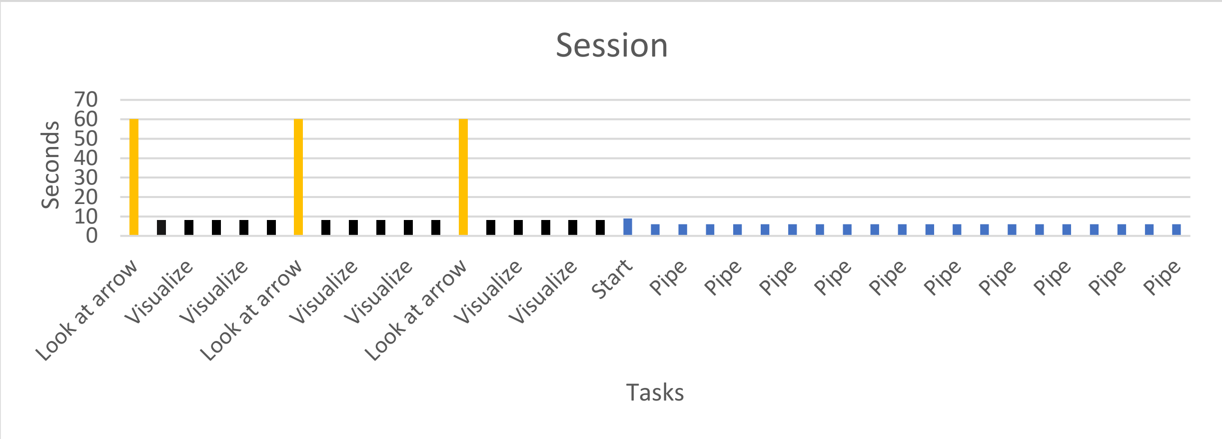

Figure 0.2 Total time spent with tasks each session

The results and the literature we have assessed suggest that difficulty and complexity is not universal accessible; unless the user experience a reasonably balanced relationship between their available personal resources, and the demands of sustained effort. It means that by eliminating challenges related to lack of understanding, we alleviate extraneous and germane load. We may boost Flow and an experience of eudemonic happiness, even as we are exerting great effort for a complex task. It means that an accessible interface induces an experience of understanding and mastery. This relates well with the third principle of Universal Design.

Limitations of the study

The greatest limitation to this study is the limited portion of the population we tested with both conditions. Although a total of 13 participants were tested, the distribution of methods across sessions failed to make all of them eligible for paired testing. A thorough factor analysis

Participants and recruitment

It might appear relevant to either measure cognitive abilities in participants to stratify the population or monitor neural activity during experiments. While this is a conservatively HCI centred study, a stratification like that may undermine the universal applicability in potential findings.

The total population in the project might not be low compared to other studies in the same field, however it is difficult to claim that the results are universally representative when it also fails to offer each participant a training programme, comparable to that of the compared studies. While other projects spent days, weeks and months to train participants to use cognitive commands, we are analysing results from the very first introduction to our interface.

Collecting training data

While the collection procedure was intended to be as similar as possible, the perception training task was self-paced, and the imagery training task was structured. The recordings were paced by the operator regardless of which method it applied. The timespan of each recording, 8 seconds, was conditioned by the control panel interface.

Future iterations of my experiment

H1

There were great individual differences that were far more significant than the variance between methods. How may we improve these individual differences? Qualitative analysis of notes or semi-structured interview for each session demands resources to analyse, but we can identify some sessions or performances that deserve extra attention. See to control chart to find the individual participants that represent outliers. Either for their complete lack of ability to perform classifiable commands, or for outliers of great performances.

H2

There are other methods of gathering UX data that may be more reliable with small populations. Collecting biometric data such as pupillary changes, sweat or accelerated heartrate are a few true and tested methods. Then again, rather than introduce a second device, that may be experienced as invasive and stressful, we might utilize data directly from the EEG headset. Offline analysis of signals provide insight into the users' emotional state during the experiment. The Emotiv EEG headset have an integrated detection suite for six signals that are monitored and available for online application. Those are Interest, Engagement, Stress, Relaxation, Excitement and Long-term excitement. There may not be a direct overlap between these factors and variables of interest, but a full licence version of Emotiv provide opportunities to customize functions with their System Development kit into the frequency range we are interested in.

H3

Researching the phenomenology of the BCI and the effect of task and performance on the users' inter-subjective experiences. Building on a correlation between the utilitarian results of H1 and the UX results in H2, we may Collecting raw data of the users' state of mind, their qualia, is a tell-tale way to correlate the users experience to their interactions in real time in offline analysis.

Would use Perceived difficulty rather than Time Pressure (Temporal demand). "[Temporal demand] ratings were only moderately correlated with [Overall Workload] ratings for individual experiments and categories of experiments". Task difficulty was dropped from the final version of NASA TLX because it was less statistically independent and provided less new information. However, of all three task related scales that were evaluated, Time pressure was the only one that remained. Simply exchanging it for Task difficulty would maintain the information and make for a more relevant factor in our data.

Identifying challenges

The design process of improved BCI prototypes require effective ways to administer and communicate contemporary challenges throughout a design process.

For an iterative development process, it would be helpful to apply qualitative methods to evaluate functionality. This may also involve qualitative analysis of quantitative data. Collecting qualitative data rarely take up resources, but the analysis is a resource demanding process. There are three models I would have preferred to incorporate effectively in the study, that could help identify which quality data to analyse, and also to convey and curate findings. Those are Control charts, Activity diagrams and Pareto diagram. Control charts with control limits may help identify outliers of specific interest. First of all, outlier sessions with exceptionally high or low performance may have inspired the participant to share experience during their sessions. Participant testimonies may include variables that have a high correlation, or even causality with their level of performance. An activity diagram could help us track potential events during test sessions, indicate ability gaps and recurring disruptions to workflows. A Pareto diagram may help direct efforts of the research and development communities towards knowledge-based development and direct resources to the most pressing challenges.

