A Novel Near-Infrared Spectroscopy Brain-Computer Interface for the Detection of Emotional Valence in Children

By

Erica Floreani

A thesis submitted in conformity with the requirements
for the degree of Master of Applied Science

Institute for Biomaterials & Biomedical Engineering
University of Toronto

A Novel Near Infrared Spectroscopy Brain-Computer Interface for the Detection of Emotional Valence in Children

Erica Floreani

Master of Applied Science

Institute for Biomaterials & Biomedical Engineering

University of Toronto

2019

Abstract

Brain-computer interfaces (BCI) are being investigated as an access pathway to communication for individuals with physical disabilities as they bypass the need for voluntary motor control to operate. However, to date, minimal research has investigated the use of BCIs for children. This thesis describes the development of a paediatric brain-computer interface to identify positive and negative emotional states from changes in hemodynamic activity of the pre-frontal cortex. To train and test the BCI, 10 typically developing children aged 8-14 underwent a session of emotion induction trials while their hemodynamic activity was measured with near-infrared spectroscopy (NIRS). BCI performance was evaluated both offline, using cross-validation, and online, using real-time prediction results. Offline differentiation rates were comparable to studies with adult populations, and online prediction was feasible, yet failed to surpass the 70% accuracy threshold of 'effective' BCI use. Interparticipant and intersession variability in BCI performance was also investigated.

Acknowledgements

I would like to extend my deepest gratitude to Dr. Tom Chau for his guidance and support over the last two years, with respect to my thesis as well as my professional development. I would like to thank Dr. Anne-Marie Guerguerian and Dr. Margot Taylor for their wisdom and valuable insights to this project as my committee members, and to Dr. Azadeh Kushki for stepping in last minute to be my external examiner. Thank you to Dr. Silvia Orlandi, Ka Lun Tam and Pierre Duez for their technical support within the lab, and the rest of my PRISM lab mates for being encouraging and dedicated colleagues. I would also like to thank my family and friends for their unwavering faith in me and for consistently motivating me over the last two years. Finally, I would like to acknowledge my cousin Teddy Floreani for being my inspiration in pursuing my academic career in biomedical engineering and assistive technology.

Table of Contents

List of Tables

Chapter 2

Chapter 3

List of Figures

List of Abbreviations

AAC	Alternative and augmentative communication
aBCI	Affective brain-computer interface
BCI	Brain-computer interface
BOLD	Blood oxygen-level dependent
dmPFC	Dorsomedial prefrontal cortex
EATQ	Early Adolescence Temperament Questionnaire
EEG	Electroencephalography/electroencephalogram
fMRI	Functional magnetic resonance imaging
fNIRS	Functional near-infrared spectroscopy
GAPED	Geneva Affective Pictures Database
[Hb]	Concentration of deoxygenated hemoglobin
[HbO]	Concentration of oxygenated hemoglobin
IAPS	International Affective Pictures System
LASSO	Least absolute shrinkage and selection operator
LDA	Linear discriminant analysis
MI	Mutual information
OASIS	Open Affective Standardized Image Set
OFC	Orbitofrontal cortex
PET	Positron emission tomography
PFC	Prefrontal cortex
pgACC	Pregenual anterior cingulate cortex
rmPFC	Rostromedial prefrontal cortex
RMS	Root mean squared
SAM	Self Assessment Manikin
SFFS	Sequential floating forward search
sgACC	Subgenual anterior cingulate cortex
vmPFC	Ventromedial prefrontal cortex

1. Introduction

1.1 Motivation

The ability to speak, express ourselves and navigate social interactions are essential parts of our lives as humans. Fostering effective communication skills is therefore a critical part of early childhood development (1,2). However, due to possible motor, language, cognitive or sensory impairments, some children are delayed compared to their peers in acquiring communication skills (3). Since these children are not able to properly convey their basic needs, wants and emotions, they are at risk of experiencing significant challenges in their daily lives. If left unaddressed, these challenges can translate to delays in cognitive development and can also compromise their formation of social relationships, their educational success and their ability to gain independence (2–4). Inadequate access to communication can prevent these children from participating fully in their communities and from being included in society (4). Throughout their lives, these children are also much more likely to rate themselves lower in measures of self-esteem and well-being (1).

If children with communication challenges receive assistance in the form of alternative and augmentative communication (AAC) technology, the barriers they face to development, inclusion and quality of life can be significantly reduced (3). AAC devices circumvent challenging physical or cognitive aspects of speech and language and allow the individual to focus on the overall goal of communication itself (5). Early intervention with AAC technologies is critical; the sooner these children receive access to communication, the sooner they will be able to participate in society and the less significant their developmental delays will be (2–4). However, despite the importance of early intervention, much of the existing AAC technology is not designed with children in mind (6). Children view and interact with the world in a significantly different way than adults and this must be reflected in the design of AAC devices in different aspects including their level of engagement, aesthetic appeal and their representation of language concepts (6). These technologies must be integrated into daily activities, such as play, which serve as rich opportunities for language learning, and must also be able to increase in complexity as the child grows and begins to develop a deeper foundation of language (6). A high rate of initial device abandonment (30%) and even higher rate of failure to use devices long-term (75%) suggests that currently, AAC technologies are not appropriately addressing the communication needs of these children (7).

This problem is only exacerbated when considering children whose communication challenges are secondary to severe physical impairments. For children born with cerebral palsy, neurodegenerative disorders or who have had strokes or traumatic brain injury, severe motor impairments can affect their ability to operate an assistive communication device (8). Devices that involve direct access, such as pointing to or touching a screen, are generally not an option for children with physical disabilities (9). Mechanical switches constitute a customizable access technology designed to utilize any functional movements a child can produce. While advances in sensor technology have allowed for the detection of even weak muscular contractions, the child must still be able to produce the motor action consistently, and for many children with physical disabilities, their functional abilities can fluctuate considerably throughout the day with changes in muscle tone, fatigue and even strength of medications (8,9). The use of switches also requires the precise timing of switch activation upon the presentation of a desired communicative output, which may also be difficult for these children (9,10). Eye tracking is another access technology that has been explored for people with physical disabilities. Eye tracking involves the detection of corneal reflections of infrared light to translate saccades and gaze fixation into point and click commands. However, infrared technology is largely affected by environmental context, including the orientation of the source and the detector, making it unreliable in changing environments (8,9). Further, eye tracking devices cannot differentiate between intentional eye movements to control the device and spontaneous changes in eye gaze (8,9).

There is a clear need for a technology that can better accommodate the unique needs of children with physical disabilities and provide them with an access pathway to communication. There are also children who present as locked-in, cognitively capable but unable to produce any functional movements (11), rendering the discussed technologies inaccessible. Recently, brain-computer interfaces (BCIs) have attracted attention as an access technology given their ability to delineate communicative or functional intent of the user through the monitoring and analysis of brain activity. As they require no voluntary motor control to operate, these devices have great potential to facilitate communication for individuals with severe physical disabilities or presenting as locked-in (12,13). Most BCI research to date, however, has been conducted with typically developing adults, with little work done to investigate BCIs for children (14). The research described in this thesis sought to develop a brain-computer interface that could address the basic communication needs of children with severe physical disabilities. This involved developing a novel BCI paradigm that draws on the existing abilities of the child and is conceptually relevant to how children interact with the world. The remainder of this thesis will describe the development and analysis of a brain-computer interface that was developed to use emotional states as an access pathway to communication. The brain-computer interface identified positive and negative emotional states in children from the activation of their prefrontal cortex, measured using near-infrared spectroscopy. This chapter will review the scientific literature supporting this work and outline the specific research questions and hypotheses of the project.

1.2 Literature Review

1.2.1 Brain-Computer Interfaces

Brain-computer interfaces (BCIs) have received attention as an access technology for people with severe physical disabilities because these devices bypass the need for voluntary motor control. To operate a BCI, the user will conduct some mental task or attend to some stimulus that has been chosen to represent a certain communicative or functional intent. The user's neurophysiological signals are then measured using an acquisition modality and analyzed using signal processing and machine learning methods to infer the mental activity or the attended stimulus. The machine's inference leads to an output control signal that can be used to drive a communicative or other functional application (12).

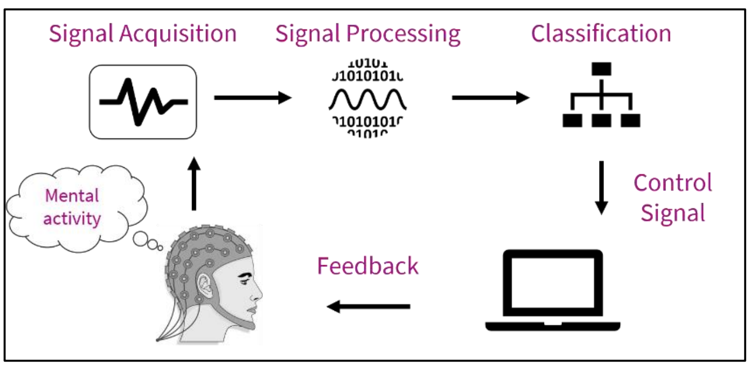


***Figure 1: A schematic overview of a brain-computer interface**. The neurophysiological brain signal that occurs as a result of a mental task or activity is acquired, digitized and then processed and analyzed to produce a control signal that drives an application.*

Neurophysiological signals can be collected from the brain using a variety of different modalities, including electroencephalography (EEG), which measures electrical activity produced by groups of neurons firing in synchrony; electrocorticography, which invasively measures neuronal activity by implanting electrodes directly on the surface of the brain; functional magnetic resonance imaging (fMRI), which measures changes in blood flow in the brain using electromagnetic fields; and functional near infrared spectroscopy (NIRS), which uses infrared light to detect changes in oxygenation levels in cerebral blood flow (15). EEG and NIRS are the two most commonly used modalities in BCI applications, due to their non-invasiveness, relative ease-of-use and portability (12,14). Once the signal is acquired, communicative intent can be decoded using a variety of physiological phenomenon manifested in these signals. One example would be the P300 spike that occurs in the EEG following the presentation of a task-related stimulus; another example would be the sensorimotor rhythms that appear during imagined movement (14). After the collected signal is processed, spatial and temporal features characterizing the physiological phenomena of interest can be extracted and fed into a classification algorithm. Based on these features, the classification algorithm will assign the acquired signal to a corresponding control command (15).

Brain-computer interfaces have been implemented in many different applications, including assistive communication, environmental control, locomotion, motor rehabilitation and even entertainment (16). One highly investigated implementation of brain-computer interfaces for assistive communication is the P300 speller. With the P300 speller, letters are displayed on a grid on the user interface. Each column and row of letters will flash in pseudo-random order. The user focuses on their desired letter, and a P300 peak will occur in their EEG when the column or row containing the desired letter flashes. The system will identify these P300 peaks and associate them with the contingent letter. This process is then repeated to spell out words and sentences (16,17). While these devices work with significant accuracy, they are limited in their information transfer rate (around two words per minute), which is impractical for communication as well as fatiguing and frustrating to the user (15). Furthermore, these spelling paradigms would not be a practical solution for children with physical disabilities who may already be facing associated delays in their language and cognitive development. For a brain-computer interface for children, we may need to look beyond the manipulation of letters and words and consider an alternative approach to communication. One such approach will be explored in the following section.

1.2.2 Emotion as an Access Pathway

The inability to communicate basic needs, wants and emotions can evoke intense feelings of fear, frustration, isolation and despair. In their case study investigating the communication barriers of critically ill patients on life support, Happ et al. strikingly described this experience as being in a state of *voicelessness* (18). The intensity of distress reported in these critically ill individuals emphasizes the emotional significance of our motivation to communicate. It is worth exploring the possibility of using emotional state as an access pathway to communication for children with physical disabilities who, like patients in critical care, face communication challenges that put them at risk for experiencing this state of voicelessness. Emotions underlie many of our basic needs, wants, preferences and opinions about others and our environments. If emotional state can be accessed through a device like a brain-computer interface, it could bypass the need for words to communicate. The following section will delve into the neurophysiological origins of emotions and how they could be accessed in a brain-computer interface.

1.2.2.1 Emotion Systems in the Brain

There is no consensus in the neuroscientific literature on a single definition and model for emotion. There are discrete models of emotion, where different affective states are categorized into separate entities such as happiness, sadness, and anger, where each is accompanied by their own distinct physiological 'fingerprint' (19,20). However, there is a vast amount of variation in these physiological templates across individuals and situations, detracting from the validity of this theory (20,21). Alternatively, dimensional models of emotion postulate that any emotional state falls somewhere along two (or more) fundamental dimensions. These dimensions are most commonly *valence*, the degree of pleasantness, and *arousal*, the degree of activation (19). For example, what we call anger may be an emotional state with low valence and high arousal, while what we call sadness would be an emotional state with both low valence and low arousal. With emerging evidence of their underlying neurophysiological and behavioural correlates, the idea of valence and arousal as 'basic properties' of emotions has been gaining support in the literature (22,23). Finally, there are also appraisal models that attempt to reconcile the complexity and variation within and across different emotional states. These models involve the deconstruction of emotional responses into different components and subsystems that change and interact in different ways depending on the emotion's context (19,24). These models fit

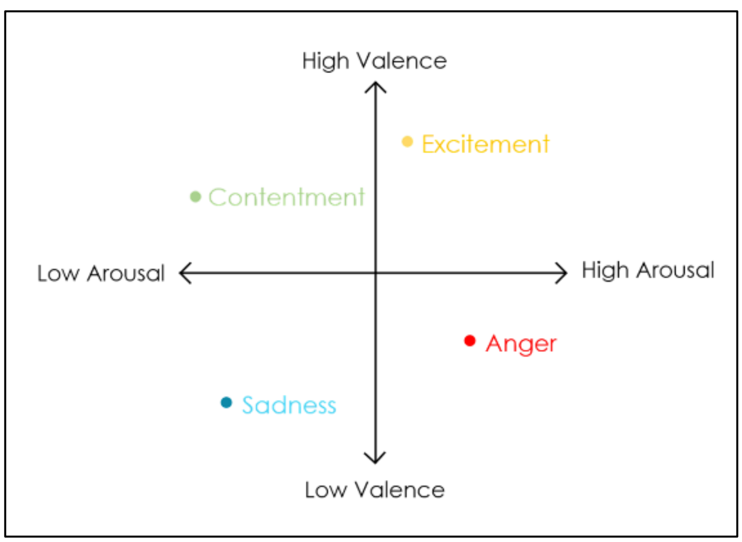


Figure 2: The dimensional model of emotions. *This model postulates that any emotional response will fall somewhere along the two dimensions of valence (degree of pleasantness) and arousal (degree of activation). Four examples, contentment, excitement, anger and sadness, can be seen in their respective quadrants of the valence-arousal plane.*

well with neuroscientific *constructionist approaches*, which attribute neurophysiological phenomena to the networking of a number of different fundamental neural processes and structures (19).

The neurophysiological experience of emotions can be described in two distinct parts; the body's physiological response to affective stimuli (the 'emotional state') and the conscious experience and interpretation of these physiological changes (the 'feeling'). The physiological emotional response is largely automatic and unconscious, and not only involves changes in endocrine, autonomic and musculoskeletal activity but modulations to executive functions such as arousal, attention, memory and decision making as well (25). Fitting with the dimensional and appraisal models, a meta-analysis of neuroimaging studies has demonstrated that large functional networks, involving several different brain regions and structures, are responsible for emotion processing in the brain. This is opposed to the simple 'one-to-one' mapping of brain structures that is typically associated with the discrete models of emotions (21). Historically, many of these structures involved in emotion processing have been referred to as the limbic system (26). After exposure to an emotionally salient stimulus, limbic structures such as the amygdala, the orbitofrontal cortex (OFC), and the anterior insula integrate the incoming sensory information with any associated memories of the stimulus to evaluate its context, or emotional value. From the amygdala, signals are distributed to the hypothalamus and brainstem, where autonomic and endocrine responses are directed. There are also extensive connections between these limbic structures and parts of the prefrontal cortex (PFC), where higher-level cognitive processes are activated in response to the contextualized emotional stimulus (25,27,28).

It is the activation of the prefrontal cortex in emotion processing that would allow emotion to be exploited as an access pathway for a brain-computer interface, as other limbic structures such as the amygdala and the hypothalamus are located too deep inside the skull for superficial detection by portable brain-imaging modalities (19). The PFC is essential for evaluating the emotional significance of a stimulus, interpreting and regulating emotional experience, and directing subsequent behaviours (29,30). In the most basic interpretation, the PFC can be understood to be evaluating the *core affect* of a stimulus – whether it is rewarding or threatening, and if it should therefore be approached or avoided, or accepted or rejected (28–30). For an extensive review of the role of the PFC in emotion processing, please refer to Dixon et al (29). In sum, the orbitofrontal cortex (OFC) evaluates incoming sensory information and appraises personal episodic memories related to the affective stimulus (28,29). The role of the ventromedial prefrontal cortex (VMPFC) can be described according to its substructures, including the subgenual anterior cingulate cortex (sgACC), which directs autonomic changes in physiological arousal; the pregenual anterior cingulate cortex (pgACC), which interprets these changes in physiological arousal to contribute to subjective 'feelings' of emotions; and the rostromedial prefrontal cortex (RMPFC), which integrates information about the self during emotion processing. Further, the dorsomedial prefrontal cortex (DMPFC) is involved with appraising the mental or emotional states of others, and the anterior-mid cingulate cortex (aMCC) evaluates and directs different behavioural actions during emotional responses (29,30). Finally, the lateral prefrontal cortex is involved in directing emotional regulation, or the conscious manipulation of an emotional response according to a desired goal (29,31,32).

Overall, if the activation of the PFC during emotional responses produces reliable neurophysiological signals that can be detected using a modality such as EEG or NIRS, a brain-computer interface can be developed to detect different emotional responses based on their inherent properties such as valence or arousal. This concept will be explored further in the following sections.

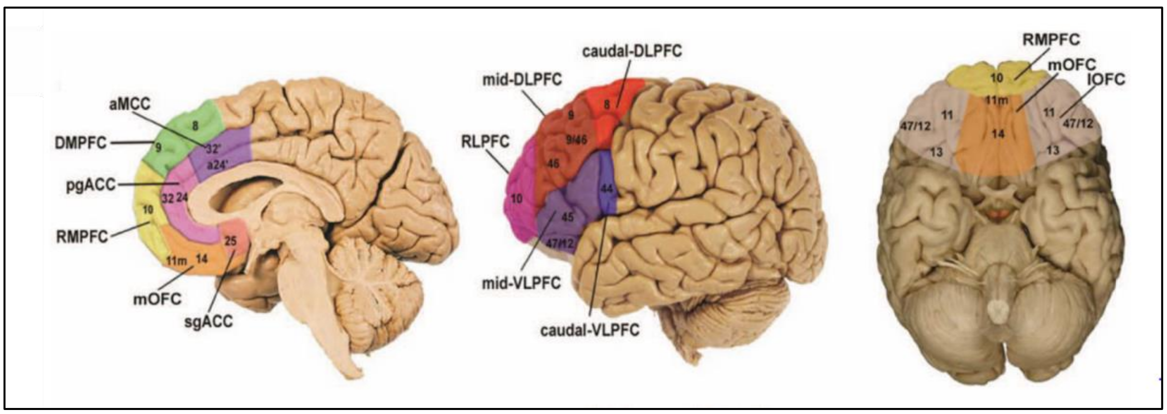


Figure 3: Parcellations of the prefrontal cortex*: aMCC = anterior mid cingulate cortex; DMPFC = dorsomedial prefrontal cortex; pgACC = pregenual anterior cingulate cortex; RMPFC = rostromedial prefrontal cortex; mOFC = medial orbitofrontal cortex; sgACC = subgenual anterior cingulate cortex; VLPFC = ventrolateral prefrontal cortex; RLPFC = rostrolateral prefrontal cortex; DLPFC = dorsolateral prefrontal cortex. Refer to Dixon et al. (29) for a complete review of the roles of these structures in emotion processing.*

1.2.2.2 Affective Brain-Computer Interfaces

Brain-computer interfaces that seek to detect and interpret affect, or emotional state, from neurophysiological signals are known as affective brain-computer interfaces (aBCIs). Affective BCIs have been investigated for the improvement of human-machine interactions (19), for medical and psychological therapies (33), and for monitoring mental workload, attention and fatigue (34). Many of the existing affective BCI studies have attempted to decode emotional states from the EEG in typically-developed adults. However, there is significant variability in the methodology of these studies, from the way emotions are defined and how emotional responses are elicited, to what features of the EEG are extracted and the algorithms used to optimize and classify these features into discrete emotional states. Thus, unsurprisingly, these affective BCI studies report a wide range of results and levels of BCI performance. For a thorough review of affective BCI research prior to 2014, see Mühl et al. (19). Following 2014, much of the work in aBCI research has been towards improving and applying novel machine learning methods. For example, Mehmood et al. (35) used Hjorth parameters as classification features rather than more traditional spectral power band features, and Yano et al. (36) were able to achieve better BCI performance than traditional approaches with their proposed novel *regularized loss minimization* strategy. Atkinson et al. (37) tested different feature selection methods and kernel-based classifiers to address a multi-class emotion identification problem, while Iacoviello et al. (38) developed a paradigm to identify emotional states in real-time. However, many of these novel methods generated a large number of EEG features and subsequently selected the most discriminatory subset via optimization methods but failed to identify the underlying physiological phenomena represented by these features. As such, these approaches had limited generalizability across the population and in fact, introduced the risk of misconstruing artefacts as emotion-related features (39). Although McFarland et al. (39) stressed the importance of using well-defined, physiologically-relevant features such as frontal alpha and frontal midline theta activity of the EEG, they reported that these features yielded less-than-optimal BCI performance when predicting emotional states. The wide variation in methodology of affective EEG-BCI studies and the corresponding uncertainty about the underlying physiological phenomena in the EEG are significant limitations of using EEG as a modality for an affective brain-computer interface. For this reason, this thesis investigated a different modality – near-infrared spectroscopy (NIRS) – to collect the neurophysiological signals underlying emotional responses.