Collecting training data

Why not just use the integrated system in Emotiv? The integrated interaction platform in Emotiv Control Panel is designed around visuospatial tasks, with manipulation of a figure in a

3D space. There could be advantages from using the integrated system, if altered to fit the tasks. However, when I initially planned to integrate functionality in customized code for a prototype, I presupposed that I would implement functionality from the native Emotiv Framework. Both options should be considered for future BCI HCI projects. If the budgets allow it, it is still preferable to implement native functionality in a customized interface, with the option of customizing machine learning models from a third-party provider, ie. TensorFlow playground.

CONCLUSION

There are is no significant difference in task performance based on cognitive commands acquired with a visual imagery against both visual imagery and visual perception. The results are inconclusive regarding both utilitarian performance and user experience as perceived workload did not different between the two dataset conditions. I have performed experiments with an Emotiv EEG headset and a game interface to find and the participants have reported perceived Workload with a NASA Tlx form. I have assessed the effectiveness and user experience of spontaneous EEG, by testing training data with different cognitive tasks and compared them with number of classifications per session, and the score achieved. There are benefits in user motivation and workload if we can present adjust tasks to skill level and customize stepwise skill acquisition to the individual user. I have hypothesised that a trade-off in dataset directionality may be outweighed by a more accessible workflow, and that different sets of training data are sufficiently equal in efficiency and accuracy. There is no significant evidence of a difference in tested efficiency or accuracy between using visual imagery and visual perception as training data with visual imagery gameplay task. There is no significant evidence of a difference in perceived workload between playing with different sets of training data. The limited sample size places a significant constrain on the conclusions that can be drawn from out data.

There is still need for base research in each technology associated with BCI and Spontaneous EEG. New approaches to machine learning and classification models will open new opportunities for us to design accessible cognitive command tasks.

REFERENCE LIST

Allison, B. Z., & Neuper, C. (2010). Could Anyone Use a BCI? In D. S. Tan & A. Nijholt (Eds.), *Brain-Computer Interfaces* (pp. 35–54). London: Springer London. https://doi.org/10.1007/978-1-84996-272-8_3

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. *Computer Graphics Forum*, *30*(3), 791–800. https://doi.org/10.1111/j.1467-8659.2011.01928.x

Ball, L. J., Beukelman, D. R., & Pattee, G. L. (2004). Acceptance of Augmentative and Alternative Communication Technology by Persons with Amyotrophic Lateral Sclerosis. *Augmentative and Alternative Communication*, *20*(2), 113–122. https://doi.org/10.1080/0743461042000216596

Bashashati, H., Ward, R. K., Birch, G. E., & Bashashati, A. (2015). Comparing Different Classifiers in Sensory Motor Brain Computer Interfaces. *PLOS ONE*, *10*(6). https://doi.org/10.1371/journal.pone.0129435

Bedny, G. Z., & Karwowski, W. (Eds.). (2011). *Human-computer interaction and operators' performance: optimizing work design with activity theory*. Boca Raton, FL: CRC Press.

Beukelman, D. R., Fager, S., Ball, L., & Dietz, A. (2007). AAC for adults with acquired neurological conditions: A review. *Augmentative and Alternative Communication*, *23*(3), 230–242. https://doi.org/10.1080/07434610701553668

Blankertz, B., Dornhege, G., Krauledat, M., Muller, K.-R., Kunzmann, V., Losch, F., & Curio, G. (2006). The Berlin Brain–Computer Interface: EEG-Based Communication Without

Subject Training. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *14*(2), 147–152. https://doi.org/10.1109/TNSRE.2006.875557

Bobrov, P., Frolov, A., Cantor, C., Fedulova, I., Bakhnyan, M., & Zhavoronkov, A. (2011). Brain-Computer Interface Based on Generation of Visual Images. *PLoS ONE*, *6*(6), e20674. https://doi.org/10.1371/journal.pone.0020674

Chavarriaga, R., Fried-Oken, M., Kleih, S., Lotte, F., & Scherer, R. (2017). Heading for new shores! Overcoming pitfalls in BCI design. *Brain-Computer Interfaces*, *4*(1–2), 60–73. https://doi.org/10.1080/2326263X.2016.1263916

Cook, A. M., & Polgar, J. M. (2015). Augmentative and Alternative Communication Systems. In *Assistive Technologies* (pp. 411–456). Elsevier. https://doi.org/10.1016/B978-0-323-09631-7.00016-8

Csikszentmihalyi, M. (1997). *Finding flow: the psychology of engagement with everyday life* (1st ed). New York: BasicBooks.

Curran, E. A., & Stokes, M. J. (2003). Learning to control brain activity: A review of the production and control of EEG components for driving brain–computer interface (BCI) systems. *Brain and Cognition*, *51*(3), 326–336. https://doi.org/10.1016/S0278-2626(03)00036-8

Dentico, D., Cheung, B. L., Chang, J.-Y., Guokas, J., Boly, M., Tononi, G., & Van Veen, B. (2014). Reversal of cortical information flow during visual imagery as compared to visual perception. *NeuroImage*, *100*, 237–243. https://doi.org/10.1016/j.neuroimage.2014.05.081

DeVellis, R. F. (2006). Classical Test Theory: *Medical Care*, *44*(Suppl 3), S50–S59. https://doi.org/10.1097/01.mlr.0000245426.10853.30

Diamond, A. (2013). Executive Functions. *Annual Review of Psychology*, *64*(1), 135–168. https://doi.org/10.1146/annurev-psych-113011-143750