1.2.3 Functional Near-Infrared Spectroscopy

1.2.3.1 Mechanisms of Near-Infrared Spectroscopy

Near-infrared spectroscopy uses light in the near-infrared range (~700-1200nm) to measure the hemodynamic activity of the brain. Near-infrared light is transmitted from a light source (e.g. LED, laser) through the tissues of the head and scalp and is absorbed by oxygenated and deoxygenated hemoglobin in cerebral blood. Oxygenated and deoxygenated hemoglobin (HbO & Hb) are biological chromophores whose spectra change with respect to their oxygenation state, which in turn fluctuates according to the brain's metabolic demands (40). Unabsorbed light is scattered throughout the tissue of the brain, some of which is eventually reflected back out of the head and can be measured by detectors on the scalp.

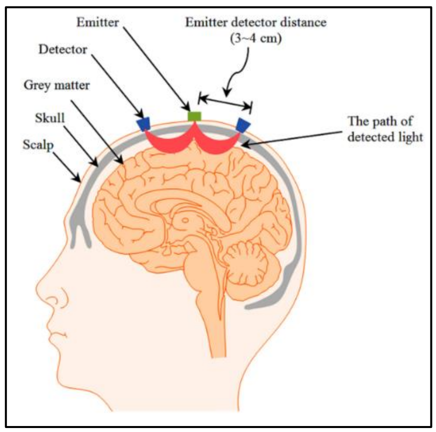


Figure 4: An example of two near-infrared light emitter-detector pairs and the corresponding path of light through the tissues of the head and into the gray matter of the cortical region of the brain. *The light is hypothesized to follow a crescent-shaped path between emitter and detector, and can reach only a few centimeters into the skull, to the upper regions of the cerebral cortex. Modified from Naseer & Hong, (79).*

At time *t,* the amount of λ-wavelength light absorbed is given by a modified version of the Beer-Lambert law:

$$OD(t,\lambda) = -\log_{10}\left(\frac{I(t,\lambda)}{I_0(t,\lambda)}\right) = \textstyle\sum_i \varepsilon_i(\lambda)c_i(t)DPF(\lambda)d + G(\lambda). \qquad (1)$$

where *OD* refers to optical density, or the attenuation of the light, I_o is the emitted light intensity, *I* is the reflected light intensity, ε_i and c_i are respectively the extinction coefficient and concentration of the i^{th} chromophore, *d* is the distance between the source and the detector, *DPF* is a differential path length factor and *G* accounts for the light lost to scattering effects. The equation is summed over all *i* chromophores of interest (41,42). The scattering effect is assumed to be time-invariant, and thus vanishes when calculating the change in optical density:

$$\Delta OD(\Delta t,\lambda) = -\log_{10}\left(\frac{I(t_1,\lambda)}{I(t_0,\lambda)}\right) = \textstyle\sum_i \varepsilon_i(\lambda)\Delta c_i DPF(\lambda)d, \qquad (2)$$

Here, Δc_i represents the temporal change in chromophore concentration. To determine the concentration changes of the two chromophores, HbO and Hb, equation (2) is solved simultaneously at two different wavelengths (41,43):

$$\begin{bmatrix} \Delta[Hhb] \\ \Delta[HbO_2] \end{bmatrix} = (d)^{-1} \begin{bmatrix} \varepsilon_{Hhb,\lambda_1} & \varepsilon_{HbO_2,\lambda_1} \\ \varepsilon_{Hhb,\lambda_2} & \varepsilon_{HbO_2,\lambda_2} \end{bmatrix}^{-1} \begin{bmatrix} \Delta OD(\Delta t, \lambda_1)/DPF(\lambda_1) \\ \Delta OD(\Delta t, \lambda_2)/DPF(\lambda_2) \end{bmatrix} \qquad (3)$$

This change in concentration of HbO and Hb can be related to brain activity. The most typical trend is that neuronal activity in a region of the brain increases the metabolic demands of that area, stimulating an increase of blood flow to the brain, and an increase in blood metabolic rate. The increase in regional blood flow is hypothesized to be greater than the increase in metabolic rate, which results in an overall increase in concentration of HbO and a decrease in concentration of Hb (40–43). Reproducibility studies have indicated that an increase in concentration of HbO is most consistently correlated with neural activation both within and across subjects (43,44). This trend, specifically the decrease in concentration of Hb during neural activation, has also been correlated with the blood-oxygen level dependent (BOLD) response seen in fMRI (44,45).

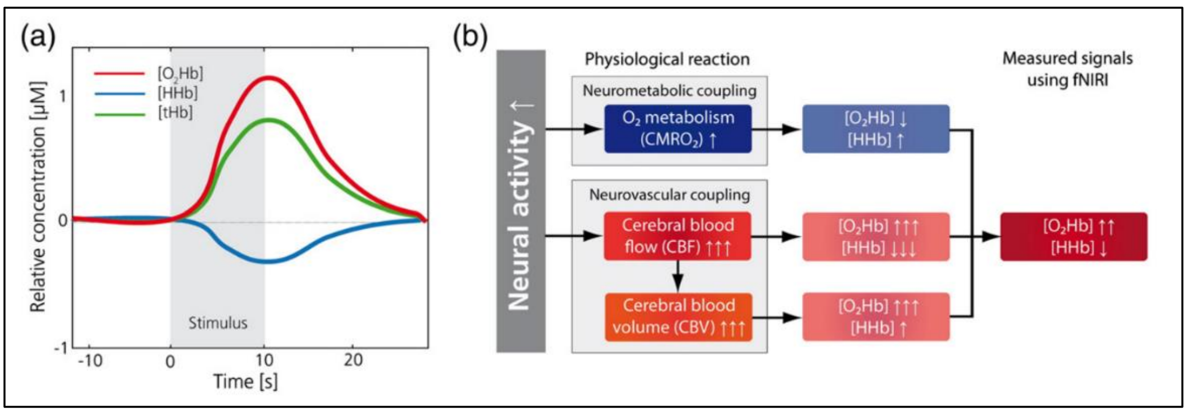


Figure 5: Typical hemodynamic response due to increased brain activity; *a) typical signal of an evoked neuronal hemodynamic responses; and b) an overview of the hemodynamic response to neuronal activity, indicating the increased cerebral blood flow and increased rate of metabolism occurring in the brain during periods of activity, overall resulting in an increase in [HbO] and a decrease in [Hb]. Taken from Scholkmann et al. (41).*

1.2.3.2 fNIRS & the Study of Emotions

Due to its relative low cost, ease of use and robustness to noise, near-infrared spectroscopy has been increasingly used for the functional mapping of brain activity (43,44). Despite the limited imaging depth of fNIRS (about 1-2 centimeters into the scalp), it remains a viable option for investigating the activation of the prefrontal cortex in studies of emotion processing (46). For detailed syntheses on the use of fNIRS to study emotion, please refer to Bendall et al. (46), Balconi & Molteni (47), and Doi et al. (48). Overall, fNIRS has been used to investigate PFC activation both in passive responses to emotional stimuli and in emotional regulation tasks. It has also been used to investigate the influence of emotions on other cognitive tasks, to examine hemispherical differences during emotional processing and to explore differences in PFC activation during emotion processing in clinical populations including major depressive disorder, bipolar disorder, social anxiety disorder and post-traumatic stress disorder (46–48). Due to its non-invasiveness and ease of use, fNIRS has also been extensively used to study emotion processing and the development of emotional networks in children, including infants below 2 years of age (49). Overall, the extensive number of studies employing fNIRS to investigate PFC activation during emotional processing supports its use as a signal acquisition modality for a brain-computer interface to identify different emotional states.

1.2.3.3 NIRS Affective BCIs

There have already been several studies investigating fNIRS as a signal acquisition modality for an affective brain-computer interface. Tai et al. (50) appear to have been the first to attempt single-trial classification of emotional state from hemodynamic activity of the brain. They were able to discriminate emotion-induced brain activation from a baseline state with at least 75% accuracy for all their participants. Hosseini et al. (51) were able to differentiate 'positive' vs 'other' and 'negative' vs 'other' emotional states at average classification accuracies of 72.9% and 68.3% respectively. Both studies used affective images to induce emotional responses in their participants. Moghimi et al. (52) used excerpts of music instead of affective images for their study, theorizing that a dynamically-induced emotional state might better represent a 'real-life' emotional experience. They were able to obtain an average classification accuracy of 72% differentiating between positive and negative responses. Heger et al. (33) investigated how the accuracy of emotion recognition changes over a longer, continuous period rather than a shorter, stimulus time-locked response. They were able to discriminate affective state from neutral at rates significantly above chance for all their participants and found that the recognition rates appeared to stabilize over the length of the trial. This is promising, considering that in real-life applications, emotional responses would be sustained rather than time-locked to a specific stimulus. Yangisawa et al. (53) investigated the activation patterns over the PFC during the viewing of high-valence and low-valence images. They found that the most significant activation occurred over the central portion of the PFC, and there were significant differences in the concentration of HbO and Hb in the two valence conditions. About two-thirds of the participants in their study achieved over 60% classification accuracy, which was closely correlated to individual subjective ratings of the stimuli. Participants with high classification accuracies (highest was 96.7%) rated the stimuli as highly emotionally salient, while participants with the lowest classification accuracies (below 60%) did not experience strong emotional reactions to the stimuli. This indicates that care must be taken to select highly salient stimuli in an affective BCI paradigm, to ensure the emotions are evoked as intended. Hu et al. (54) believed that classifying emotions as 'positive' or 'negative' is an oversimplification of the variation seen in different emotional states. Instead, they investigated three different subsets of positive emotions – encouragement, playfulness and harmony. They were able to differentiate between these different positive emotional states at an average accuracy of 73.8%. Aranyi et al. designed an experiment where users could interact with a virtual character by expressing either positive emotions (55) or anger (56). The 'behaviour' of the virtual character was controlled by the real-time evaluation of asymmetry of the hemodynamic response in the left and right hemispheres of the prefrontal cortex. 11 of 17 participants were able to achieve at least 50% accuracy controlling the character with positive emotions, and 8 of 11 participants achieved at least 50% controlling the character with negative emotions. A table summarizing these affective NIRS-BCI studies can be found in Appendix A.

It is important to note three key distinctions between these eight studies and the work that will be described in the remainder of this thesis. Firstly, six of the eight of these studies used unimodal stimuli (either affective images or music) to evoke emotional responses. Secondly, all of these studies were conducted with a population of typically developed adults. Finally, all but two of these studies were conducted offline, meaning the data were collected, analyzed and BCI performance was evaluated post-hoc. In offline studies, the participant takes more of a passive role as their brain activity is recorded, receiving no meaningful feedback of their task performance or performance of the BCI itself. The importance of neurofeedback in brain-computer interface research will be highlighted in the following section of this literature review.

1.2.4 Self-Regulation and Neurofeedback

1.2.4.1 Emotional Regulation

As suggested by the findings of Yanagisawa et al. (53), it is important to evoke strong emotional responses and consistent brain signals from all participants when developing an affective BCI. This can be a challenge, given the subjectivity of an individual's responses to affective stimuli. This subjectivity can be a result of personality, disposition, experience, biological sex, and even genotype (57). For example, an individual with arachnophobia would react more strongly to a picture of spiders than someone without this phobia. In developing an affective brain-computer interface, this variation could be accounted for by tailoring the stimuli material to each participant and collecting subjective ratings of their emotional responses to each stimulus.

However, there can also be subjectivity in an individual's awareness and understanding of their own feelings, or their *trait emotional awareness* (58). The practice of *emotional regulation* can be used to heighten one's emotional awareness and provide a means of controlling or adjusting an emotional response in an adaptive way, to help meet one's goals (59). Emotional regulation can affect the intensity, time course, quality and type of emotion experienced (60). There are many types of emotional regulation strategies, including *selective attention*, where the individual focuses (or avoids focusing) on salient emotional features of an affective stimulus; *attentional distraction*, where the individual uses a distracting secondary stimulus to draw their attention away from the affective stimulus; *controlled generation*, where the individual uses mental imagery and reasoning to attribute (or mitigate) emotional meaning to a stimulus; and *controlled regulation*, where the individual actively reinterprets the meaning of the stimulus to alter its emotional saliency (59,61,62). These strategies have also been shown to activate regions of the prefrontal cortex involved in emotion processing (62). Using these strategies to intensify emotional responses could be a way to strengthen prefrontal cortex activation, producing more robust signals that could eventually improve the performance of an affective brain-computer interface.

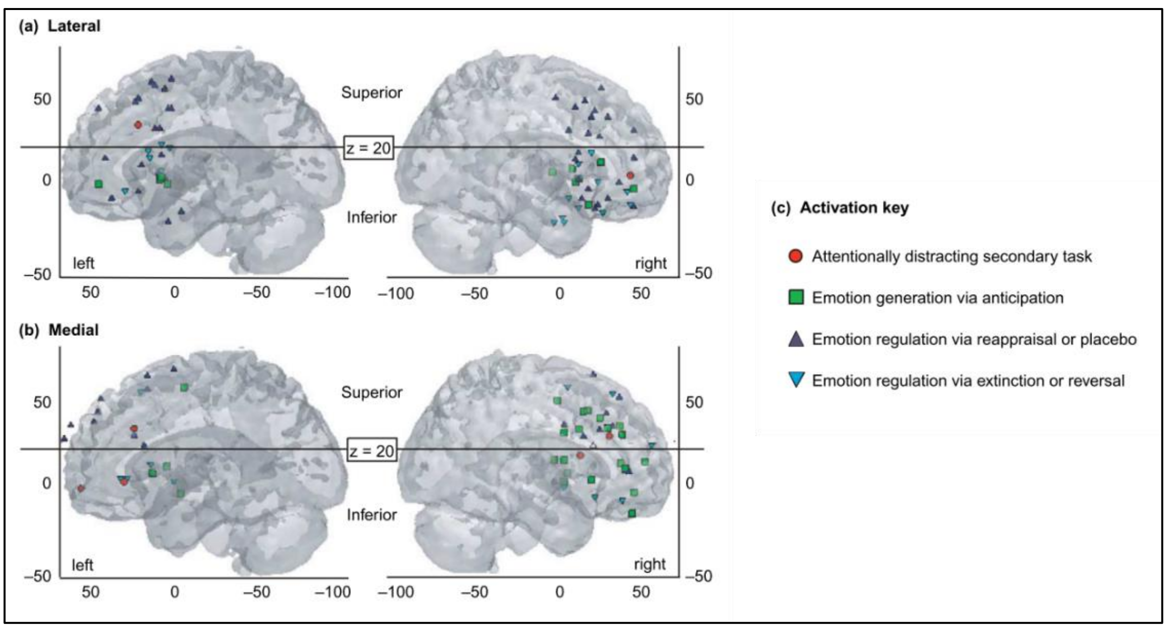


Figure 6: Activation of the PFC in different emotional regulation tasks*. Each point represents an area of activation focus, compiled from Oschner & Gross's review of emotional regulation studies (62). The points are colour and shape-coded for the different tasks (attentional distraction, controlled generation, and controlled regulation through reappraisal and extinction).*

1.2.4.2 Self-Regulation & Neurofeedback

Emotional regulation could be incorporated in an affective BCI by providing the individual with neurofeedback, as in some sensory representation (e.g. visual or auditory) of their real-time emotion-based neural activity. Neurofeedback allows the individual to visualize or have a sense of the 'level' of their brain activity, facilitating the conscious modulation of this activity. Neurofeedback is based on the principles of operant conditioning – changes of brain activity in the desired direction are rewarded, which 'teaches' the brain to repeat that behaviour in the future in the hopes of generating another reward (63). In clinical settings, neurofeedback has been used to train patients to self-regulate their own brain activity in a way that might ameliorate a particular behaviour or 'rewire' certain pathological networks, and has led to long-lasting functional reorganization of the brain (64–66).

Self-regulation has also been used as an access pathway for BCIs. The Birbaumer group (67) pioneered this idea, developing a *Thought Translation Device* that operated on the volitional control of slow cortical potentials (SCPs), which are electrical shifts in the EEG that reflect the priming of the brain's attentional resources. Through several training sessions, participants were taught to produce positive (upward) or negative (downward) shifts in their SCPs, visually displayed as the height of a cursor on the user interface. In trials where the participant was successful in keeping their SCP above a certain threshold for a specified duration, they were 'rewarded' with a reinforcing image. Once able to consistently control their SCPs, participants could use this approach to control a cursor, a spelling device, and even a web browsing application (68). Grosse-Wentrup et al. (69) similarly trained participants using neurofeedback to operate a BCI using self-regulated shifts in the gamma-band oscillations of the EEG. Also, Weyand et al. (70) demonstrated that participants could gradually transition from using mental tasks to self-regulation to adjust the hemodynamic activity of their prefrontal cortex.

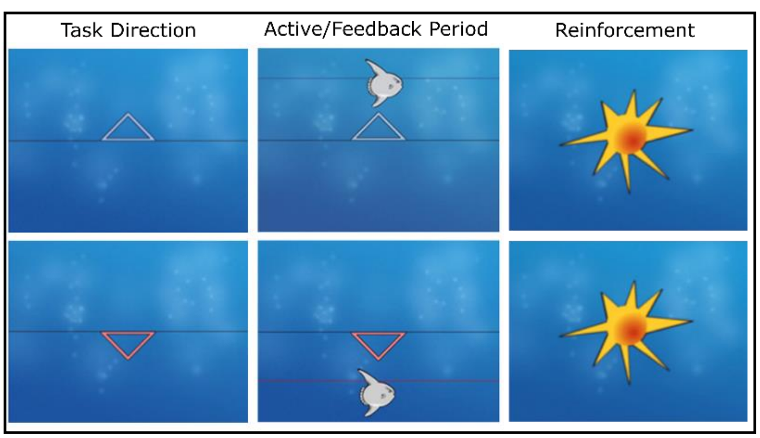


Figure 7: An example of a neurofeedback interface for self-regulation training. *The height of the icon (the fish) on the screen fluctuates with the user's brain activity. For a given trial, the participant is direct to either move the fish up or down and maintain its height above or below the threshold line. If the participant can do this for a certain period of time, a reward in the form of an encouraging graphic is displayed. Modified from Mayer et al. (66).*

The success of these studies suggests that it would be highly beneficial to incorporate self-regulation and neurofeedback in future BCI paradigms, especially one whose access pathway (i.e. emotional state) has been shown to be highly sensitive to conscious manipulation (i.e. emotional regulation strategies).

1.2.5 Neurodevelopment of Emotion

It is well known that higher-order processing areas of the brain, such as the prefrontal cortex, are some of the last neural structures to fully mature (71). When developing a brain-computer interface for children, it is important to consider whether the evoked neurophysiological responses typically seen in adults will also be present in the younger target population. The final section of this literature review will describe the role of emotions in childhood and how emotional awareness and processing develops from infancy to adolescence.

Emotions are essentially adaptive behavioural mechanisms that that help us adjust our behaviour in response to incoming sensory information to achieve our goals. This adaptive role of emotional processing, in its most rudimentary form, is present already early in infancy, even before any communication skills are developed (e.g. babies crying in response to hunger, stimulation or changes in their environment) (72). Emotional awareness and regulation also play a key part in childhood in the development of social competence and the formation of relationships (73).

During infancy and early childhood, emotions are regulated externally by caregivers and other adults. Caregivers soothe and manage emotional distress and arousal, monitor sensory input, and model socially appropriate expressions of emotion (72–74). Children attend to the emotional cues of their caregivers and begin to learn how to regulate their own emotional states. This increasing amount of self-control is facilitated by developing cognitive abilities, language skills, increasing emotional awareness and understanding, and a heightened sense of self that becomes more refined with age (72–74). By the preschool age, children have developed an understanding of which basic emotions will be evoked in specific situations, know the consequences of different emotional responses, and can employ some simple emotional regulation strategies. By grade-school, children have developed a deeper understanding of emotion and can rely on a variety of strategies to regulate their emotions successfully and independently without much external assistance. This process continues to mature with adolescence, where a more sophisticated sense of self is established and contributes to the understanding of emotions as personalized experiences that are evoked and expressed differently among individuals (72,73).

These developments in emotional awareness, understanding and regulation are paralleled by the functional neural development of the cortex throughout childhood. It is believed that in early infancy, emotions are directed mainly by *cortical excitatory processes* (e.g. the cause of sudden outbursts of crying) associated with sympathetic and parasympathetic regulation (72). These processes decline as *cortical inhibitory controls* emerge with the development of the frontal lobe and other cortical networks that facilitate more complex cognitive abilities (71,74,75). These areas associated with higher-order cognitive abilities and executive functioning, such as the prefrontal cortex, start to develop within the first year and continue to mature throughout childhood (71). This maturation process involves a reduction in neuronal density, synaptogenesis, branching of dendrites and increased myelination (71,76). Overall, these areas become more fractionated, developing into specialized functional networks that can carry out the complex cognitive tasks that are required for a comprehensive awareness and understanding of emotions and effective emotional regulation (76).

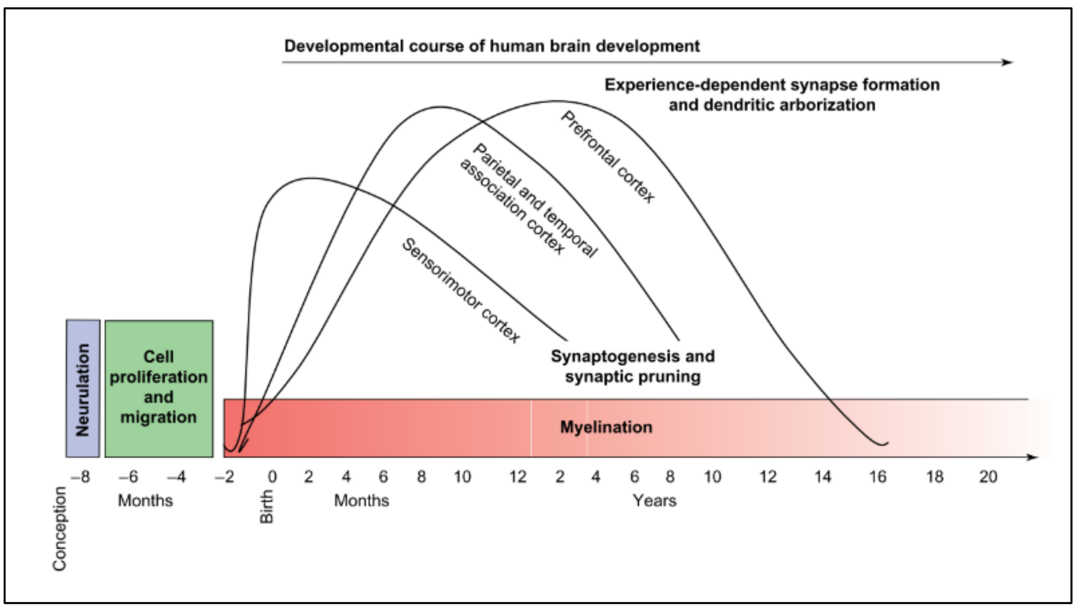


Figure 8: Human brain development throughout childhood. *The brain reaches 90% of its full adult size by approx. 6 years of age but continues to change dynamically throughout childhood and adolescence. Development is a combination of progressive (e.g. synaptogenesis) and regressive (e.g. synaptic pruning) changes. The prefrontal cortex is one of the last structures to develop and fully mature. Taken from Casey et al. (71).*

The early emergence of emotional awareness and regulation abilities suggest that emotion can be used as an access pathway for a brain-computer interface for children. Following the maturation and refinement of emotion processing and other cognitive systems with age, it is possible that older children may have more robust changes in their hemodynamic activity in response to emotional stimuli and thus we may see a trend of increasing BCI success with age.

1.3 Project Overview & Research Objectives

1.3.1 Project Overview

The primary goal of this research was to develop a brain-computer interface that could eventually provide children with severe physical disabilities with an alternative means of communication. The ability to express emotion is an essential part of communication. An awareness and understanding of emotion starts to develop early in childhood, making it a viable option for a BCI paradigm for young children. The BCI developed in this work used near-infrared spectroscopy to decode positive and negative emotional states from changes in the hemodynamic activity of the prefrontal cortex, an essential component of the brain's emotional circuitry. The BCI also incorporated emotional regulation through visual neurofeedback to heighten awareness and attention of the emotional response, thereby increasing signal robustness and increasing the likelihood of achieving above-chance classification accuracy of emotional states.

1.3.2 Research Questions & Aims

The main questions that were addressed by this research are as follows:

> Q1 – Can positive and negative emotional valence be differentiated in children aged 8-14 from changes in hemodynamic activity of the prefrontal cortex using an NIRS-based BCI?
>
> Q2 – Can above-chance accuracy of emotional valence classification be achieved in an online, real-time implementation of the proposed BCI?
>
> Q3 – What level of interparticipant and intraparticipant variability exists in the hemodynamic responses during evoked emotional states on different days and what impact does this have on BCI performance?

Three specific aims were drawn from these research questions:

> Aim 1: Develop, train and test an NIRS-based BCI to classify changes in hemodynamic activity evoked in response to the presentation of positively- and negatively-valenced emotional stimuli.
>
> Aim 2: Test the BCI online, incorporating neurofeedback and emotional regulation strategies to facilitate online classification.
>
> Aim 3: Analyze differences in the hemodynamic response and the corresponding impact on BCI performance between participants and for each participant on different session days.

1.3.3 Thesis Outline

This first chapter outlined the motivation behind this project and reviewed the relevant literature. The second chapter describes the specific methodology that was used to conduct this research. The third chapter presents the results, the fourth chapter provides a discussion of these results, and the fifth chapter summarizes the contributions of this thesis. The final chapter is the journal article produced from this research, intended for publication in a journal such as the Journal of Neural Engineering or NeuroImage.

2. Methodology

2.1 Participants

10 typically developing children between the ages of 8 and 14 years old (mean age 11.5 ± 1.8 years, 3 males) participated in this study. The participants were recruited from families, staff and volunteers at the Holland Bloorview Kids Rehabilitation Hospital and the neighboring community. Based on a caregiver-reported assessment, participants were screened for any neurological, psychological, cardiovascular, respiratory or drug or alcohol related conditions, as well as for any history of trauma-related brain injury. Due to the emotional nature of the stimuli, participants were also screened for any history of emotional trauma. Written consent was obtained from all participants' caregivers and assent was obtained from all participants prior to the start of data collection.

2.2 Study Design

The research study took place in two parts, in accordance with the first two research aims.

Phase 1 (Aim I): Offline BCI Training

The first part of the study involved collecting data to train the BCI to identify the valence of evoked emotional responses. Each participant attended a single session where they underwent a series of emotion-induction trials. On each trial they were presented with a set of emotionally salient auditory and visual stimuli while their hemodynamic activity was measured using near-infrared spectroscopy. Participants were instructed to let themselves respond naturally to the stimuli and to try to maintain an awareness of the emotions they experienced as the stimuli were presented. The stimuli included an equal number of positively- and negatively-valenced sets, whose presentation was counterbalanced and randomized, and punctuated with a short rest period. Refer to section 2.4.2 for more details on the stimuli. The collected hemodynamic signals from the emotion-induction trials were labeled and then used to train and test a classifier to recognize positive and negative emotional responses.

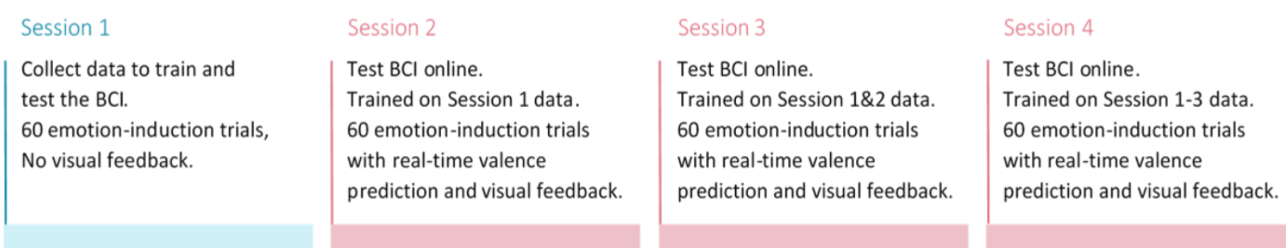


Figure 93: Overview of the study design.

Phase 2 (Aim II): Online BCI Testing

After the BCI was trained on the data from the first session, the participants were invited back for three more sessions, each about a month apart, to test the BCI online. In these sessions, each participant again underwent a series of emotion-induction trials. In this case, they were also presented with visual feedback representing their emotional state as predicted by real-time classification results of the BCI. The participants were instructed to monitor the feedback and use emotional regulation strategies such as selective attention or controlled generation (see section 1.2.4a) to strengthen their emotional response. The feedback was updated every 2s, allowing the participants to visualize how successful they were in regulating their emotions. Again, the presentation of positive and negative stimuli was counterbalanced and randomized. Section 2.4.4 elaborates more on the visual feedback and online classification paradigm.

2.3 Instrumentation & Data Collection

Cerebral hemodynamic activity was measured using the Hitachi ETG-4000 NIRS system (Hitachi Medical Systems, Tokyo, Japan). A 3x5 grid of 8 light emitters and 7 light detectors was secured over the prefrontal cortex using a custom-made headband, with the bottom row of optodes sitting just above the eyebrows and centered at the nose. A diagram of the arrangement can be seen below. Detectors 11 and 12 (figure 11) were approximately aligned with the Fp1 and Fp2 sites of the 10-20 International System of electrode placement (77,78). Each emitter and detector were separated by 3cm, corresponding to a measurement depth of 2-3cm (79–81), reaching the cortical surface (82). This arrangement resulted in 22 integrated measurement sites, as indicated in figure 11. Each emitter contains two laser diodes

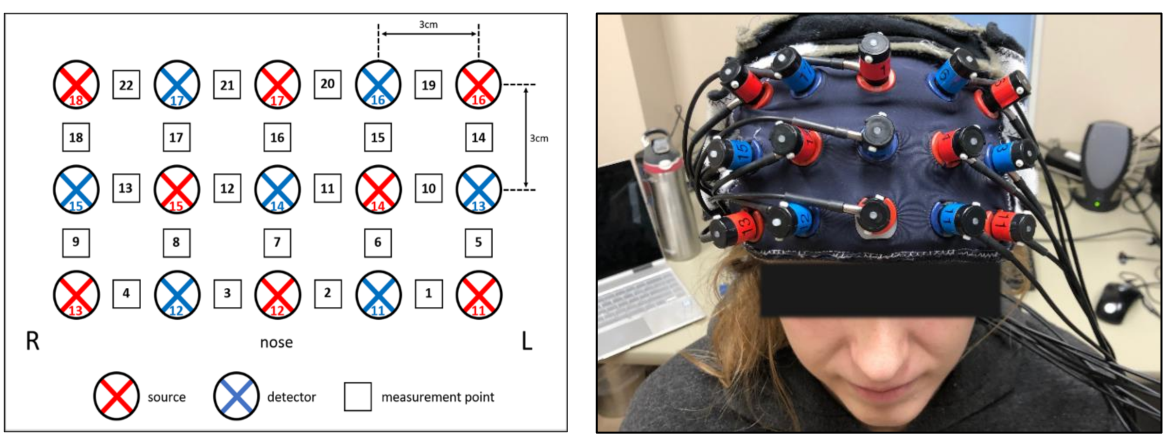


Figure 40: NIRS optode configuration over the prefrontal cortex. *The diagram on the left shows the configuration of the eight light sources (red) and eight light detectors (blue) arranged in a 3x5 grid over the forehead, resulting in 22 measurement sites, indicated by the black and white squares. L and R corresponding to the left and right sides of the forehead, respectively. The image on the right shows the optodes, mounted in the headpiece and placed over the forehead.*

simultaneously emitting light at two different wavelengths (690nm and 830nm). Data were sampled at 10Hz.

2.4 Experimental Protocol

2.4.1 Session Structure

A single session was composed of five blocks of emotion-induction trials. At the start of each block, a 30-second baseline recording was taken, during which the participant was instructed to relax and look at a fixation point on the screen. Then, for each trial, a set of emotional stimuli was presented for a 20s response period (83). The set of stimuli used for a single trial was matched for valence and arousal. A prompt on the screen labeled each trial as ‘positive’ or ‘negative’, confirming for the participants the intended valence for each trial. For the first session, the participants were instructed to respond naturally to the stimuli. In sessions two through four, visual feedback was presented during the response period and the participants were instructed to use the feedback as a guide to regulate the strength of their emotional response. Each active response period was followed by a 20s rest period, allowing the hemodynamic activity to return to baseline levels (77,78,84–87). During the rest period, a set of emotionally-neutral visual and auditory stimulus was presented as a control task (52).There were twelve trials within one block, for a total of 60 trials within one session. When a block was completed, participants would rate the intensity of the emotions they felt over the last set of trials and then self-select when to proceed, allowing for a break in between blocks to reduce mental fatigue. Each session took approximately 40 to 50 minutes to complete.

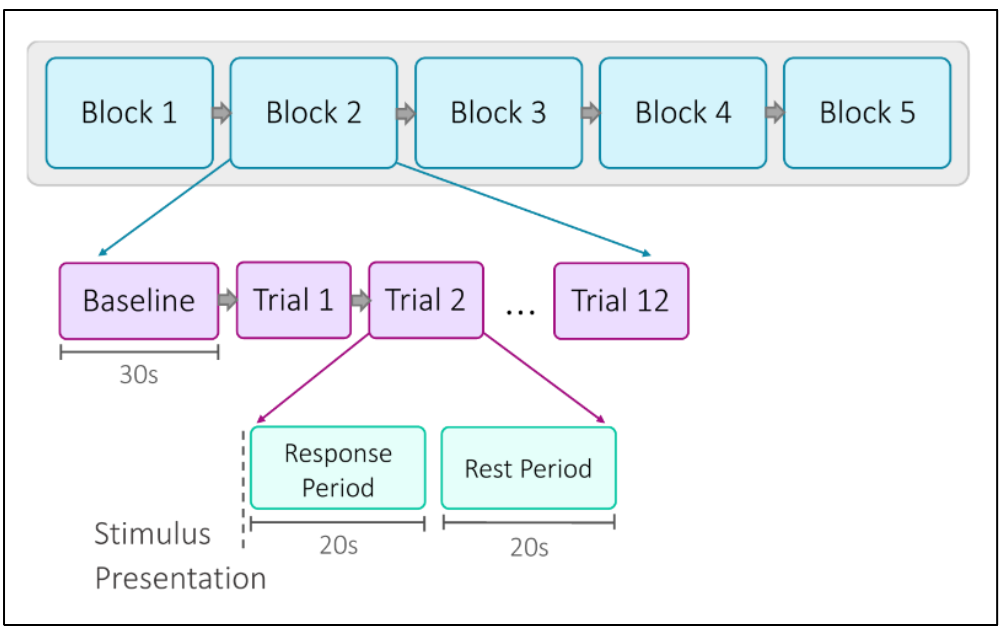


Figure 51: Overview of session structure.

2.4.2 Affective Stimuli

To induce emotional responses, a set of bimodal affective stimuli was presented for each trial. It has been shown that bimodal stimulation can enhance brain activation in processing emotional responses compared to unimodal stimulation (88,89). The set included both visual and auditory stimuli presented simultaneously. The visual stimuli consisted of pictures drawn from several standardized databases: the International Affective Pictures System (IAPS), the Geneva Affective Pictures Database (GAPED) and the Open Affective Standardized Image Set (OASIS). These databases are collections of colour photographs from a wide range of semantic categories that have been rated to be reproducible on their affective quality on scales of valence and arousal (90–92). Several studies have confirmed that affective images from these standardized databases can be used to evoke emotional responses reliably in preadolescent children (93–95). A subset of images was chosen from each database and then reviewed with a parent representative from the Family Leadership program at Holland Bloorview to ensure that the content in the selected images was appropriate for the target age group. The final set consisted of 100 images in each of four categories; high valence and high arousal, high valence and low arousal, low valence and high arousal, and low valence and low arousal.

The auditory stimuli consisted of 20s excerpts of music. Music has been shown to evoke strong emotional responses in listeners (96) and activate brain regions implicated in emotion processing (97). The excerpts of music were chosen from a wide range of genres and were sampled without lyrics to reduce potential brain activation due to mental singing. The music excerpts included modern songs from genres that would familiar to the target population to ensure they would be highly salient (e.g. pop music rather than classical music). The excerpts were rated by the researcher for valence and arousal based on their tempo and mode, which has been shown to be used by pre-adolescent children to distinguish affect in music (98,99). The selected music samples were reviewed by a music therapist at Holland Bloorview to confirm the researcher’s ratings of valence.

One 20s musical excerpt and 5 affective images comprised a stimulus set for a single trial. For each trial, the 5 images would be displayed sequentially for 4s each, while the musical excerpt played. The music and 5 images for a single trial were matched for their levels of valence and arousal. Either a positively- or negatively-valenced set of stimuli was presented for each trial, in a counter-balanced and pseudo-randomized order. For the 20s rest period in between trials, a clip of brown noise was played and 5 emotionally-neutral images (also selected from the databases) were displayed as a neutral control stimulus (52).

2.4.3 Stimuli Personalization

Images from the IAPS, GAPED and OASIS databases have been rated on their valence and arousal by averaging the personal ratings of thousands of individuals. While this gives a good indication of the overall affective quality of each image, these ratings can still vary significantly between individuals, especially when personal preferences, experiences and contexts are considered. For this reason, at the start of the study the participants were asked to go through the stimulus set and confirm the affective ratings. The images from the chosen stimulus set were displayed to each participant in the form of a slideshow, and upon presentation of each image, the participant was instructed to rate it as either positive or negative by pressing a button on the keyboard. Each image was displayed for only 4s to ensure the participant was rating the image based on their initial reaction. A screenshot of the interface used to display and rate the images can be seen below in figure 13. Any image that did not match the standardized rating was discarded from that participant's stimulus set. In this way, a personalized stimulus set was developed for each participant, helping to ensure that the emotions evoked would be of the intended valence.