EEG > introduction Biomedical Signals Acquisition. (2005). Retrieved June 18, 2018, from https://www.medicine.mcgill.ca/physio/vlab/biomed_signals/eeg_n.htm

Friedrich, E. V. C., Scherer, R., & Neuper, C. (2013). Long-term evaluation of a 4-class imagery-based brain–computer interface. *Clinical Neurophysiology*, *124*(5), 916–927. https://doi.org/10.1016/j.clinph.2012.11.010

Galy, E., Paxion, J., & Berthelon, C. (2018). Measuring mental workload with the NASA-TLX needs to examine each dimension rather than relying on the global score: an example with driving. *Ergonomics*, *61*(4), 517–527. https://doi.org/10.1080/00140139.2017.1369583

Ganis, G., Thompson, W. L., & Kosslyn, S. M. (2004). Brain areas underlying visual mental imagery and visual perception: an fMRI study. *Cognitive Brain Research*, *20*(2), 226–241. https://doi.org/10.1016/j.cogbrainres.2004.02.012

Gramann, K., Fairclough, S. H., Zander, T. O., & Ayaz, H. (2017). Editorial: Trends in Neuroergonomics. *Frontiers in Human Neuroscience*, *11*. https://doi.org/10.3389/fnhum.2017.00165

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In *Advances in Psychology* (Vol. 52, pp. 139–183). Elsevier. https://doi.org/10.1016/S0166-4115(08)62386-9

Krauledat, M., Tangermann, M., Blankertz, B., & Müller, K.-R. (2008). Towards Zero Training for Brain-Computer Interfacing. *PLoS ONE*, *3*(8), e2967. https://doi.org/10.1371/journal.pone.0002967

Kronegg, Julian, Voloshynovskiy, S., & Pun, T. (2005). Analysis of bit-rate definitions for Brain-Computer Interfaces. Presented at the Int. Conf. on Human-computer Interaction (HCI'05), Las Vegas, Nevada, USA.

Kronegg, Julien, Alecu, T., & Pun, T. (2003). Information theoretic bit-rate optimization for average trial protocol Brain-Computer Interfaces. In *HCI International 2003, 10th International Conference on Human-Computer Interaction*. Retrieved from https://archive-ouverte.unige.ch/unige:47809

Kübler, A., & Birbaumer, N. (2008). Brain–computer interfaces and communication in paralysis: Extinction of goal directed thinking in completely paralysed patients? *Clinical Neurophysiology*, *119*(11), 2658–2666. https://doi.org/10.1016/j.clinph.2008.06.019

Kübler, Andrea, Holz, E. M., Riccio, A., Zickler, C., Kaufmann, T., Kleih, S. C., … Mattia, D. (2014). The User-Centered Design as Novel Perspective for Evaluating the Usability of BCI-Controlled Applications. *PLoS ONE*, *9*(12), e112392. https://doi.org/10.1371/journal.pone.0112392

Leeb, R., Perdikis, S., Tonin, L., Biasiucci, A., Tavella, M., Creatura, M., … Millán, J. d.R. (2013). Transferring brain–computer interfaces beyond the laboratory: Successful application control for motor-disabled users. *Artificial Intelligence in Medicine*, *59*(2), 121–132. https://doi.org/10.1016/j.artmed.2013.08.004

Light, J. (1989). Toward a definition of communicative competence for individuals using augmentative and alternative communication systems. *Augmentative and Alternative Communication*, *5*(2), 137–144. https://doi.org/10.1080/07434618912331275126

Light, J., & McNaughton, D. (2013). Putting People First: Re-Thinking the Role of Technology in Augmentative and Alternative Communication Intervention. *Augmentative and*

Alternative Communication, *29*(4), 299–309. https://doi.org/10.3109/07434618.2013.848935

Lotte, F., Larrue, F., & Mühl, C. (2013). Flaws in current human training protocols for spontaneous Brain-Computer Interfaces: lessons learned from instructional design. *Frontiers in Human Neuroscience*, *7*. https://doi.org/10.3389/fnhum.2013.00568

McFarland, D. J., & Wolpaw, J. R. (2011). Brain-computer interfaces for communication and control. *Communications of the ACM*, *54*(5), 60. https://doi.org/10.1145/1941487.1941506

McKinley, R. A., McIntire, L. K., Schmidt, R., Repperger, D. W., & Caldwell, J. A. (2011). Evaluation of Eye Metrics as a Detector of Fatigue. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *53*(4), 403–414. https://doi.org/10.1177/0018720811411297

McMahan, T., Parberry, I., & Parsons, T. D. (2015). Modality specific assessment of video game player's experience using the Emotiv. *Entertainment Computing*, *7*, 1–6. https://doi.org/10.1016/j.entcom.2015.03.001

Mladenović, J., Frey, J., Bonnet-Save, M., Mattout, J., & Lotte, F. (2017). The Impact of Flow in an EEG-based Brain Computer Interface.

Moroney, W. F., Biers, D. W., Eggemeier, F. T., & Mitchell, J. A. (1992). A comparison of two scoring procedures with the NASA task load index in a simulated flight task (pp. 734–740). IEEE. https://doi.org/10.1109/NAECON.1992.220513

Nicolas-Alonso, L. F., & Gomez-Gil, J. (2012). Brain Computer Interfaces, a Review. *Sensors*, *12*(12), 1211–1279. https://doi.org/10.3390/s120201211

Proctor, R. W., & Vu, K.-P. L. (2010). Cumulative Knowledge and Progress in Human Factors. *Annual Review of Psychology*, *61*(1), 623–651. https://doi.org/10.1146/annurev.psych.093008.100325

Ramadan, R. A., & Vasilakos, A. V. (2017). Brain computer interface: control signals review. *Neurocomputing*, *223*, 26–44. https://doi.org/10.1016/j.neucom.2016.10.024

Shourie, N., Firoozabadi, M., & Badie, K. (2014). Analysis of EEG Signals Related to Artists and Nonartists during Visual Perception, Mental Imagery, and Rest Using Approximate Entropy. *BioMed Research International*, *2014*, 1–10. https://doi.org/10.1155/2014/764382

stikling/FlapPyBird: A Flappy Bird Clone using python-pygame. (n.d.). Retrieved November 21, 2017, from https://github.com/stikling/FlapPyBird

Taylor, G. S., & Schmidt, C. (2012). Empirical Evaluation of the Emotiv EPOC BCI Headset for the Detection of Mental Actions. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *56*(1), 193–197. https://doi.org/10.1177/1071181312561017

VanLehn, K. (1996). COGNITIVE SKILL ACQUISITION. *Annual Review of Psychology*, *47*(1), 513–539. https://doi.org/10.1146/annurev.psych.47.1.513

Winkler, I., Haufe, S., & Tangermann, M. (2011). Automatic Classification of Artifactual ICA-Components for Artifact Removal in EEG Signals. *Behavioral and Brain Functions*, *7*(1), 30. https://doi.org/10.1186/1744-9081-7-30

Zickler, C., Halder, S., Kleih, S. C., Herbert, C., & Kübler, A. (2013). Brain Painting: Usability testing according to the user-centered design in end users with severe motor paralysis. *Artificial Intelligence in Medicine*, *59*(2), 99–110. https://doi.org/10.1016/j.artmed.2013.08.003

Zydney, J. M., Hai-Jew, S., Renninger, K. A., List, A., Hardy, I., Koerber, S., … Blumschein, P. (2012). Spatial Ability. In N. M. Seel (Ed.), *Encyclopedia of the Sciences of Learning* (pp. 3154–3154). Boston, MA: Springer US. https://doi.org/10.1007/978-1-4419-1428-6_2426

s

Appendix 1. Nasa TLX

Figure 8.6

NASA Task Load Index

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

Name	Task	Date

Mental Demand — How mentally demanding was the task?

Very Low — Very High

Physical Demand — How physically demanding was the task?

Very Low — Very High

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

Very Low — Very High

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

Perfect — Failure

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

Very Low — Very High

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

Very Low — Very High

Appendix 2. DATA FROM GAMETASK IN ALL SESSIONS

Session	Participant	Method	Mental Load	Physical Load	Temporal Load	Effort	Frustration	Performance	Score	Bitrate
1	1	1	20	1	10	20	8	16	0	148
4	3	1	20	20	15	19	20	20	2	103
5	3	2	20	20	14	20	19	20	1	120
6	4	2	12	5	13	15	8	20	1	108
7	4	2	12	8	8	9	7	9	4	193
49	8	2	5	11	11	6	7	3	4	406
15	8	1	18	2	9	17	17	18	2	83
16	9	2	0	0	0	0	0	0	0	100
17	9	2	19	13	10	18	15	5	4	234
18	10	1	11	1	2	11	3	19	0	22
19	10	1	11	2	1	11	3	19	0	24
20	11	1	13	9	6	10	4	7	4	381
47	12	2	17	8	6	17	18	16	0	28
48	12	1	12	7	11	17	18	11	0	36
27	14	1	11	5	14	16	16	17	0	85
26	14	2	11	5	14	16	16	17	1	279
46	16	1	18	1	1	20	1	11	0	62
45	16	2	18	1	1	20	1	11	6	387
35	18	2	19	12	2	19	19	13	7	354
36	18	1	18	10	3	20	20	14	3	171
38	20	2	18	5	14	18	18	19	0	42
39	20	1	20	4	16	17	20	20	0	21
43	22	2	20	1	12	20	20	20	0	20
42	22	1	20	1	9	20	15	12	1	148

Method 1 = Visual perception. Method 2 = Visual imagery

Appendix 3. NSD

MELDESKJEMA

Meldeskjema (versjon 1.6) for forsknings- og studentprosjekt som medfører meldeplikt eller konsesjonsplikt
(jf. personopplysningsloven og helseregisterloven med forskrifter).