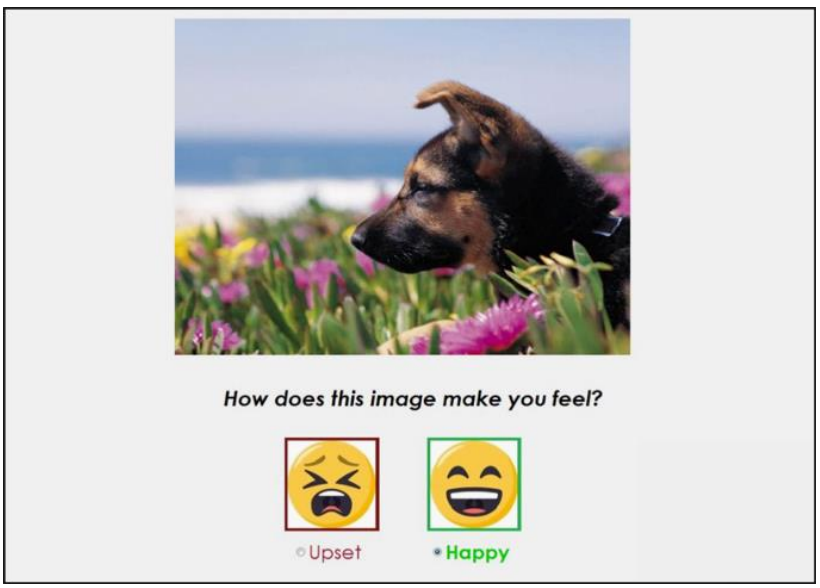


Figure 62: Screenshot of the interface used to rate affective stimuli. *Images from the stimulus set would appear on the screen for 4s, and participants could select either 'happy' or 'upset' in response to the question, 'How does this image make you feel?'. Any missed image or any image that did not match the standardized rating was discarded from that participant's stimulus set.*

2.4.4 Visual Feedback & Online Classification Paradigm

During the online sessions visual feedback was displayed to the participants so they could monitor the strength of their emotional responses and use emotional regulation strategies to adjust the intensity of their emotions. The feedback was in the form of a vertical bar that filled with colour according to the predicted valence of the response (see figure 14). The top of the vertical bar represented a strong *positive* valence, and the bottom of the vertical bar represented a strong *negative* valence. Each trial started with the feedback bar at a neutral middle position. The feedback bar would increase in height, filling with a green colour, if the predicted emotional valence was positive and would decrease in height, filling with a red colour, if the predicted emotional valence was negative. The participants were instructed to try to increase or decrease the height of this bar as much as possible over the 20s trial response period, according to the valence of emotion they experienced for that trial.

Figure 13: Screenshots of the experiment interface. *The top row shows three screenshots, taken at 4s, 12s and 20s respectively, for a positive trial; the middle row shows a neutral rest period; and the bottom row shows a negative trial. Each screenshot shows the displayed image at that time in the trial, the valence prompt above it, and the feedback bar on the right. The positive trial shows the feedback bar increasing in valence, and the negative trial shows the feedback bar decreasing in valence. The neutral rest period has no visual feedback.*

The height of the feedback bar was determined using the real-time classification output from the BCI. The BCI would process, analyze and classify 2-second segments of incoming data from the current trial response period and update the height of the feedback bar based on the classification output. The specific height of the feedback reflected the probability that the incoming segment of the hemodynamic signal belonged to either the positive or negative valence class. The BCI would analyze each 2-second segment cumulatively; that is, the first segment would be the first two seconds of the trial response period, the second segment would be the first four seconds, and so on until the classification of the entire 20s hemodynamic signal from that trial. This final classification output, based on the entire 20s of the trial, was used to evaluate the BCI's performance for that trial (e.g. a correct classification was tallied when the BCI-predicted classification of the 20s signal matched the valence of the stimuli for that trial).

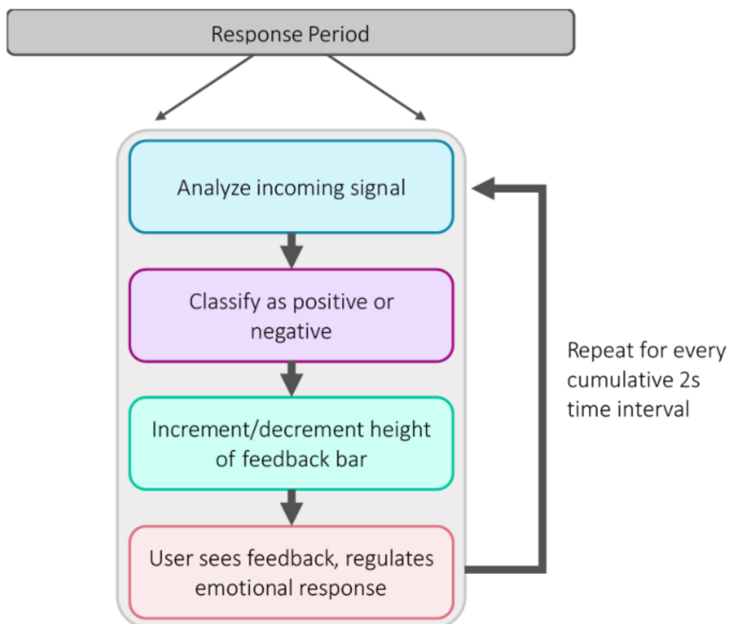


Figure 74: Real-time emotional valence prediction paradigm for generating visual feedback. *During the response period, 2s of the incoming signal is analyzed and classified as either positive or negative. The height of the feedback bar is incremented according to the results of the classification. The user sees this feedback and regulates their emotional response accordingly. Subsequently, data from the next 2s time interval are appended to the signal (e.g. the second segment is 4s long, the third is 6s, and so on) and classification and feedback provision iterate as above.*

2.4.5 Subjective Emotional Intensity Rating

At the end of each block, the participants were asked to rate the intensity, or strength, of the emotions they experienced for that set of trials using a modified version of the Self-Assessment Manikin (SAM). The SAM is a visual tool that allows for the rating of an emotional experience based on three axes – valence, arousal and dominance on multiple-point scales (100). When using this method with children, it has been suggested to that the SAM be modified to have fewer rating levels (e.g. 5 or 7 scale points rather than 9, which is commonly used for adults) and to use highly relatable visual representations for these options (e.g. emoji images rather than words or diagrams) (101). For this study, a modified SAM was used with 5 emoji images depicting a range of faces from low to high intensity (figure 16). The modified SAM was displayed on the screen at the end of each block, and the participant selected their desired rating before moving on to the next block.

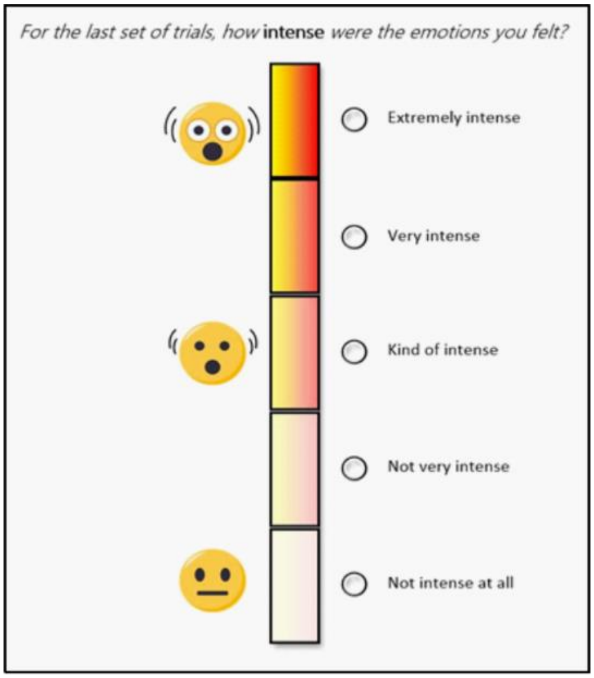


Figure 15: Modified self-assessment manikin (SAM) for rating the intensity of each block of emotion-induction trials. *Participants could choose from one of five options, from 'not intense at all' to 'extremely intense'.*

2.4.6 End-of-Session Subjective Experience Questionnaire

At the end of each session, the participants were asked to answer a short questionnaire about their subjective experience. The questionnaire included questions on their mood, alertness, perceived effort and frustration experienced throughout the session. The results from this questionnaire were used to qualitatively evaluate the intersession variability for each participant. The questionnaire can be found in Appendix B.

2.4.7 Temperament Questionnaire

The participants' caregivers were also asked to complete a temperament assessment questionnaire for their child. Temperament, a fundamental part of one's personality, refers to how an individual reacts to their environment within emotional, physical and attentional domains (102). An individual's temperament can affect how they experience and regulate emotions. In this study, differences in temperament were used to investigate differences in participant performance using the affective BCI. The parent-version of the validated Early Adolescent Temperament Questionnaire (EATQ), developed by Capaldi and Rothbart (103,104), was used to assess the participants' temperaments using 10 sub-scales: activation control, affiliation, aggression, attention, depressive mood, fear, frustration, inhibitory control, shyness and surgency. A description of these measures can be seen in Table 1. These 10 sub-scales can also be consolidated into 3 'super-measures': effortful control, which combines activation control, attention and inhibitory control; negative emotionality, which combines aggression, frustration and depressive mood; and extraversion, which combines shyness and fear, reverse-scored, as well as surgency (103). The EATQ was found to have acceptable levels of reliability and stability (105). A copy of the EATQ and its assessment metrics can be found in Appendix C.

***Table 1**: **Descriptions of temperament measures in the Early Adolescent Temperament Questionnaire (EATQ).** Definitions taken from Capaldi & Rothbart (103).*

Measure	Definition
Activation Control	"Ability to perform an action when there is a strong tendency to avoid it."
Affiliation	"The desire for warmth and closeness with others."
Attention	"The ability to focus and shift attention when desired."
Fear	"The tendency to experience unpleasant affect related to anticipation of distress."
Frustration	"The tendency to experience unpleasant affect related to the interruption of ongoing tasks or goal-blocking."
Inhibitory Control	"The ability to plan, and to suppress inappropriate responses."
Shyness	"Behavioral inhibition to novelty and challenge, especially social."
Surgency	"The pleasure derived from activities involving high intensity or novelty."
Aggression	"Hostile and aggressive actions, including person- and object-directed physical violence, direct and indirect verbal aggression, and hostile reactivity."
Depressive Mood	"Unpleasant affect and lowered mood, loss of enjoyment and interest in activities."

2.5 Data Analysis

Data analysis for an NIRS brain-computer interface refers to the steps required to recognize patterns in the acquired hemodynamic signal. These steps involve processing the acquired signal to remove unwanted artefacts, extracting discriminatory features, and finally classifying the signal on the basis of these features (15). These steps will be described further in the sections below and are summarized in figure 17.

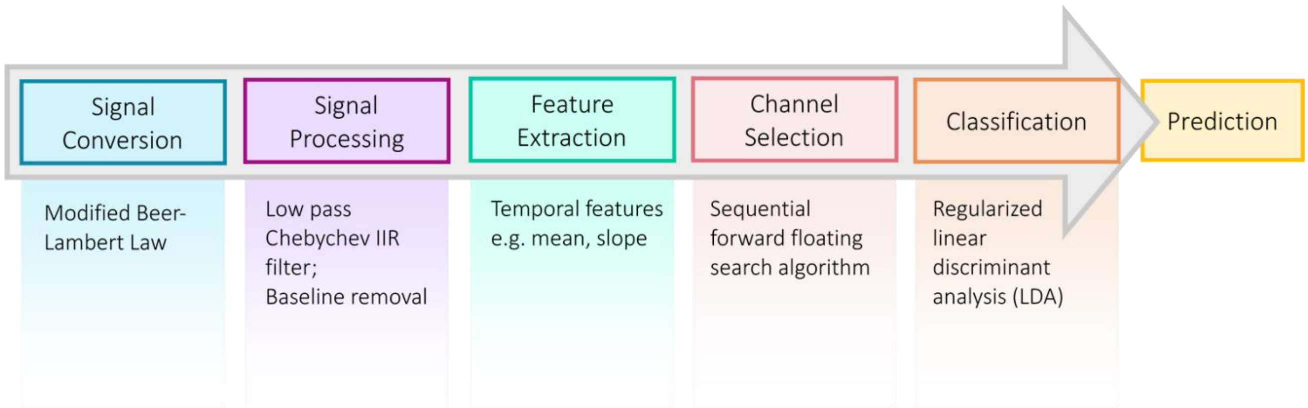


***Figure 86: Data analysis pipeline** showing the processing steps for the incoming raw NIRS signal, ultimately yielding a prediction of the associated emotional valence after the classification step.*

2.5.1 Signal Preprocessing

The Hitachi NIRS System used in this study converts the changes in light intensity for each source-detector pair to changes in concentration of hemoglobin using the modified Beer-Lambert Law (refer to section 1.2.3). Within the acquired hemodynamic signal, there are several contaminating spontaneous oscillations of physiological origin that can produce variability that is not task or stimulus related. This physiological noise can usually be filtered by removing their respective frequency components from the acquired signal. Cardiac activity (0.8-1.2Hz), respiration (0.2-0.4Hz), and Mayer waves, or fluctuations due to arteriole pulsations (0.1Hz), are all physiological phenomena producing noise with frequency content higher than that of the hemodynamic response of interest (41,79,106). These effects were removed using a 3rd order type II Chebyshev low-pass infinite impulse response filter with a passband of 0-0.1Hz, transition band of 0.1-0.5Hz and a stopband cut off frequency of 0.5Hz, with a ripple of 0.1db and minimum attenuation of 50dB (77,78,84–87).

2.5.2 Baseline Removal

The baseline, or resting state hemodynamic signal, can vary on a day-to-day basis and can even fluctuate throughout a session due to system drift and minor shifts in optode positions (107,108). For this reason, a 30s baseline recording was taken at the beginning of each block to compensate for any potential baseline fluctuations. The mean of the 30s baseline recording was calculated and subtracted from the signal of each subsequent trial in that block (85).

2.5.3 Feature Extraction & Selection

2.5.3.1 Feature Types

To differentiate classes of emotional valence, a set of discriminatory features must be extracted from the acquired hemodynamic signals. Temporal features that capture the morphology of the signal are commonly used in NIRS-BCI studies, with signal mean or signal slope being used in approximately half of the reported studies (79). These features can be calculated for each individual measurement site/channel, for each chromophore of interest (HbO and Hb) and for every defined time interval. Time-frequency features from wavelet decomposition have also been used in NIRS-BCI studies, however, were found to offer only marginal value when used in combination with temporal features compared to temporal features alone (33,50). Previous affective NIRS-BCI studies have also considered laterality features that compare differences in the hemodynamic signal between the right and left sides of the prefrontal cortex, to exploit potential hemispherical differences (52). However, preliminary analysis for the present study showed that laterality features were not useful in classifying emotional valence. Instead, seven temporal signal features were investigated for use with the BCI – mean, slope, moving slope, variance, root mean squared (RMS), skewness, and kurtosis. These features were all calculated over the entire 20s trial response period.

2.5.3.2 Feature Selection

Considering multiple possible feature types, multiple chromophores and multiple measurement channels generates a high-dimensional feature set (7 feature types x 2 chromophores x 22 channels = 308 features). It is likely that such a high dimensional feature set contains redundant information, which is not helpful in classification problems and can in fact be detrimental to classifier performance (15,109–111). It is therefore necessary to select the optimal subset of features that provides the most discriminatory information about the acquired signal. In this study, dimensionality reduction of the feature set was conducted in two steps. First, based on the offline data collected from session one, BCI performance was evaluated using each of the seven possible feature types for each participant and the feature type that yielded the highest performance was chosen as the most discriminatory feature for that participant. Secondly, a channel selection procedure was conducted to determine the 5 best channels for each chromophore, resulting in a set of 10 features. The size of this feature set (p = 10 features) was chosen based on preliminary analyses. The best 5 channels for each chromophore were selected using a sequential forward floating search (SFFS) algorithm, which is recommended for classification problems with modest data sets (109) and has been used successfully in several NIRS-BCI studies (77,78,86). The SFFS algorithm (figure 18) systematically searches through the available features to create an optimal subset, adding the best *l* features and removing the worst *r* features each iteration based on a fitness criterion (15,112). The Fisher criterion was used as the fitness criterion; refer to refs. (77,86,113) for more information on this method.

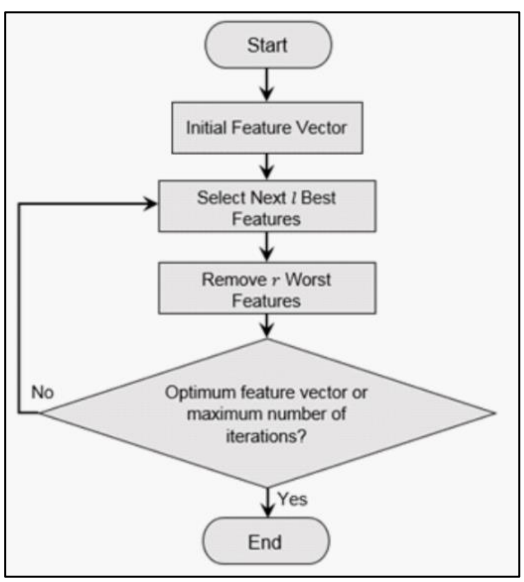


***Figure 9: The sequential forward floating search algorithm**. Adapted from Chandrashekar et al. (150).*

2.5.4 Classification

A linear discriminant analysis classifier (LDA) was used to discriminate the acquired hemodynamic signals based on their emotional valence. LDA is a commonly used classification algorithm in online BCI work due to its speed and low computational cost (15). LDA involves defining a linear decision boundary that separates the data into two classes, maximizing the distance between the class means while minimizing the variance within each class (15,79). To further mitigate the effects of a high-dimensional feature set with a small training dataset ($p = 10$ features, $n = 60$ training samples), a regularized version of LDA was used (110,114). Regularized LDA involves defining a penalty parameter, gamma (γ), to reduce the high variance of models fitted to small training sets, thereby yielding a more generalizable model (114,115). Please refer to Rezazadeh et al. (85) for more details on the regularized LDA method. The regularization parameter can be a value between 0 and 1, with values closer to 1 providing higher generalizability, and values closer to 0 approximating traditional LDA. Potential values of gamma (from 0 to 1 in 0.05 increments) were assessed through 10-fold cross-validation, and the value of γ yielding the highest classification accuracy was selected and used in the subsequent classifier model.

2.5.5 Online Classifier Retraining

A regularized LDA classifier was trained for each participant using their most discriminatory feature and 5 best channels of each chromophore based on the data collected from session 1. This classifier was then used to predict, in real-time, the emotional valence of the incoming hemodynamic signals during the next online session. Due to potential intersession variability in hemodynamic activity, it has been suggested that same-day training data can improve classifier performance (85,107,108). For this reason, after each block, the classifier was retrained to incorporate the new data from the current session. Retraining involved running all the collected data (from both the previous and the current session) through the data analysis pipeline, selecting 5 new 'best channels' for each chromophore, determining a new regularization parameter, and then training a new LDA classifier. This new classifier would then be used to predict the emotional valence of the incoming hemodynamic signals for the subsequent online block of trials. For each new online session, data from all previous sessions were combined to train the classifier.

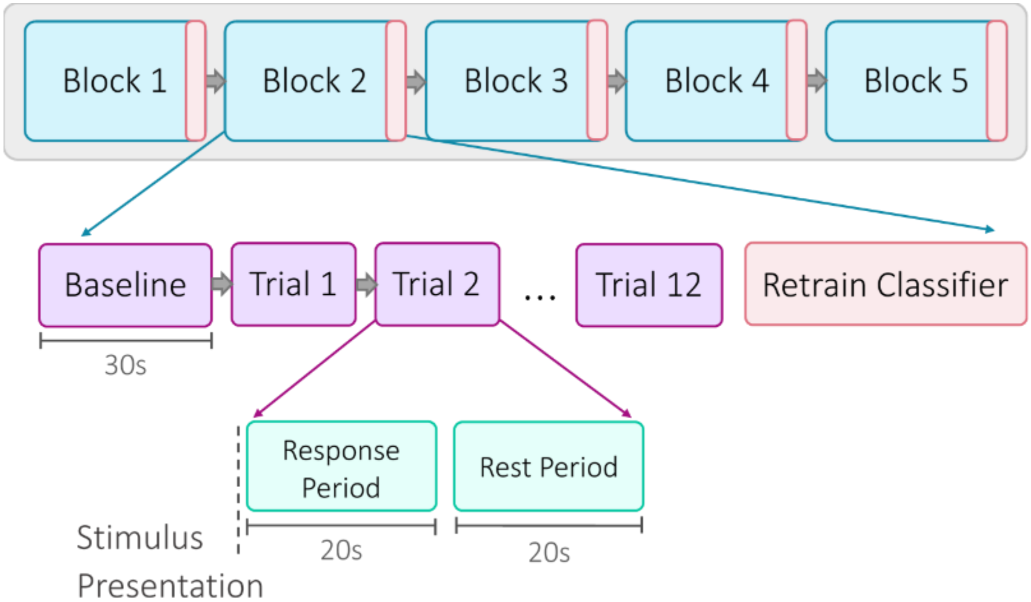


Figure 10: Overview of session structure for online sessions. *The classifier is retrained after each block of 12 trials. Classifier retraining involves selecting five new 'best channels' using SFFS for each chromophore, determining a new regularization parameter (γ) and training a new LDA classifier based on the accumulation of data from previous and current sessions.*

3. Results

3.1 Session 1 Offline Results

The first aim of this thesis was to develop, train and test a brain-computer interface to determine the feasibility of identifying emotional valence in children from their hemodynamic activity. Data for training the BCI were collected from a single session where participants underwent a series of emotion-induction trials. The collected data were analyzed post-hoc using the data analysis pipeline described in section 2.5, investigating seven possible feature types. BCI performance was evaluated based on its classification accuracy, or the percentage of correct predictions out of all the predictions made. Classification accuracy was calculated using 10 iterations of 10-fold cross-validation, where the data set was randomly partitioned into 10 folds, with 9 being used to train the classifier model (n=54 training samples for 1 session) and the 10th used to test the model (n = 6 test samples for 1 session). This process was repeated 10 times, using each possible combination of the training/test set. Ten iterations of this procedure were conducted, using new partitions of the data set each time. All 100 accuracies were averaged to produce an overall measure of BCI performance (78,116). The average cross-validated classification accuracies for each participant, for each investigated feature type, from this first session can be seen in table 2. The most discriminatory feature, the feature that yielded the highest classification accuracy, for each participant is bolded and accuracies exceeding the upper limit of the 95%, 99% and 99.9% confidence intervals of chance are marked with *, ** and ***, respectively. All chance levels were calculated using the binomial distribution (117).