1. Intro		
Samles det inn direkte personidentifiserende opplysninger?	Ja ● Nei ○	En person vil være direkte identifiserbar via navn, personnummer, eller andre personentydige kjennetegn. Les mer om hva personopplysninger er. NB! Selv om opplysningene skal anonymiseres i oppgave/rapport, må det krysses av dersom det skal innhentes/registreres personidentifiserende opplysninger i forbindelse med prosjektet. Les mer om hva behandling av personopplysninger innebærer.
Hvis ja, hvilke?	■ Navn □ 11-sifret fødselsnummer □ Adresse ■ E-post ■ Telefonnummer □ Annet	
Annet, spesifiser hvilke		
Skal direkte personidentifiserende opplysninger kobles til datamaterialet (koblingsnøkkel)?	Ja ● Nei ○	Merk at meldeplikten utløses selv om du ikke får tilgang til koblingsnøkkel, slik fremgangsmåten ofte er når man benytter en databehandler.
Samles det inn bakgrunnsopplysninger som kan identifisere enkeltpersoner (indirekte personidentifiserende opplysninger)?	Ja ○ Nei ●	En person vil være indirekte identifiserbar dersom det er mulig å identifisere vedkommende gjennom bakgrunnsopplysninger som for eksempel bostedskommune eller arbeidsplass/skole kombinert med opplysninger som alder, kjønn, yrke, diagnose, etc. NB! For at stemme skal regnes som personidentifiserende, må denne bli registrert i kombinasjon med andre opplysninger, slik at personer kan gjenkjennes.
Hvis ja, hvilke		
Skal det registreres personopplysninger (direkte/indirekte/via IP-/epost adresse, etc) ved hjelp av nettbaserte spørreskjema?	Ja ● Nei ○	Les mer om nettbaserte spørreskjema.
Blir det registrert personopplysninger på digitale bilde- eller videoopptak?	Ja ○ Nei ●	Bilde/videoopptak av ansikter vil regnes som personidentifiserende.
Søkes det vurdering fra REK om hvorvidt prosjektet er omfattet av helseforskningsloven?	Ja ○ Nei ●	NB! Dersom REK (Regional Komité for medisinsk og helsefaglig forskningsetikk) har vurdert prosjektet som helseforskning, er det ikke nødvendig å sende inn meldeskjema til personvernombudet (NB! Gjelder ikke prosjekter som skal benytte data fra pseudonyme helseregistre). Les mer. Dersom tilbakemelding fra REK ikke foreligger, anbefaler vi at du avventer videre utfylling til svar fra REK foreligger.
2. Prosjekttittel		
Prosjekttittel	Comparing the Reliability of Perceived visual stimuli and Imagined visual stimuli as a Data Entry strategy, for commands with spontaneous EEG	Oppgi prosjektets tittel. NB! Dette kan ikke være «Masteroppgave» eller liknende, navnet må beskrive prosjektets innhold.
3. Behandlingsansvarlig institusjon		
Institusjon	Høgskolen i Oslo og Akershus	Velg den institusjonen du er tilknyttet. Alle nivå må oppgis. Ved studentprosjekt er det studentens tilknytning som er avgjørende. Dersom institusjonen ikke finnes på listen, har den ikke avtale med NSD som personvernombud. Vennligst ta kontakt med institusjonen. Les mer om behandlingsansvarlig institusjon.
Avdeling/Fakultet	Fakultet for teknologi, kunst og design	
Institutt	Institutt for informasjonsteknologi	
4. Daglig ansvarlig (forsker, veileder, stipendiat)		

Side 1

Participants and sessions

		Participant	Condition 1-session	Participant	Condition 2-session
1 *	Paired	3	4	3	5
2*	Paired	8	15	8	49
3*	Paired	14	27	14	26
4*	Paired	18	36	18	35
5*	Paired	20	39	20	38
6*	Paired	22	42	22	43
7*	Paired	16	46	16	45
8*	Paired	12	48	12	47
9**	Second sessions	16	46	16	30
10		8	14		
11**	Grouped	10	18	4	6
12**	Grouped	10	19	4	7
13**	Grouped Single try /Grouped	1	1	9	17
14**	Grouped Single try /Grouped	11	20	9	16
15	Grouped Single try	19	34		

APPENDIX 5. PROTOTYPE CODE

```
from itertools import cycle

import random

import sys

import pygame

from pygame.locals import *

FPS = 30

SCREENWIDTH  = 288

SCREENHEIGHT = 512

# amount by which base can maximum shift to left

PIPEGAPSIZE  = 100 # gap between upper and lower part of pipe

BASEY        = SCREENHEIGHT * 0.79

# image, sound and hitmask  dicts

IMAGES, SOUNDS, HITMASKS = {}, {}, {}

# list of all possible players (tuple of 3 positions of flap)

PLAYERS_LIST = (

    # red bird

    (

        'assets/sprites/redbird-upflap.png',

        'assets/sprites/redbird-midflap.png',

        'assets/sprites/redbird-downflap.png',

    ),
```

```
    # blue bird
    (
        # amount by which base can maximum shift to left
        'assets/sprites/bluebird-upflap.png',
        'assets/sprites/bluebird-midflap.png',
        'assets/sprites/bluebird-downflap.png',
    ),
    # yellow bird
    (
        'assets/sprites/yellowbird-upflap.png',
        'assets/sprites/yellowbird-midflap.png',
        'assets/sprites/yellowbird-downflap.png',
    ),
)
# list of backgrounds
BACKGROUNDS_LIST = (
    'assets/sprites/background-day.png',
    'assets/sprites/background-night.png',
)
# list of pipes
PIPES_LIST = (
    'assets/sprites/pipe-green.png',
    'assets/sprites/pipe-red.png',
```

```
)
try:
    xrange
except NameError:
    xrange = range
def main():
    global SCREEN, FPSCLOCK
    pygame.init()
    FPSCLOCK = pygame.time.Clock()
    SCREEN = pygame.display.set_mode((SCREENWIDTH, SCREENHEIGHT))
    pygame.display.set_caption('Flappy Bird')
    # numbers sprites for score display
    IMAGES['numbers'] = (
        pygame.image.load('assets/sprites/0.png').convert_alpha(),
        pygame.image.load('assets/sprites/1.png').convert_alpha(),
        pygame.image.load('assets/sprites/2.png').convert_alpha(),
        pygame.image.load('assets/sprites/3.png').convert_alpha(),
        pygame.image.load('assets/sprites/4.png').convert_alpha(),
        pygame.image.load('assets/sprites/5.png').convert_alpha(),
        pygame.image.load('assets/sprites/6.png').convert_alpha(),
        pygame.image.load('assets/sprites/7.png').convert_alpha(),
        pygame.image.load('assets/sprites/8.png').convert_alpha(),
        pygame.image.load('assets/sprites/9.png').convert_alpha()
```

```
)
# game over sprite
IMAGES['gameover'] = pygame.image.load('assets/sprites/gameover.png').convert_alpha()
# message sprite for welcome screen
IMAGES['message'] = pygame.image.load('assets/sprites/message.png').convert_alpha()
# base (ground) sprite
IMAGES['base'] = pygame.image.load('assets/sprites/base.png').convert_alpha()
# sounds
if 'win' in sys.platform:
    soundExt = '.wav'
else:
    soundExt = '.ogg'
SOUNDS['die']    = pygame.mixer.Sound('assets/audio/die' + soundExt)
SOUNDS['hit']    = pygame.mixer.Sound('assets/audio/hit' + soundExt)
SOUNDS['point']  = pygame.mixer.Sound('assets/audio/point' + soundExt)
SOUNDS['swoosh'] = pygame.mixer.Sound('assets/audio/swoosh' + soundExt)
SOUNDS['wing']   = pygame.mixer.Sound('assets/audio/wing' + soundExt)
while True:
    # select random background sprites
    randBg = random.randint(0, len(BACKGROUNDS_LIST) - 1)
    IMAGES['background'] = pygame.image.load(BACKGROUNDS_LIST[randBg]).convert()

    # select random player sprites
```

```
randPlayer = random.randint(0, len(PLAYERS_LIST) - 1)
IMAGES['player'] = (
    pygame.image.load(PLAYERS_LIST[randPlayer][0]).convert_alpha(),
    pygame.image.load(PLAYERS_LIST[randPlayer][1]).convert_alpha(),
    pygame.image.load(PLAYERS_LIST[randPlayer][2]).convert_alpha(),
)
# select random pipe sprites
pipeindex = random.randint(0, len(PIPES_LIST) - 1)
IMAGES['pipe'] = (
    pygame.transform.rotate(
        pygame.image.load(PIPES_LIST[pipeindex]).convert_alpha(), 180),
    pygame.image.load(PIPES_LIST[pipeindex]).convert_alpha(),
)
# hismask for pipes
HITMASKS['pipe'] = (
    getHitmask(IMAGES['pipe'][0]),
    getHitmask(IMAGES['pipe'][1]),
)
# hitmask for player
HITMASKS['player'] = (
    getHitmask(IMAGES['player'][0]),
    getHitmask(IMAGES['player'][1]),
    getHitmask(IMAGES['player'][2]),
```

```
        )
        movementInfo = showWelcomeAnimation()
        crashInfo = mainGame(movementInfo)
        showGameOverScreen(crashInfo)
def showWelcomeAnimation():
    """Shows welcome screen animation of flappy bird"""
    # index of player to blit on screen
    playerIndex = 0
    playerIndexGen = cycle([0, 1, 2, 1])
    # iterator used to change playerIndex after every 5th iteration
    loopIter = 0
    playerx = int(SCREENWIDTH * 0.2)
    playery = int((SCREENHEIGHT - IMAGES['player'][0].get_height()) / 2)
    messagex = int((SCREENWIDTH - IMAGES['message'].get_width()) / 2)
    messagey = int(SCREENHEIGHT * 0.12)
    basex = 0
    # amount by which base can maximum shift to left
    baseShift = IMAGES['base'].get_width() - IMAGES['background'].get_width()
    # player shm for up-down motion on welcome screen
    playerShmVals = {'val': 0, 'dir': 1}

    while True:
        for event in pygame.event.get():
```

```
        if event.type == QUIT or (event.type == KEYDOWN and event.key == K_ESCAPE):
            pygame.quit()
            sys.exit()
        if event.type == KEYDOWN and (event.key == K_SPACE or event.key == K_UP):
            # make first flap sound and return values for mainGame
            SOUNDS['wing'].play()
            return {
                'playery': playery + playerShmVals['val'],
                'basex': basex,
                'playerIndexGen': playerIndexGen,
            }
    # adjust playery, playerIndex, basex
    if (loopIter + 1) % 5 == 0:
        playerIndex = next(playerIndexGen)
    loopIter = (loopIter + 1) % 30
    basex = -((-basex + 4) % baseShift)
    playerShm(playerShmVals)
    # draw sprites
    SCREEN.blit(IMAGES['background'], (0,0))
    SCREEN.blit(IMAGES['player'][playerIndex],
                (playerx, playery + playerShmVals['val']))
    SCREEN.blit(IMAGES['message'], (messagex, messagey))
    SCREEN.blit(IMAGES['base'], (basex, BASEY))
```

```
        pygame.display.update()
        FPSCLOCK.tick(FPS)
def mainGame(movementInfo):
    score = playerIndex = loopIter = 0
    playerIndexGen = movementInfo['playerIndexGen']
    playerx, playery = int(SCREENWIDTH * 0.2), movementInfo['playery']
    basex = movementInfo['basex']
    baseShift = IMAGES['base'].get_width() - IMAGES['background'].get_width()
    # get 2 new pipes to add to upperPipes lowerPipes list
    newPipe1 = getRandomPipe()
    newPipe2 = getRandomPipe()
    # list of upper pipes
    upperPipes = [
        {'x': SCREENWIDTH + 200, 'y': newPipe1[0]['y']},
        {'x': SCREENWIDTH + 200 + (SCREENWIDTH / 2), 'y': newPipe2[0]['y']},
    ]
    # list of lowerpipe
    lowerPipes = [
        {'x': SCREENWIDTH + 200, 'y': newPipe1[1]['y']},
        {'x': SCREENWIDTH + 200 + (SCREENWIDTH / 2), 'y': newPipe2[1]['y']},
    ]
    pipeVelX = -4
    # player velocity, max velocity, downward accleration, accleration on flap
```

```
playerVelY    =  -9   # player's velocity along Y, default same as playerFlapped
playerMaxVelY =  10   # max vel along Y, max descend speed
playerMinVelY =  -8   # min vel along Y, max ascend speed
playerAccY    =   1   # players downward accleration
playerRot     =  45   # player's rotation
playerVelRot  =   3   # angular speed
playerRotThr  =  20   # rotation threshold
playerFlapAcc =  -9   # players speed on flapping
playerFlapped = False # True when player flaps
while True:
    for event in pygame.event.get():
        if event.type == QUIT or (event.type == KEYDOWN and event.key == K_ESCAPE):
            pygame.quit()
            sys.exit()
        if event.type == KEYDOWN and (event.key == K_SPACE or event.key == K_UP):