Table 2: Cross-validated classification accuracies from session 1, for each investigated feature type*. The most discriminatory feature for each participant is bolded in blue. Accuracies exceeding the upper limit of the 95%, 99% and 99.9% confidence intervals of chance are marked with *, ** and *** respectively. Chance levels were calculated using the binomial distribution Combrisson & Jerbi (117).*

	Feature Type						
	Mean	Slope	Moving Slope	RMS	Variance	Skewness	Kurtosis
P1	65.0*	**74.3****	65.2*	63.7	63.8	69.5*	68.2*
P2	80.2***	**90.2*****	80.5***	67.7*	70.0**	77.0**	60.8
P3	63.0	73.3**	59.7	65.2*	64.8	70.3**	**76.0****
P4	70.8**	**74.2****	71.7**	63.7	64.2	62.8	66.8*
P5	81.5***	**92.7*****	90.8***	65.3*	67.2*	68.2*	63.0
P6	76.3**	**80.5*****	76.0**	68.8*	66.8*	75.2**	60.0
P7	62.5	63.5	**71.7****	63.3	65.2*	57.7	61.5
P8	60.5	72.7**	**82.8*****	67.3*	65.5*	62.7	74.8**
P9	66.2*	63.5	**71.0****	65.0*	64.8	62.8	65.2*
P10	**73.3****	65.8*	62.2	72.8**	72.0**	60.4	69.7*
Average	**69.9***	**75.1****	**73.2****	**66.3***	**66.4***	**66.7***	**66.6***
Std. Dev.	**±7.6**	**±10.2**	**±9.7**	**±2.9**	**±2.7**	**±6.4**	**±5.6**

Slope was found to be the most discriminatory feature for 5 of the 10 participants and moving slope (4s intervals with a 0.5s overlap) was most discriminatory for another 3 of the 10 participants. This aligns with previous work, as over half of reported NIRS-BCI studies have used either mean or slope as their feature of interest (79). Interestingly, mean was the best feature for only 1 of the 10 participants. Kurtosis yielded the highest classification accuracy for Participant 3, but slope was chosen instead for their online classification due to the poor performance of kurtosis as a feature across the rest of the participants. Variance, RMS and skewness also consistently performed less robustly as classification features than mean, slope and moving slope.

The average cross-validated classification accuracy for each participant, using their best feature, is displayed in figure 19. Using their best feature, every participant surpassed the chance level threshold. In response to the first research question of this thesis, these findings suggest that positive and negative emotional valence can be differentiated (offline) in children based on changes in hemodynamic activity in the prefrontal cortex.

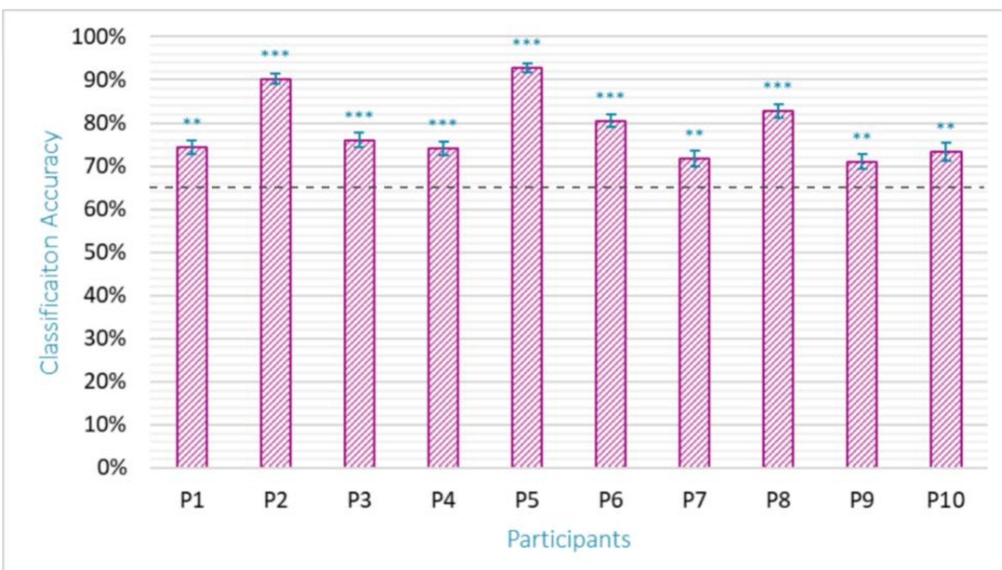


Figure 19: Offline cross-validated classification accuracies from session 1, using each participant's best feature. *Accuracies exceeding the upper limit of the 95%, 99% and 99.9% confidence intervals of chance are marked with *, ** and *** respectively.*

3.2 Session 2-4 Online Results

The second aim of this thesis was to test the BCI online to investigate if emotional valence could be predicted in real-time. A classifier was trained for each participant using their most discriminatory feature and the data collected from session 1. The participants then attended three more sessions where the trained classifier was used to predict emotional valence of the incoming, real-time changes in hemodynamic activity. The classifier was retrained after each block of 12 trials to incorporate same-day session data. For the online sessions, BCI performance was evaluated based on the accuracy of the real-time predictions. A correct classification was tallied if the predicted classification matched the valence of the stimuli for that trial. Classification accuracy was again defined as the percentage of correct classifications out of all the classification made (i.e. 60 classifications made for the 60 trials in a single session). The average online classification accuracies for each participant for each online session can be seen in figure 20 and are broken down by block in table 3. All participants achieved above-chance online classification accuracy for at least one of the online sessions, and 8 of the 10 participants achieved above-chance accuracy for two of the three online sessions. However, none of the participants achieved above-chance accuracy for all three online sessions.

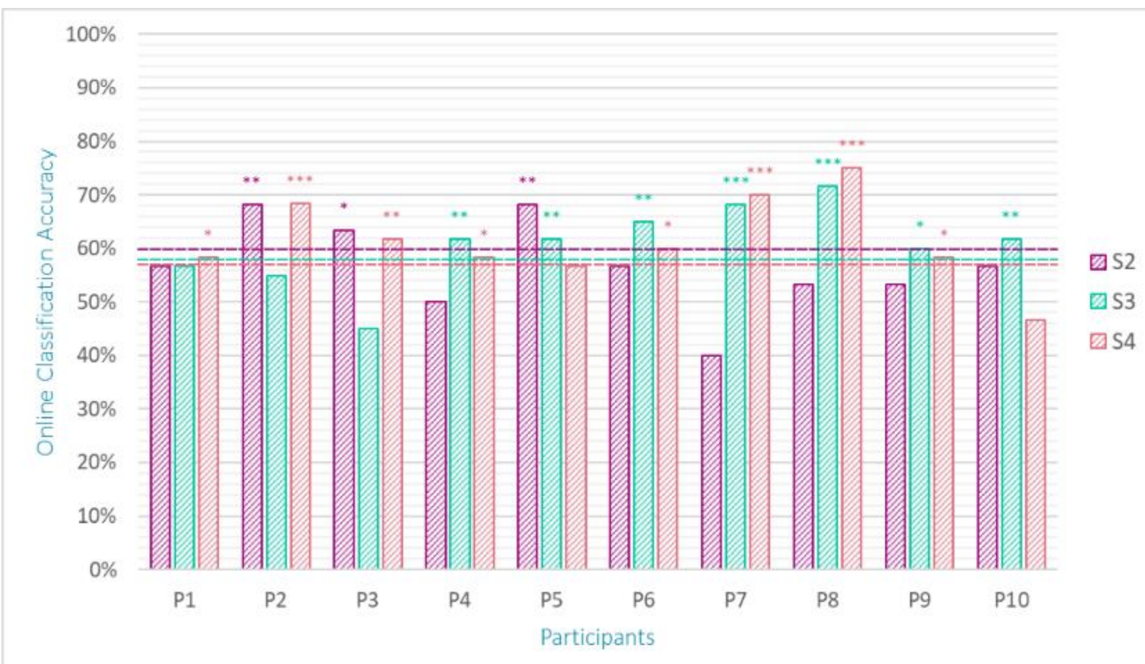


Figure 20: Average classification accuracies for each online session. *The hatched lines represent the 95% confidence level threshold for each session. This threshold decreases with each subsequent session, as the amount of data used to train the classifier increases with each session (117).*

BCI performance appeared to improve throughout a single session. On average, the last block of each session was more accurate than the first block of that session (e.g. $mean_{21}$ = 55.8% and $mean_{52}$ = 66.7%, where the first subscript denotes the session and the second denotes the block). However, these differences were not significant for any of the sessions ($p_{session2}$ = 0.28, $p_{session3}$= 0.12, $p_{session4}$ = 0.42, two-tailed t-test). As more same-day data are included in retraining a classifier, it is expected that accuracy will improve (107). In the current study, the length of the data collection session was limited to 5 blocks to prevent fatigue. It is likely that a significant improvement in classifier performance would have been achieved if the session were longer and more same day data could have been included in the classifier model.

Table 3: Average online classification accuracies for each session, broken down by block*. The entire session average is bolded. Accuracies exceeding the upper limit of the 95%, 99% and 99.9% confidence intervals of chance are marked with *, ** and *** respectively.*

Session 2											
	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	Average
Block 1	41.7	41.7	66.7**	33.3	75.0***	58.3	58.3	33.3	50.0	41.7	50.0
Block 2	58.3	83.3***	66.7**	83.3***	83.3***	50.0	41.7	50.0	50.0	41.7	60.8*
Block 3	25.0	66.7**	66.7**	50.0	58.3	58.3	41.7	66.7**	50.0	75.0***	55.8
Block 4	75.0***	66.7**	66.7**	41.7	58.3	66.7**	25.0	66.7**	75.0***	41.7	58.3
Block 5	83.3***	83.3***	50.0	41.7	66.7**	50.0	33.3	50.0	41.7	83.3***	58.3
Average	**56.7**	**68.3****	**63.3***	**50.0**	**68.3****	**56.7**	**40.0**	**53.3**	**53.3**	**56.7**	**56.7**
Session 3											
	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	Average
Block 1	66.7**	58.3*	41.7	50.0	50.0	50.0	41.7	75.0***	75.0***	50.0	55.8
Block 2	75.0***	41.7	25.0	58.3*	58.3*	100.0***	75.0***	66.7**	75.0***	58.3*	63.3**
Block 3	41.7	75.0***	58.3*	50.0	58.3*	66.7**	83.3***	75.0***	41.7	66.7**	61.7**
Block 4	41.7	33.3	50.0	66.7**	58.3*	33.3	66.7**	58.3*	75.0***	75.0***	55.8
Block 5	58.3*	66.7**	50.0	83.3***	83.3***	75.0***	75.0***	83.3***	33.3	58.3**	66.7**
Average	**56.7**	**55.0**	**45.0**	**61.7****	**61.7****	**65.0****	**68.3*****	**71.7*****	**60.0***	**61.7****	**60.7***
Session 4											
	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	Average
Block 1	41.7	66.7**	75.0***	75.0***	41.7	33.3	58.3*	100.0***	58.3*	58.3*	61.1*
Block 2	66.7**	66.7**	75.0***	33.3	50.0	66.7**	83.3***	83.3***	66.7**	50.0	64.2**
Block 3	50.0	66.7**	91.7***	83.3***	66.7**	58.3*	58.3*	66.7**	33.3	50.0	62.5**
Block 4	66.7**	66.7**	16.7	50.0	41.7	66.7**	75.0***	58.3*	66.7**	58.3*	56.7
Block 5	66.7**	75.0***	50.0	50.0	83.3***	75.0***	75.0***	66.7**	66.7**	16.7	67.6***
Average	**58.3***	**68.3*****	**61.7****	**58.3***	**56.7**	**60.0***	**70.0*****	**75.0*****	**58.3***	**46.7**	**62.4****

On a per-session basis, we see that only 3 of the 10 participants achieved above-chance online classification accuracies for the first online session, while 7 of the 10 and 8 of the 10 participants achieved above-chance level accuracies for the second and third online sessions respectively. Overall, there is a general trend of improvement in performance across the three online sessions (session 2 mean: 56.7%; session 3 mean: 60.7%; session 4 mean: 62.4%). This improvement could be due to enhanced classifier performance from an increasing amount of training data and/or the participants becoming more familiar with the task of emotional regulation. However, a one-way ANOVA analysis of the average classification accuracies did not reveal a significant difference among the sessions ($p = 0.38$, Shapiro-Wilkes test for normality). This could be attributed to the fact that while we see a general increase in online BCI performance for some participants (e.g. Participants 7 & 8), we observed a decrease in performance for others (e.g. Participant 5), while others appeared to have consistent performance, except for one anomalous session (e.g. Participants 2 & 3). Interparticipant and intraparticipant (intersession) variability will be explored in the following sections.

3.3 Hemodynamic Response Functions

Figure 21 shows the trial-averaged hemodynamic response function (HRF) for Δ[HbO] and Δ[Hb] for Participant 4 for the first two sessions. The response signal from each measurement channel is shown mapped according to its position over the forehead. A clear distinction can be seen between the positive and negative response for both Δ[HbO] and Δ[Hb] in both sessions. However, both the positive and negative responses seem to vary across the different session days. Figure 22 shows the trial-averaged HRFs for Δ[HbO] for Participant 2 for all four sessions. Once again, we can see variability in the response across the different sessions. Interestingly, the session that appears most dissimilar, session 3, was also Participant 2's worst session in terms of online classification performance. It is also interesting to note that the variability between participants appears to be more significant than the variability between sessions for an individual participant, although these conclusions are based on visual inspection only.

3.4 Intensity Ratings

Participants were asked to rate the intensity of their emotions after each block of trials on a 5-point scale (from 1, representing 'not intense at all', to 5, representing 'extremely intense'). The intensity ratings for each block, for each participant, can be seen in the heatmaps in figure 23. The heatmaps are coloured according to the intensity rating and labeled with the corresponding classification accuracy for that block. Below the heatmap for each participant is a figure showing the average intensity rating alongside the average classification accuracy for each session. It was expected that a higher emotional intensity rating would indicate stronger evoked emotions for that block, resulting in a stronger hemodynamic response and yielding more accurate classification. However, no significant correlation between intensity rating and classification accuracy was found for any of the participants. Many of the participants rated each block consistently within and across sessions, with each of their ratings differing only by 1-2 points on the scale.

3.5 Session Experience Questionnaire

At the end of each session, the participants were asked to complete a questionnaire assessing their mental state before, after and throughout the session (Appendix B). There were four main categories of mental state assessed in the questionnaire – alertness/fatigue, mood, effort and frustration. It was hypothesized that differences in mental state could provide insight on variations in participant performance across different session days.

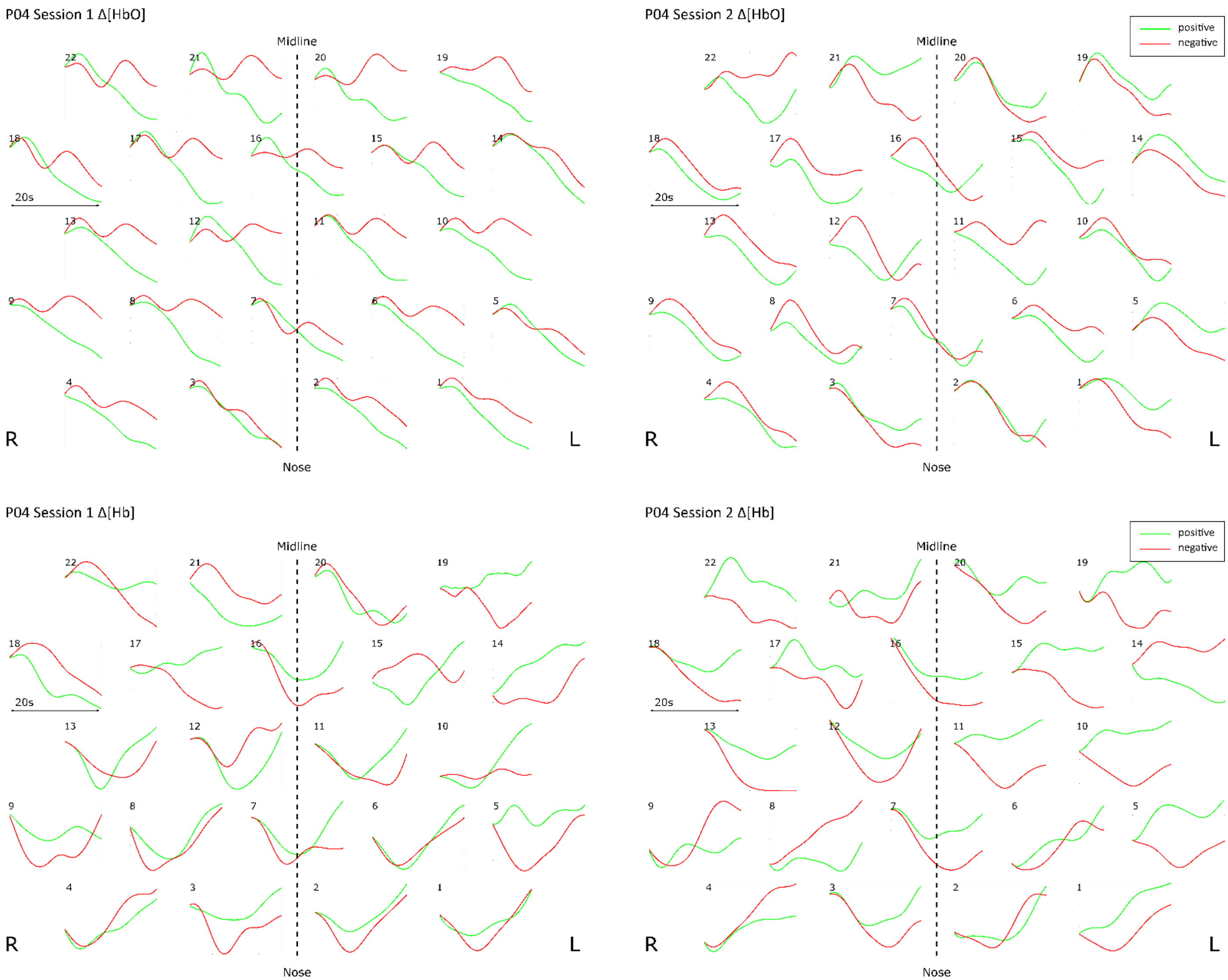


Figure 21: Trial-averaged hemodynamic response functions for Participant 4, Sessions 1 & 2. *The upper left figure shows Δ[HbO] for Session 1. The upper right figure shows Δ[HbO] for Session 2 The bottom left figure shows Δ[Hb] for Session 1, and the bottom right figure shows Δ[Hb] for Session 2. The response function for each measurement channel is shown mapped according to its position over the forehead. Left and right sides of the head as well as the midline and the nose are provided as landmarks. Green indicates the average of all 30 positive trials, and red indicates the average of all 30 negative trials from the respective session. Each response function is 20s in length.*