            if playery > -2 * IMAGES['player'][0].get_height():
                playerVelY = playerFlapAcc
                playerFlapped = True
                SOUNDS['wing'].play()
    # check for crash here
    crashTest = checkCrash({'x': playerx, 'y': playery, 'index': playerIndex},
                           upperPipes, lowerPipes)
    if crashTest[0]:
```

```
    return {
        'y': playery,
        'groundCrash': crashTest[1],
        'basex': basex,
        'upperPipes': upperPipes,
        'lowerPipes': lowerPipes,
        'score': score,
        'playerVelY': playerVelY,
        'playerRot': playerRot
    }
# check for score
playerMidPos = playerx + IMAGES['player'][0].get_width() / 2
for pipe in upperPipes:
    pipeMidPos = pipe['x'] + IMAGES['pipe'][0].get_width() / 2
    if pipeMidPos <= playerMidPos < pipeMidPos + 4:
        score += 1
        SOUNDS['point'].play()
# playerIndex basex change
if (loopIter + 1) % 3 == 0:
    playerIndex = next(playerIndexGen)
loopIter = (loopIter + 1) % 30
basex = -((-basex + 100) % baseShift)
# rotate the player
```

```
if playerRot > -90:
    playerRot -= playerVelRot
# player's movement
if playerVelY < playerMaxVelY and not playerFlapped:
    playerVelY += playerAccY
if playerFlapped:
    playerFlapped = False
    # more rotation to cover the threshold (calculated in visible rotation)
    playerRot = 45
playerHeight = IMAGES['player'][playerIndex].get_height()
playery += min(playerVelY, BASEY - playery - playerHeight)
# move pipes to left
for uPipe, lPipe in zip(upperPipes, lowerPipes):
    uPipe['x'] += pipeVelX
    lPipe['x'] += pipeVelX
# add new pipe when first pipe is about to touch left of screen
if 0 < upperPipes[0]['x'] < 5:
    newPipe = getRandomPipe()
    upperPipes.append(newPipe[0])
    lowerPipes.append(newPipe[1])
# remove first pipe if its out of the screen
if upperPipes[0]['x'] < -IMAGES['pipe'][0].get_width():
    upperPipes.pop(0)
```

```
            lowerPipes.pop(0)

        # draw sprites

        SCREEN.blit(IMAGES['background'], (0,0))

        for uPipe, lPipe in zip(upperPipes, lowerPipes):

            SCREEN.blit(IMAGES['pipe'][0], (uPipe['x'], uPipe['y']))

            SCREEN.blit(IMAGES['pipe'][1], (lPipe['x'], lPipe['y']))

        SCREEN.blit(IMAGES['base'], (basex, BASEY))

        # print score so player overlaps the score

        showScore(score)

        # Player rotation has a threshold

        visibleRot = playerRotThr

        if playerRot <= playerRotThr:

            visibleRot = playerRot

        playerSurface = pygame.transform.rotate(IMAGES['player'][playerIndex], visibleRot)

        SCREEN.blit(playerSurface, (playerx, playery))

        pygame.display.update()

        FPSCLOCK.tick(FPS)

def showGameOverScreen(crashInfo):

    """crashes the player down ans shows gameover image"""

    score = crashInfo['score']

    playerx = SCREENWIDTH * 0.2

    playery = crashInfo['y']

    playerHeight = IMAGES['player'][0].get_height()
```

```
playerVelY = crashInfo['playerVelY']

playerAccY = 2

playerRot = crashInfo['playerRot']

playerVelRot = 7

basex = crashInfo['basex']

upperPipes, lowerPipes = crashInfo['upperPipes'], crashInfo['lowerPipes']

# play hit and die sounds

SOUNDS['hit'].play()

if not crashInfo['groundCrash']:

    SOUNDS['die'].play()

while True:

    for event in pygame.event.get():

        if event.type == QUIT or (event.type == KEYDOWN and event.key == K_ESCAPE):

            pygame.quit()

            sys.exit()

        if event.type == KEYDOWN and (event.key == K_SPACE or event.key == K_UP):

            if playery + playerHeight >= BASEY - 1:

                return

    # player y shift

    if playery + playerHeight < BASEY - 1:

        playery += min(playerVelY, BASEY - playery - playerHeight)

    # player velocity change

    if playerVelY < 15:
```

```
            playerVelY += playerAccY

        # rotate only when it's a pipe crash
        if not crashInfo['groundCrash']:
            if playerRot > -90:
                playerRot -= playerVelRot

        # draw sprites
        SCREEN.blit(IMAGES['background'], (0,0))
        for uPipe, lPipe in zip(upperPipes, lowerPipes):
            SCREEN.blit(IMAGES['pipe'][0], (uPipe['x'], uPipe['y']))
            SCREEN.blit(IMAGES['pipe'][1], (lPipe['x'], lPipe['y']))
        SCREEN.blit(IMAGES['base'], (basex, BASEY))
        showScore(score)
        playerSurface = pygame.transform.rotate(IMAGES['player'][1], playerRot)
        SCREEN.blit(playerSurface, (playerx,playery))
        FPSCLOCK.tick(FPS)
        pygame.display.update()
def playerShm(playerShm):
    """oscillates the value of playerShm['val'] between 8 and -8"""
    if abs(playerShm['val']) == 8:
        playerShm['dir'] *= -1
    if playerShm['dir'] == 1:
         playerShm['val'] += 1
    else:
```

```
        playerShm['val'] -= 1

def getRandomPipe():

    """returns a randomly generated pipe"""

    # y of gap between upper and lower pipe

    gapY = random.randrange(0, int(BASEY * 0.6 - PIPEGAPSIZE))

    gapY += int(BASEY * 0.2)

    pipeHeight = IMAGES['pipe'][0].get_height()

    pipeX = SCREENWIDTH + 10

    return [

        {'x': pipeX, 'y': gapY - pipeHeight},  # upper pipe

        {'x': pipeX, 'y': gapY + PIPEGAPSIZE}, # lower pipe

    ]

def showScore(score):

    """displays score in center of screen"""

    scoreDigits = [int(x) for x in list(str(score))]

    totalWidth = 0 # total width of all numbers to be printed

    for digit in scoreDigits:

        totalWidth += IMAGES['numbers'][digit].get_width()

    Xoffset = (SCREENWIDTH - totalWidth) / 2

    for digit in scoreDigits:

        SCREEN.blit(IMAGES['numbers'][digit], (Xoffset, SCREENHEIGHT * 0.1))

        Xoffset += IMAGES['numbers'][digit].get_width()

def checkCrash(player, upperPipes, lowerPipes):
```

```
"""returns True if player collders with base or pipes."""
pi = player['index']
player['w'] = IMAGES['player'][0].get_width()
player['h'] = IMAGES['player'][0].get_height()
# if player crashes into ground
if player['y'] + player['h'] >= BASEY - 1:
    return [True, True]
else:
    playerRect = pygame.Rect(player['x'], player['y'],
                  player['w'], player['h'])
    pipeW = IMAGES['pipe'][0].get_width()
    pipeH = IMAGES['pipe'][0].get_height()
    for uPipe, lPipe in zip(upperPipes, lowerPipes):
        # upper and lower pipe rects
        uPipeRect = pygame.Rect(uPipe['x'], uPipe['y'], pipeW, pipeH)
        lPipeRect = pygame.Rect(lPipe['x'], lPipe['y'], pipeW, pipeH)
        # player and upper/lower pipe hitmasks
        pHitMask = HITMASKS['player'][pi]
        uHitmask = HITMASKS['pipe'][0]
        lHitmask = HITMASKS['pipe'][1]
        # if bird collided with upipe or lpipe
        uCollide = pixelCollision(playerRect, uPipeRect, pHitMask, uHitmask)
        lCollide = pixelCollision(playerRect, lPipeRect, pHitMask, lHitmask)
```

```
            if uCollide or lCollide:
                return [True, False]
    return [False, False]
def pixelCollision(rect1, rect2, hitmask1, hitmask2):
    """Checks if two objects collide and not just their rects"""
    rect = rect1.clip(rect2)
    if rect.width == 0 or rect.height == 0:
        return False
    x1, y1 = rect.x - rect1.x, rect.y - rect1.y
    x2, y2 = rect.x - rect2.x, rect.y - rect2.y
    for x in xrange(rect.width):
        for y in xrange(rect.height):
            if hitmask1[x1+x][y1+y] and hitmask2[x2+x][y2+y]:
                return True
    return False
def getHitmask(image):
    """returns a hitmask using an image's alpha."""
    mask = []
    for x in xrange(image.get_width()):
        mask.append([])
        for y in xrange(image.get_height()):
            mask[x].append(bool(image.get_at((x,y))[3]))
    return mask
```

```
if __name__ == '__main__':
    main()
```

RESULTATER

mdl1 =

Linear regression model:

y ~ 1 + x1

Estimated Coefficients:

	Estimate	SE	tStat	pValue
(Intercept)	-1.2643	2.1049	-0.60065	0.57005
x1	0.027405	0.024972	1.0974	0.31453

Number of observations: 8, Error degrees of freedom: 6

Root Mean Squared Error: 1.18

R-squared: 0.167, Adjusted R-Squared 0.0284

F-statistic vs. constant model: 1.2, p-value = 0.315

mdl2 =

Linear regression model:

y ~ 1 + x1

Estimated Coefficients:

	Estimate	SE	tStat	pValue
(Intercept)	7.979	3.5723	2.2336	0.066938
x1	-0.07027	0.043299	-1.6229	0.15574

Number of observations: 8, Error degrees of freedom: 6

SAMPLE T-TEST

Two-Sample *t*-test

The two-sample *t*-test is a parametric test that compares the location parameter of two independent data samples.

The test statistic is

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_x^2}{n} + \frac{s_y^2}{m}}},$$

where $\bar{x}$ and $\bar{y}$ are the sample means, s_x and s_y are the sample standard deviations, and n and m are the sample sizes.
In the case where it is assumed that the two data samples are from populations with equal variances, the test statistic under the null hypothesis has Student's *t* distribution with $n + m - 2$ degrees of freedom, and the sample standard deviations are replaced by the pooled standard deviation

$$s = \sqrt{\frac{(n-1)s_x^2 + (m-1)s_y^2}{n+m-2}}.$$

In the case where it is not assumed that the two data samples are from populations with equal variances, the test statistic under the null hypothesis has an approximate Student's *t* distribution with a number of degrees of freedom given by Satterthwaite's approximation. This test is sometimes called Welch's *t*-test.

Figure 0.2 TwoSample t-test definition from MatLab

h1 =	h2 =
0	0
p1 =	p2 =
0.2322	0.7817
ci =	ci =
-3.7363	-24.7051

Figure 0.1 Text outout of ttests for h1 and h2