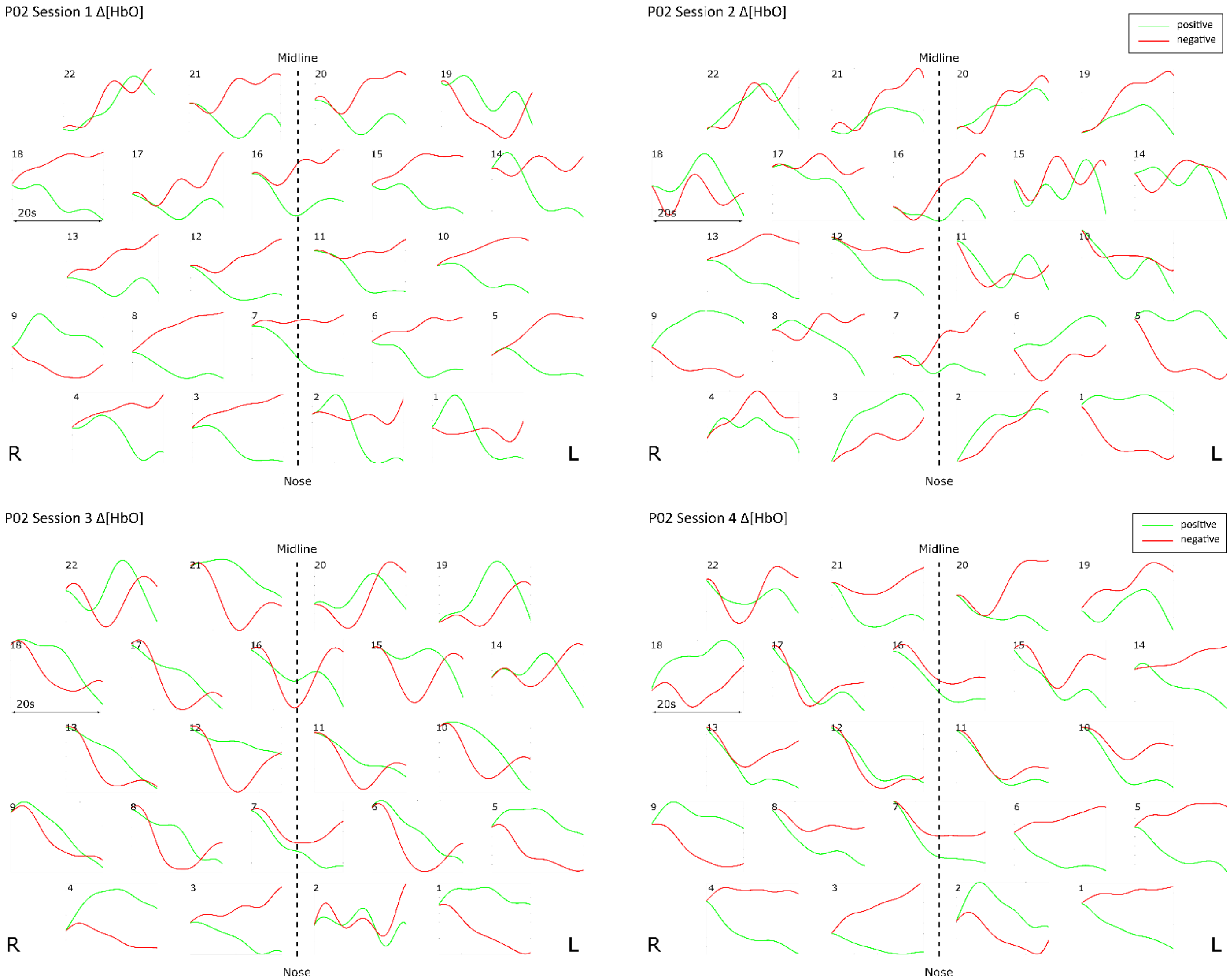


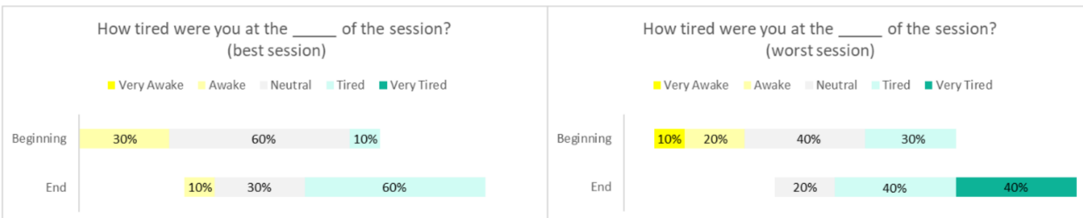
Figure 22: Trial-averaged hemodynamic response functions (Δ[HbO]) for Participant 2 for all four sessions*. The upper left figure shows Δ[HbO] for Session 1. The upper right figure shows Δ[HbO] for Session 2. The bottom left figure shows Δ[HbO] for Session 3, and the bottom right figure shows Δ[HbO] for Session 4. The response function for each measurement channel is shown mapped according to its position over the forehead. Left and right sides of the head as well as the midline and the nose are provided as landmarks. Green indicates the average of all 30 positive trials, and red indicates the average of all 30 negative trials from the respective session. Each response function is 20s in length.*

Figure 23: End-of-block intensity ratings compared to online classification accuracy. *The top row of figures shows heatmaps of the intensity rating for each block across all three online sessions for each participant. Each rectangle represents one block of one session. The colour of the rectangle indicates the intensity rating, with cooler colours indicating a low-intensity rating (dark green = 1, "not intense at all"), and warmer colours indicating a high-intensity rating (red = 5, "extremely intense"). Each rectangle is also labeled with the corresponding average online classification accuracy from that block. For example, P1 (upper left) rated the first block of session 2 as "5 - highly intense" (red color), and the classification accuracy for that block was 41.7%, while they rated the second block of session 2 as "3 – kind of intense" (yellow), and the accuracy for that block was 58.7%. The bottom row of figures displays the average intensity rating for each session, alongside the average online classification accuracy for that session. For example, P1's average classification accuracy (solid yellow line) for each session was relatively consistent, while their average intensity rating (dotted yellow line) was the lowest in session 3 and the highest in the consecutive session 4.*

3.5.1 Alertness/Fatigue

Participants were asked how tired they were at the start and at the end of the session, on a scale from 1 to 5, with 1 representing "very tired" and 5 representing "very awake." Figure 24A shows the responses to this question for the participants' best and worst sessions. Overall, participants were more tired at the end of the session on their worst session days (mean fatigue = 2.5 at the end of their best session, mean fatigue = 1.8 at the end of their worst session) and experienced a greater decrease in alertness from start to finish on their worst session days (mean difference in fatigue = 0.7 for their best session, mean difference in fatigue = 1.3 for their worst session). However, these differences were not significant (Figure 24B). Figure 24C shows the relationship between change in alertness across the session and average classification accuracy, although this relationship is uncorrelated.

A)



B)

	Alertness Rating		
	Session Start	Session End	Difference
Mean Best Session	3.2	2.5	0.7
Mean Worst Session	3.1	1.8	1.3
P-value	0.91	0.11	0.33

C)

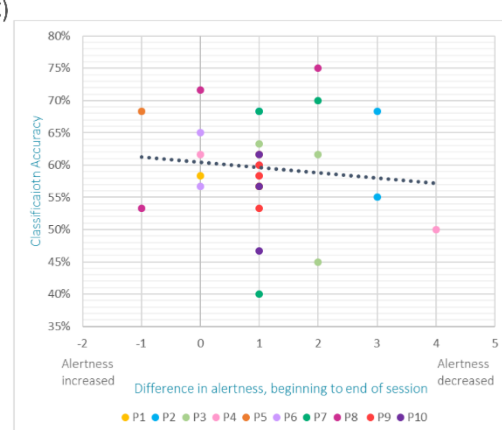
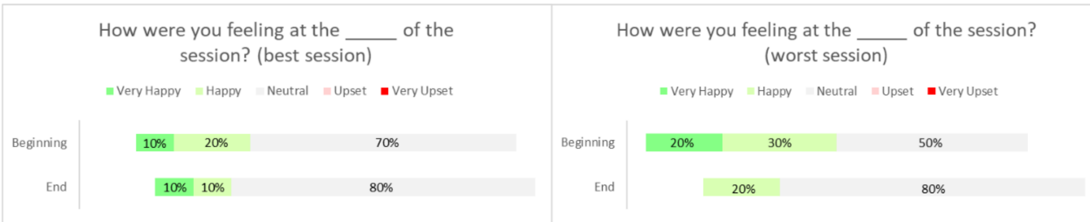


Figure 24: Questionnaire responses to questions regarding alertness/fatigue.
A) Percentage of participants who responded with each possible answer to the questions "How tired were you at the [beginning/end] of the session?" on their best and worst session days;
B) Average alertness ratings for the best and worst sessions at the start and end of the session as well as the difference between them (note that 1 = very tired, 5 = very awake). The p-values for a Wilcoxon signed-rank test assessing the differences between the ratings for best and worst sessions are also shown;
C) Scatterplot of average online classification accuracy plotted against the difference in alertness from beginning to the end of the session (note that 1 corresponds to a drop in alertness rating by one point on the scale, 2 corresponds to a drop in two points, etc.).

3.5.2 Mood

Participants were also asked how they were feeling at the start and at the end of the session, on a scale from 1 to 5, with 1 representing "very upset" and 5 representing "very happy". Figure 25A shows the responses to this question for the participants' best and worst sessions. On the day of their worst session, participants experienced a larger downward shift in their mood compared to the day of their best session (mean change in mood = 0.1 for the best session, mean change in mood = 0.5 for the worst session). Once again, this difference was not significant (Figure 25B). The relationship between classification accuracy and change in mood across the session can be seen in Figure 25C, although once again the results are uncorrelated.

A)



B)

	Mood Rating		
	Session Start	Session End	Difference
Mean Best Session	3.4	3.3	0.1
Mean Worst Session	3.7	3.2	0.5
P-value	0.50	1.00	0.22

C)

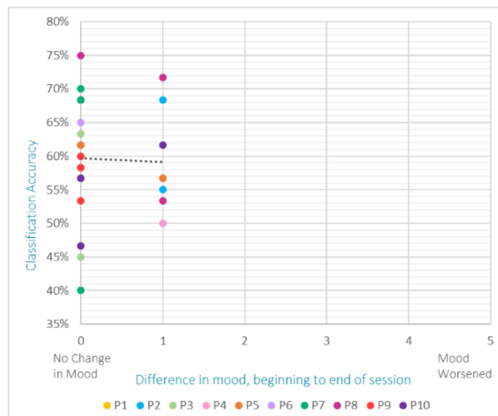


Figure 25: Questionnaire responses to questions regarding mood.
A) Percentage of participants who responded with each possible answer to the questions "How were you feeling at the [beginning/end] of the session?" on their best and worst session days;
B) Average mood ratings for the best and worst sessions at the start and end of the session as well as the difference between them (note that 1 = very upset, 5 = very happy). The p-values for a Wilcoxon signed-rank test assessing the differences between the ratings for best and worst sessions are also shown;
C) Scatterplot of average online classification accuracy plotted against change in mood from beginning to the end of the session (note that 0 corresponds to no change in mood, 1 corresponds to a drop in mood by 1 scale point).

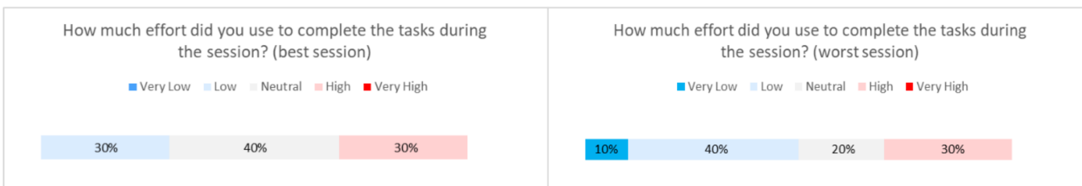
3.5.3 Effort

The third question asked how much effort the participants had to exert to complete the tasks within the session. The possible answers ranged from 1, representing "very low", to 5, representing "very high". Figure 26A shows the responses, for each participant's best and worst sessions. On average, participants used slightly more effort on their best session day compared to their worst session day (mean effort = 3.0 for the best session, mean effort = 2.7 for the worst session), although the difference is very small and not significant (Figure 26B). No significant correlation can be seen between classification accuracy and perceived effort exerted during the task of emotional regulation (Figure 26C).

3.5.4 Frustration

Finally, the last question asked the participants how much frustration they experienced during the session. Once again, they were given possible answers from 1; “very low”, to 5; “very high”. Figure 27A shows these responses for the best and worst sessions. On average, participants experienced greater frustration on their worst session days compared to their best session days (mean frustration = 2.1 for the best session, mean frustration = 2.5 for the worst session). However, the difference is small and not significant (Figure 27B). There was no significant correlation between classification accuracy and frustration experienced during the session (Figure 27C).

A)



B)

	Effort Rating
	Entire Session
Mean Best Session	3.0
Mean Worst Session	2.7
P-value	0.38

C)

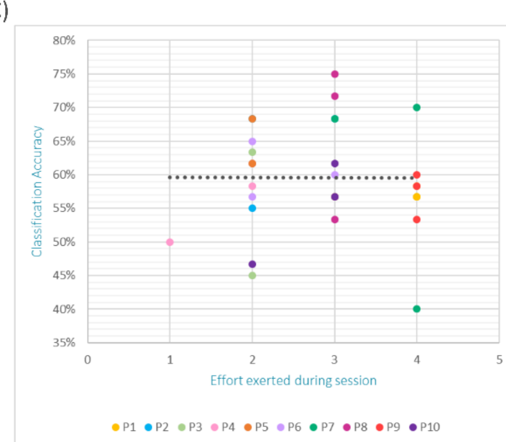
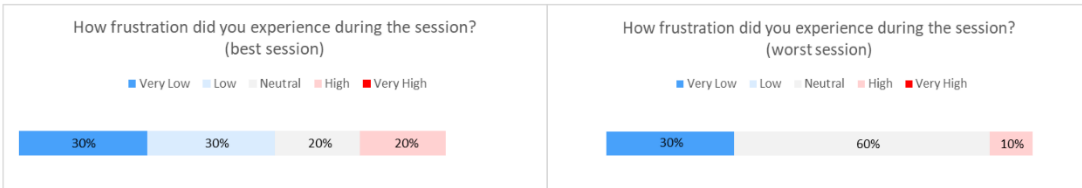


Figure 26: Questionnaire responses to questions regarding effort.
A) Percentage of participants who responded with each possible answer to the question “How much effort did you use to complete the tasks during the session?” for their best and worst session days;
B) Average effort ratings for the best and worst session (note that 1 = very low effort, 5 = very high effort). The p-values for a Wilcoxon signed-rank test assessing the differences between the ratings for best and worst sessions are also shown;
C) Scatterplot of average online classification accuracy plotted against effort exerted during the session.

A)



B)

	Frustration Rating
	Entire Session
Mean Best Session	2.1
Mean Worst Session	2.5
P-value	0.34

C)

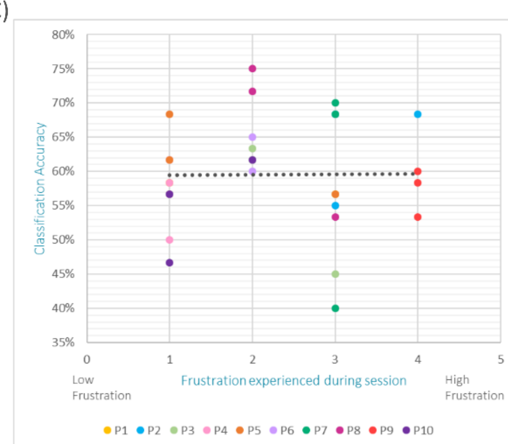


Figure 27: Questionnaire responses to questions regarding frustration.
A) Percentage of participants who responded with each possible answer to the question "How much frustration did you experience during the session?" for their best and worst session days;
B) Average frustration ratings for the best and worst session (note that 1 = very low frustration, 5 = very high frustration). The p-values for a Wilcoxon signed-rank test assessing the differences between the ratings for best and worst sessions are also shown;
C) Scatterplot of average online classification accuracy plotted against frustration experienced during the session. 27

3.5.5 Individual Participant Trends

Differences in mental state and their impact on session performance were also observed on an individual participant basis. There is evidence for many of the participants that a distracting mental state, such as fatigue, a poor mood, or frustration impacted BCI performance on different session days. For example, Participant 4 experienced the most fatigue during their session with their lowest BCI performance (session 2), and the least fatigue during their session where they obtained their highest BCI performance (session 3), suggesting that the fatigue they experienced on the day of session 2 impacted their ability to use the BCI effectively. More participant-specific observations can be seen in Table 4 below.

***Table 4: Participant-specific observations in alertness, mood, effort and frustration on different session days** in relation to their online BCI performance.*

Participant	Observation
P1	Experienced the least fatigue on session with highest performance (S4).
P2	No trends observed.
P3	Experienced the most frustration on the only below-chance session (S3).
P4	Experienced the least fatigue on session with highest performance (S3) and the most fatigue on session with the lowest performance (S2). Experienced the greatest decrease in mood on session with lowest performance (S2). Used the least effort on the session with the lowest performance (S2).
P5	Experienced the least fatigue on session with highest performance (S2). Experienced the greatest decrease in mood on only below-chance session (S4). Used the most effort on session on only below-chance session (S4). Experienced the most frustration on only below-chance session (S4).
P6	Experienced the most frustration on only below-chance session (S2).
P7	No trends observed.
P8	Experienced the most frustration on only below-chance session (S2).
P9	No trends observed.
P10	Experienced the greatest decrease in mood on session with highest performance (S3).

3.6 Age & Sex

Figure 28 shows the average online classification accuracy for each participant, arranged by age from youngest to oldest. A significant positive correlation can be seen between online BCI performance and age (r=0.71, p=0.02). Three of the highest performers (P5, P7 and P8) were the three oldest participants, all at least 13 years old. The two lowest performers (P1 and P9) were also among the youngest participants, at ages 8 and 10 respectively.

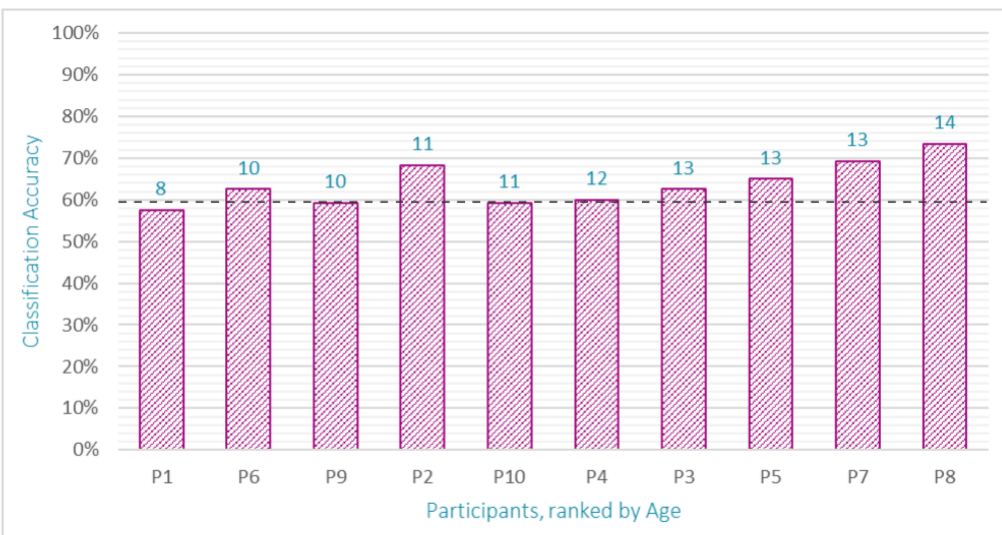


***Figure 28: Average online classification accuracy, arranged by age**. The hatched line marks the 95% confidence level threshold (for n=60 samples for each class). The participant's age is labeled on top of their respective bars.*

Figure 29 shows the average online classification accuracy for each participant, arranged by sex. There are no evident trends between online BCI classification accuracy and sex (mean acc = 63.6% for males, mean acc = 63.7% for females; p = 0.9813, Welch's t-test).

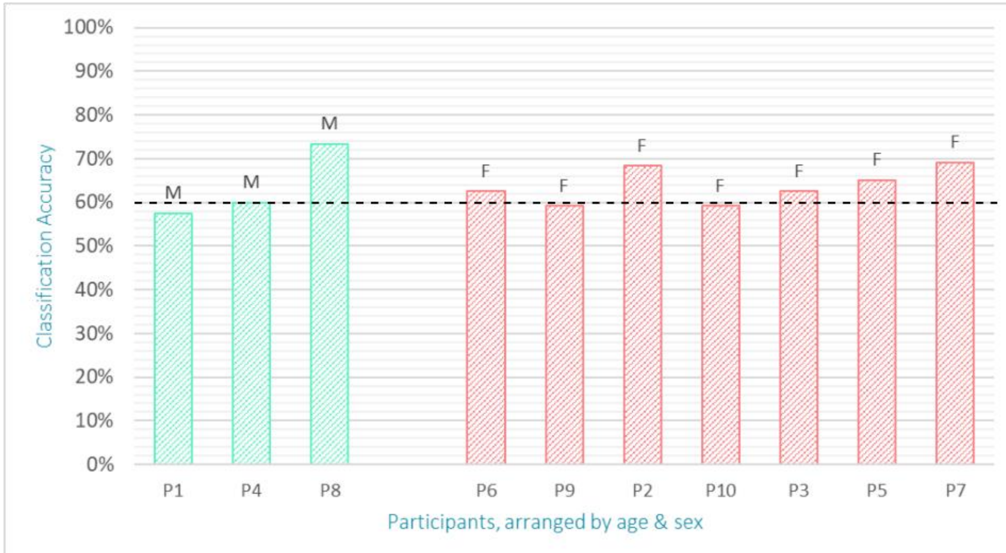


Figure 29: Average online classification accuracy, arranged by age and gender. *The hatched line marks the 95% confidence level threshold (for n=60 samples for each class).*

Figures 30 and 31 show the average online classification sensitivity and specificity for each participant, arranged by sex. Sensitivity represents the number of correctly predicted positive trials out of the total number of positive trials, and specificity represents the number of correctly predicted negative trials out of the total number of negative trials (see Appendix D for formulae). Thereby, sensitivity can be seen as

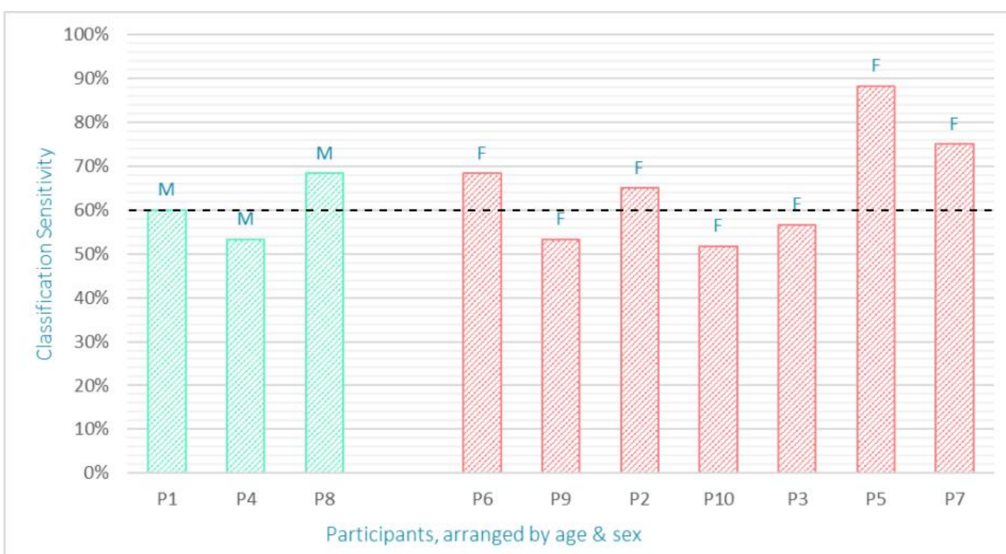


Figure 11: Average online classification sensitivity, arranged by age and sex. *The hatched line marks the 95% confidence level threshold (for n=60 samples for each class).*

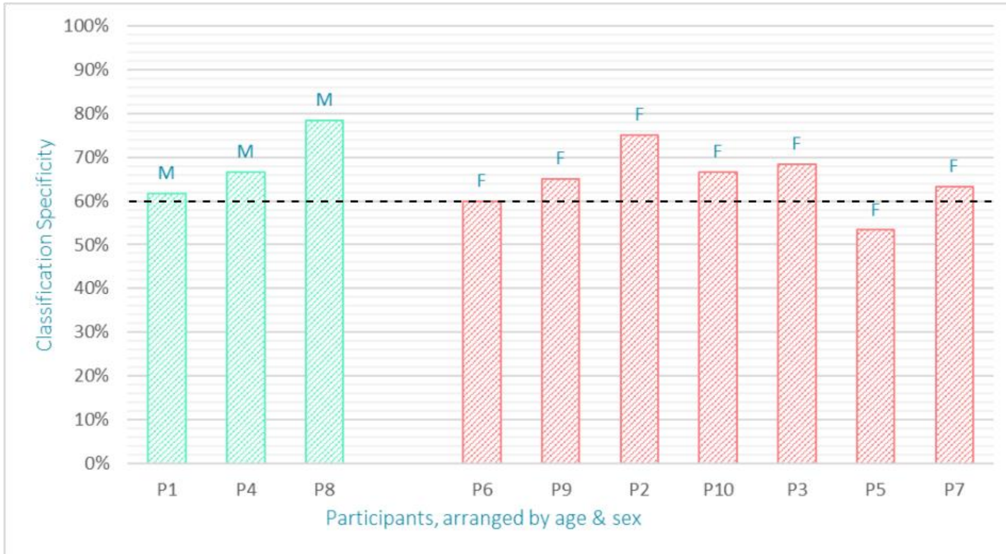


Figure 31: Average online classification specificity, arranged by age and sex. *The hatched line marks the 95% confidence level threshold (for n=60 samples for each class).*

a measure of how well the participants were able to respond to/regulate positive emotional states, and specificity a measure of how well they were able to respond to/regulate negative emotional states. Females appeared to be better at regulating positive emotional states (mean sens = 60.6% for males, mean sens = 65.5% for females) and males appeared to be better at regulating negative emotional states (mean spec = 68.9% for males, mean spec = 64.5% for females). However, the differences were not significant (p=0.4376, p=0.3176 for sensitivity, specificity respectively, Welch's t-test). Table 5 summarizes these results. It should be noted that the small sample size (n=3 males) makes it difficult to draw any statistical conclusions on BCI performance based on sex.

Table 5: Sex differences between average online classification accuracy, sensitivity and specificity. *Group means are shown as well as p-values from a Welch's t-test (note that n=3 males and n=7 females).*

	Sex Differences		
	Male	Female	P-value
Accuracy	63.6%	63.7%	0.98
Sensitivity	60.6%	65.5%	0.44
Specificity	68.9%	64.5%	0.32

3.7 Temperament Questionnaire

The parent-reported EATQ assessed the participants' temperament along ten different dimensions: activation control, affiliation, aggression, attention, depressive mood, fear, frustration, inhibitory control, shyness and surgency. Each question on the EATQ was answered with a five-point Likert scale (1 = "almost never true", 5 = "almost always true") and was categorized under one of the ten dimensions. Mean scores for each dimension were calculated by averaging the responses for all the questions pertaining to that dimension, resulting in a score from 1-5 for each of the ten dimensions. These scores were compared to online BCI performance (classification accuracy, sensitivity and specificity) for each participant. Pearson's correlation between accuracy and score for each EATQ score (118) is reported in Table 6A. The only significant correlation was between aggression and sensitivity (r = +0.63, p = 0.05) and specificity (r = -0.66, p = 0.04), suggesting that participants who were more aggressive in temperament were better at regulating positive emotional valence and worse at regulating negative emotional valence.

Table 6: Temperament measures - Pearson's correlation coefficient (r) and p-value (p) for: *A) Classification accuracy, sensitivity and specificity with respect to each of the ten measures of temperament; and B) Classification accuracy, sensitivity and specificity with respect to each of the three super-measures of temperament.*

A)

	Activation Control		Affiliation		Aggression		Attention		Depressive Mood	
	r	*p*	*r*	*p*	*r*	*p*	*r*	*p*	*r*	*p*
Accuracy	-0.05	0.89	0.06	0.86	0.23	0.52	-0.19	0.59	0.36	0.31
Sensitivity	-0.02	0.95	0.19	0.60	0.63	0.05*	-0.11	0.75	0.20	0.59
Specificity	0.23	0.53	0.28	0.43	-0.66	0.04*	0.07	0.84	0.22	0.55
	Fear		**Frustration**		**Inhibitory Control**		**Shyness**		**Surgency**	
	r	*p*	*r*	*p*	*r*	*p*	*r*	*p*	*r*	*p*
Accuracy	-0.31	0.39	0.26	0.47	-0.23	0.53	0.43	0.22	-0.07	0.85
Sensitivity	-0.27	0.46	0.33	0.35	-0.46	0.18	0.41	0.24	0.45	0.19
Specificity	0.22	0.54	0.00	1.00	0.35	0.32	-0.34	0.33	-0.28	0.43

B)

	Effortful Control		Emotionality/ Negative Affect		Extraversion/ Surgency	
	r	*p*	*r*	*p*	*r*	*p*
Accuracy	-0.11	0.77	0.41	0.24	-0.03	0.93
Sensitivity	-0.13	0.73	0.64	0.05*	0.09	0.80
Specificity	0.15	0.67	-0.35	0.32	0.07	0.84

The ten dimensions from the EATQ were also be combined into three super-scales; effortful control, emotionality/negative affect and extraversion/surgency. These scores were determined for each participant and compared with online BCI performance. Pearson's correlation coefficients for these comparisons can be found in Table 6B. The only significant correlation was found between emotionality and sensitivity ($r = +0.64$, $p = 0.05$), suggesting that participants with high levels of negative emotionality were better at regulating positive emotional states.

To further investigate the influence of temperament on BCI performance, the participants were divided into two groups – high BCI performers (BCI+) and low BCI performers (BCI-). Participants whose average classification accuracy across the three online sessions was greater than the group mean were assigned to the BCI+ group (P2, P5, P6, P7, P8), while participants whose average classification accuracy was lower than the group mean were assigned to the BCI- group (P1, P3, P4, P9, P10). Group scores for each of the ten temperament measures and three aggregated measures were compared using Welch's t-test and can be seen in Figure 32. Group means and p-values are labeled on the plots. No statistical differences were found between the two groups in any of the measures of temperament.

In the previous section, it was demonstrated that there is a positive correlation between BCI performance and age. However, there were two younger participants (P2 and P6) who achieved high classification accuracies, despite their age. To investigate whether temperamental differences could have contributed to their success with the BCI, the high BCI performers and low BCI performers groups were further divided by age. Participants in each of the BCI+ and BCI- groups whose age was older than the group mean were assigned to the 'older' group, and those whose age was younger than the group mean was assigned to the 'younger' group, creating four different groups – BCI+ older, BCI+ younger, BCI- older, BCI- younger. Group scores for each of the ten temperament measures and three aggregated measures were compared between the BCI+ younger and BCI- younger groups, and the BCI+ older and BCI- older groups. These results can be seen in Figure 33. Group means and p-values (Welch's t-test) are labeled on the plots.

Figure 122: Differences in temperament group scores between high-performing (BCI+) and low-performing BCI (BCI-) users in *A) the ten temperament measures of activation control, affiliation, aggression, attention, depressive mood, fear, frustration, inhibitory control, shyness and surgency; and B) the three super-measures of effortful control, negative affect/emotionality and extraversion.*
Group means are labeled with an 'x' and their value is displayed underneath each box-whisker plot. The upper and lower bounds of the rectangles represent the third and first quartiles respectively, and the middle line represents the median. The group means were compared using Welch's t-test and the resulting p-values are displayed on each plot.

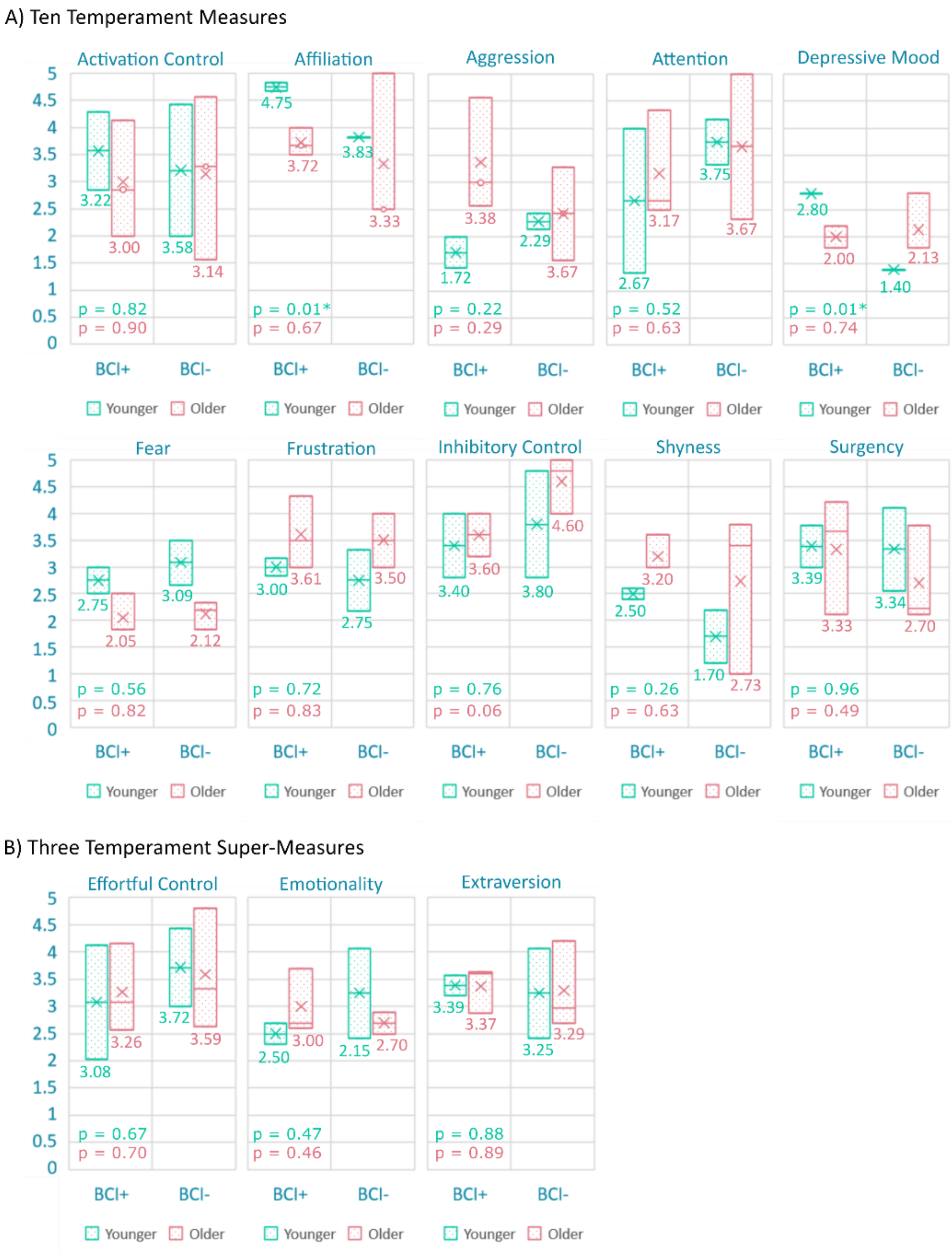


Figure 33: Differences in temperament group scores between high-performing (BCI+) and low-performing BCI (BCI-) users, divided by age *in A) the ten temperament measures of activation control, affiliation, aggression, attention, depressive mood, fear, frustration, inhibitory control, shyness and surgency; and B) the three super-measures of effortful control, negative affect/emotionality and extraversion.*

The young participant groups are shown in green and the older participant groups are shown in red. Group means are labeled with an 'x' and their value is displayed underneath each box-whisker plot. The upper and lower bounds of the rectangles represent the third and first quartiles respectively, and the middle line represents the median. The group means were compared using Welch's t-test and the resulting p-values are displayed on each plot.

A significant difference was found between the high-performing younger participants and the low-performing younger participants in the measures of affiliation (means 4.75, 3.83; p=0.01) and depressive mood (means 2.80, 1.40; p=0.01), with the high-performing younger participants scoring higher in both measures. No significant differences were found in any of the measures between the high and low performing older participant groups. However, these conclusions should only be considered tentatively, as the sample size for both groups is very small (n=2 for the younger age group, n=3 for the older age group), limiting statistical power. A larger sample population is needed to fully understand how temperament affects emotional regulation ability and BCI performance across different age groups.

3.8 Offline Cross-Validated Classification Results, All Sessions

In addition to the real-time, online classification accuracies, 10 iterations of 10-fold cross-validation were used to evaluate the BCI performance offline for sessions 2-4. The resulting average classification accuracies, using each participant's best feature, can be found in Figure 34. Accuracies exceeding the

A)

Session	Participant										
	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	Average
1	74.3**	90.2***	73.3**	74.2**	92.7***	80.5***	71.7**	82.8***	71.0**	73.3**	78.4**
2	75.7**	80.5***	78.2**	71.7**	78.8**	73.2**	80.0***	74.3**	77.7**	83.8***	77.4**
3	62.3	80.0***	60.0	72.5**	91.3***	83.3***	78.5**	79.3**	81.2***	78.7**	76.7**
4	69.2*	85.5***	68.5*	78.5**	78.5**	79.2**	82.8***	88.2***	72.8**	65.3*	76.9**
Average	**70.4****	**84.0*****	**70.0****	**74.2****	**85.3*****	**79.0****	**78.3****	**81.2*****	**75.7****	**75.3****	**77.3****

B)

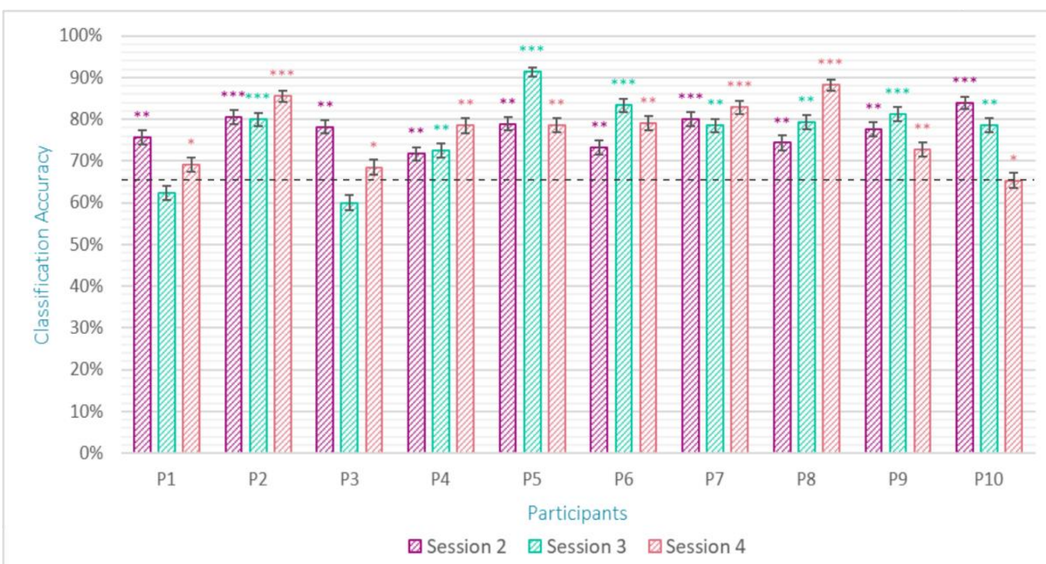


*__Figure 134: Offline average cross-validated classification accuracies for all four sessions individually__. A) Table of average offline classification accuracies from 10 iterations of 10-fold cross-validation for each session individually; and B) Bar chart of average offline cross-validated classification accuracies for sessions 2, 3 and 4 individually. The hatched line marks the highest 95% confidence level threshold (n=30 samples for each class) and accuracies exceeding the upper limit of the 95%, 99% and 99.9% confidence intervals of chance are marked with *, ** and *** respectively (117).*

upper limit of the 95%, 99% and 99.9% confidence intervals of chance are marked with *, ** and ***, respectively. Again, all chance levels were calculated using the binomial distribution (117). Comparing the offline results of all four sessions, 38 of the 40 sessions exceeded the chance level of 65% (excluding P1 session 2 and P3 session 2). 35 of the 40 sessions exceeded the 99% confidence level, and 15 of the 4 sessions exceeded the 99.9% confidence interval. The five 'top-performing' participants had average classification accuracies across all four sessions of 84.0%, 85.3%, 79.0%, 78.3% and 81.2% for Participants 2, 5, 6, 7, and 8 respectively. The five 'low-performing' participants had average classification accuracies across all four sessions of 70.4%, 70.0%, 74.2%, 75.7% and 75.3% for Participants 1, 3, 4, 9, and 10 respectively. For each session, the average offline classification accuracy across all participants was 78.4%, 77.4%, 76.7% and 76.9% for sessions 1 to 4 respectively. This provides strong evidence that emotional valence can consistently be differentiated in children from changes in hemodynamic activity evoked during an emotion-induction task.

3.8.1 Consistency of Feature Performance

The 'most discriminatory feature' chosen for each participant after the first offline session was used to train all the classifiers for the remaining three online sessions. Looking retrospectively, we can investigate whether this choice of feature performed consistently as best feature across different session days. Figure 35 shows the frequency at which each feature type performed as 'best feature', 'second best feature' and 'third best feature'. The size of the rectangle indicates frequency, and the bolded box indicates the feature that was chosen to be the 'most discriminatory feature' after session 1. Figure 35 divides the participants into two groups – the high BCI performers (BCI+), who achieved higher than average classification accuracies in their online sessions, and low BCI performers (BCI-), who achieved lower than average classification accuracies in their online sessions. The BCI+ group shows greater consistency in feature performance; for these participants, their chosen 'most discriminatory feature' remained the best or second-best performing feature across all four sessions. For this group, we also see the same three or four features appear consistently in the top three across all sessions. In comparison, in the BCI- group, we see a greater variety in the top three performing features, and the chosen 'most discriminatory feature' was not necessarily the best performing feature, or even in the top three, across all four sessions. For example, P10's chosen feature, mean, only appeared in the top three once out of all four sessions. With such variability in their responses, it may have been insufficient to select a single 'most discriminatory feature' type for these participants based only on a single session of data, contributing to their lower performance.

Figure 145: The frequency at which each feature type appeared as most discriminatory for each participant across the four sessions. *BCI performance was evaluated using cross-validation for each investigated feature type (mean, slope, moving slope, variance, RMS, skewness and kurtosis) for each of the four sessions individually. The feature type that yielded the highest, second-highest and third-highest classification accuracies for each session were deemed the best, second-best and third-best feature for that session respectively. The size of the rectangle indicates the frequency (i.e. the larger the rectangle, the more frequent) at which that feature was found to be either the best feature (top row), the second-best feature (middle row) or the third-best feature (bottom row) out of all seven feature types. The bolded rectangle indicates which feature was chosen after session 1 as the 'most discriminatory feature' for each participant to train their online classifiers for the three online sessions.*

3.8.2 Offline vs. Online Classification Results

For all participants, the offline cross-validated accuracies for each session were found to be higher than their respective online accuracies. Figure 36 shows a direct comparison between the offline and online classification accuracies. Although the offline accuracies are higher, they follow a similar trend to that of the online accuracies. That is, for most participants, the session with the best online performance was also the session with the highest offline cross-validated classification accuracy, and correspondingly, the session with the worst online performance was also the session with the lowest offline cross-validated classification accuracy. So overall, if the participant was having a 'good' or 'bad' session, this was reflected in both the real-time and the offline results.

3.8.3 Multiple Session Cross-Validation Results

To investigate how more training data impacted the BCI's ability to predict emotional valence, the performance of the BCI was evaluated through 10-fold cross-validation using every possible grouping of two, three and all four sessions combined into one training data set. The resulting classification accuracies can be seen in Table 7. Because the chance-level threshold decreases as the classifier is trained on more samples (e.g. 65% for n=30 samples of each class in one session, and 60% for n=60 samples of each class in two sessions), the classification accuracies from one session, two sessions, three sessions and four sessions of combined data cannot be compared directly. Instead, the accuracies were normalized with respect to each chance-level threshold ($normalized\ accuracy = \frac{(accuracy\% - threshold\%)}{(100\% - threshold\%)}$). These normalized accuracies are plotted against the number of sessions of training data in Figure 37. A slight decrease (or no change) in classification accuracy can be seen as the number of sessions of training data increases. This means that classifier performance did not improve, despite being trained on more data. This contributes to the evidence that that the hemodynamic response during an emotion-induction task, while generally machine-discernible at above-chance, is not consistent across different session days.

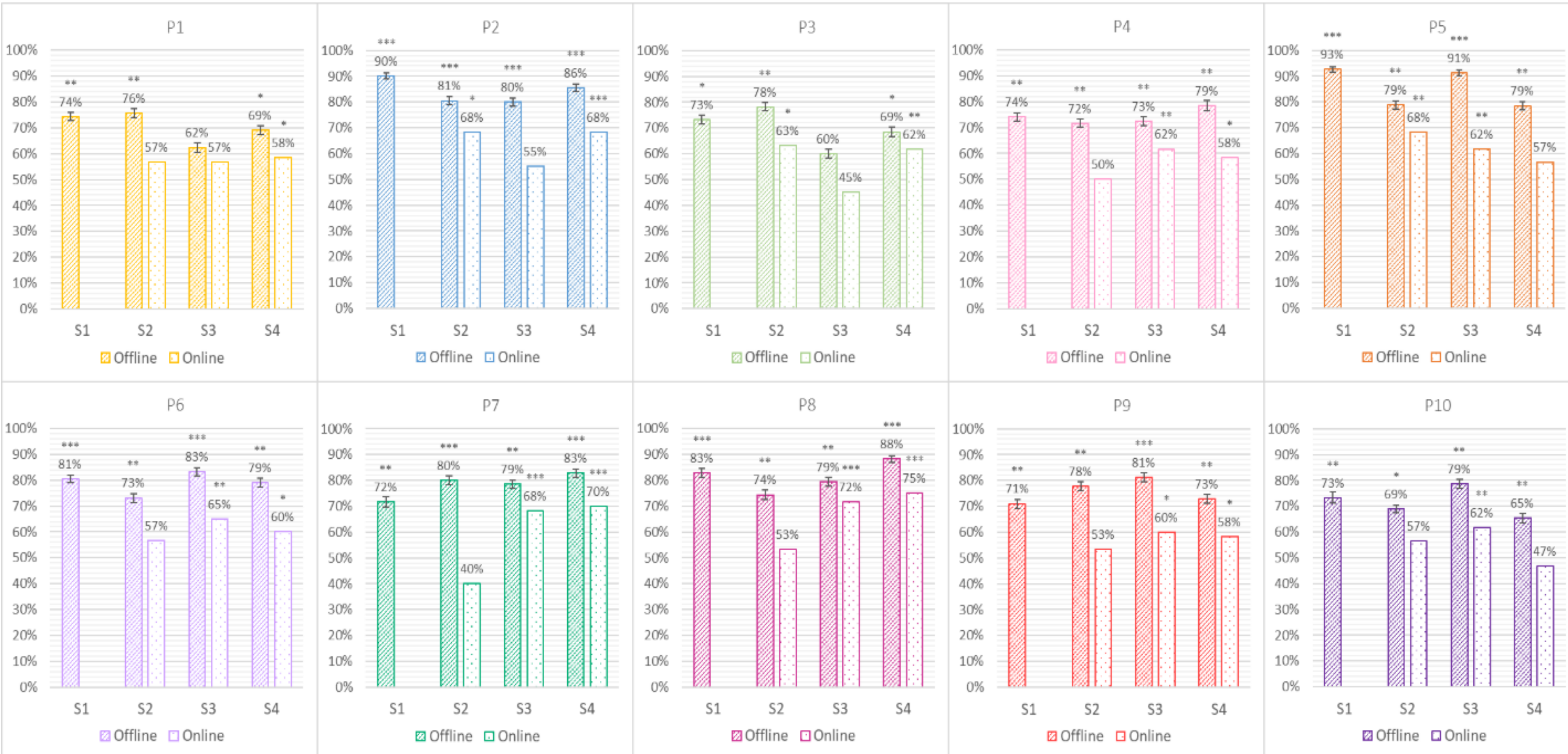


*__Figure 156: Offline average cross-validated classification accuracies plotted against the online average classification accuracies for each session__. The hatched bars show the offline accuracies and the dotted bars show the online accuracies. Accuracies exceeding the upper limit of the 95%, 99% and 99.9% confidence intervals of chance are marked with *, ** and *** respectively. Note that once again, these thresholds depend on the amount of data used to train the classifier (117).*

Table 7: Offline average cross-validated classification accuracies from every possible combination of two, three and four sessions of training data*. Accuracies exceeding the upper limit of the 95%, 99% and 99.9% confidence intervals of chance are marked with *, ** and *** respectively (117).*

Session	Participant									
2 Sessions	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	Average
1&2	65.6**	79.8***	67.1**	67.5**	83.1***	72.1***	68.0**	65.3**	67.6**	68.7**	70.5**
2&3	61.1*	75.4***	65.0**	71.9***	78.1***	73.1***	76.0***	73.0***	72.2***	77.9***	72.4***
1&3	66.8**	80.2***	60.3*	65.6**	83.4***	83.3***	71.7***	69.8**	71.3***	66.5**	71.9***
1&4	62.7*	86.6***	61.6*	69.8**	72.2***	76.5***	72.3***	79.8***	67.8**	55.1	70.4**
2&4	67.4*	80.8***	71.8***	70.7***	77.2***	66.9**	78.4***	80.0***	68.9**	66.3**	72.8***
3&4	65.8**	75.3***	61.6*	73.9***	83.3***	79.9***	73.3***	82.2***	67.3**	67.7**	73.0***
3 Sessions											
1&2&3	65.2**	74.6***	60.2*	66.5***	80.3***	73.8***	67.4***	64.3**	66.6***	69.0***	68.8***
1&2&4	61.7**	81.2***	66.9***	67.1***	78.0***	70.8***	68.5***	68.4***	64.5**	61.9**	68.9***
1&3&4	64.2**	79.6***	60.2*	69.6***	77.9***	78.5***	66.4***	73.9***	65.4**	64.0**	70.0***
2&3&4	65.9**	78.6***	65.1**	70.1***	74.8***	73.0***	73.9***	72.9***	69.4***	71.4***	71.5***
4 Sessions											
1&2&3&4	62.5**	75.7***	61.7**	69.0***	75.3***	73.0***	64.5***	67.4***	63.5*	68.3***	68.1***

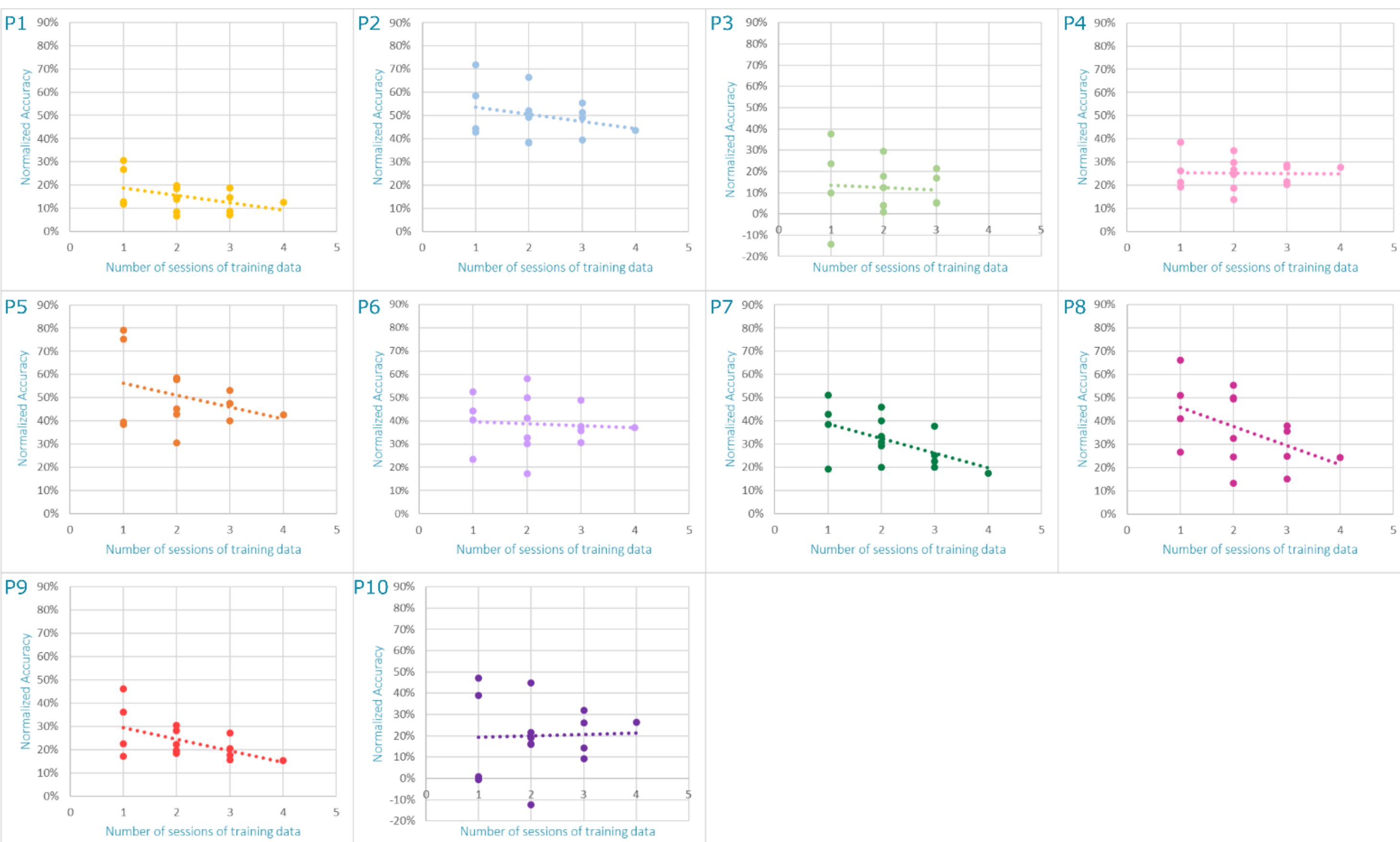


Figure 167: Average cross-validated classification accuracies, normalized against the chance-level threshold for the corresponding number of sessions of training data*, plotted against the number of sessions of training data for each participant. The trend of normalized classification accuracy in relation to number of sessions of training data is displayed.*

4 Discussion

4.1 Feasibility of an NIRS-based Affective BCI for Children

Overall, the results of this study provide evidence that it is possible to discriminate emotional valence from hemodynamic activity in children during a bimodal emotion-induction task. The offline cross-validated classification accuracies for each of the four sessions, averaged across all participants (78.4%, 77.4%, 76.7%, and 76.9% for S1-S4 respectively), exceeded the theoretical levels of chance (65.0%, calculated from the binomial distribution for a bi-class problem with n=30 samples from each class) (117). Out of all 40 sessions, there were only two sessions that did not exceed chance level. The average offline classification accuracies for each participant across their four sessions ranged from 70.0% to 85.3%.

This thesis is the first, to the author's knowledge, to investigate an affective brain computer interface for children using near infrared spectroscopy. The present results can be compared to affective NIRS-based BCI studies involving adult populations. Tai et al. (50) achieved higher classification accuracies, (range 75.0-96.7%, mean 84.6%); however, their classification task was between a presumably more distinct set of classes, emotional vs. neutral state, rather than between positive and negative valence. The results of the current study are comparable to that of Hosseini et al., (range 58.0-83.0%, mean 70.6%) (51) and Moghimi et al. (range 62.0-86.9%, mean 71.9%) (52) who both sought to classify positive and negative valence, but using unimodal stimuli; affective images and affective music respectively. Heger et al. (33) tackled the multi-class problem of discriminating emotional arousal as well as valence, and reported slightly lower accuracies than those of the current study (range 44.0%-74.4%). Yanagisawa et al. (53) saw considerable variability in the classification accuracy of their participants, with some participants reaching as high as 96.7%, but with approximately one-third of their participants not achieving above-chance results. Finally, Hu et al. (54) reported similar accuracies (mean, 73.8%) to the current study; however, their classification task was to discriminate between different subcategories of positive emotions rather than positive and negative valence. Overall, the results of the current study fit well with that of the existing affective NIRS-BCI literature, and although comparisons are limited considering the methodological differences, most notably the differences in the age of study populations and classification tasks.

The ability to differentiate positive and negative emotional states from hemodynamic activity in a brain-computer interface suggests underlying neurophysiological differences in the processing of emotional valence. In the literature, these differences have been attributed to two independent neural systems –

the approach system, which primes an individual for approach, attachment or appetitive behaviours, and the withdrawal system, which primes an individual for avoidance, flight or defense (27). Emotions can be seen as the by-product of an information processing system (119) that evaluates incoming sensory information to identify stimuli as a being either helpful or harmful for an individual to achieve their goals (28,29). Stimuli that produce negative emotions such as fear or anxiety would activate brain regions involved with the withdrawal system, facilitating avoidance of the aversive stimulus. Stimuli that produce positive emotions such as happiness or excitement would recruit the approach system, facilitating advancement towards the rewarding stimulus (120). Many studies have found evidence supporting this approach-withdrawal theory. Differential activation with respect to emotional valence has been found in the amygdala (121), the orbito-frontal cortex (122), and the dorsolateral prefrontal cortex (120) in fMRI and positron emission tomography (PET) studies, as well as in the theta and delta frequency bands of the EEG (123). There is evidence that while differentiable, positive and negative valence are not completely independent; substantial overlap has been demonstrated within these activation patterns (124). This could be attributed to the fact that positive and negative valence are not directly analogous to approach and withdrawal. For example, fear, a negatively valenced emotion, would stimulate the withdrawal system, while anger, also a negative emotion, may instead stimulate the approach system. Furthermore, individual differences in temperament can also affect the likelihood that either the approach or withdrawal systems will be activated in response to different affective stimuli (121,125). Overall, the approach-withdrawal model is one way to perceive differences in valence processing, but more work is required to fully understand the neurophysiological underpinnings of emotional valence.

4.2 Feasibility of Real-Time Emotion Prediction

A secondary goal of this thesis was to determine if emotional valence could be predicted in real-time using the developed affective BCI. Real-time emotion prediction would be crucial for future use of an affective BCI as an assistive communication device. In a natural communicative interaction, the BCI would need to operate asynchronously; that is, the user’s brain activity would need to be monitored continuously and be able to generate an output whenever an emotional state is identified (15). Investigating the feasibility of real-time emotion prediction is a necessary first step towards such asynchronous use of a BCI. In this study, capability for the BCI to predict emotional valence in real-time was tested in three online sessions. Data collected from the previous sessions were used to train a classifier for online prediction, and the classifier was retrained periodically throughout each session to incorporate the newly collected same-day data. The participants were provided with neurofeedback, a real-time visual representation of their predicted emotional state, to increase their awareness, attention, and to encourage them to use emotional regulation strategies to strengthen their responses.

The study demonstrated moderate success with real-time emotional prediction. Every participant was able to achieve above-chance accuracies ($p < 0.05$) with the online BCI at least once out of the three sessions, and eight of the ten participants achieved above-chance accuracies for two of the three sessions. However, there was a considerable amount of interparticipant variability in online BCI performance, with some participants barely achieving above-chance accuracies (e.g. P1) and others achieving online accuracies as high as 75% (e.g. P8). Overall, BCI performance seemed to improve across the three sessions (across-participant averages of 56.7%, 60.7%, and 62.4% for sessions 2, 3 and 4 respectively). It is generally expected for BCI performance to improve both as more training data are accumulated and the user learns how to better modulate their brain activity according to the provided neurofeedback, which in the case of this study would mean becoming more effective at emotional regulation (110,126). This trend of improvement across different session days was seen for some, but not all the participants. Several of the participants exhibited consistent or improving performance, with the exception of one anomalous session where they performed more poorly (e.g. Participant 2, who achieved accuracies of 68% in sessions 2 and 4 but only 55% in session 3). This inconsistency in performance warranted further investigation into interparticipant and intraparticipant (intersession) variability for each participant across different session days.

It appears that Aranyi et al. were the first to attempt to identify affective state in real-time using near-infrared spectroscopy. Participants were shown a virtual character and were instructed to interact with it using either positive emotions (55) or anger (56). They monitored the incoming NIRS signal for hemispheric asymmetry and defined classification success if there was a statistically significant increase in average asymmetry when the participants were 'interacting' with the character compared to the preceding rest trial. The real-time NIRS signal was also used to update the virtual character's 'behaviour'. For the positive condition, they found that 11 out of 17 participants achieved a statistically significant increase for an average of 50% of their trials. For the negative condition, 8 of 11 participants achieved this threshold for an average of approximately 67% of their trials. As in the current study, none of their participants completed all trials successfully. These results are comparable to those of the current study, suggesting moderate success with real-time emotion prediction. However, neither the accuracies reported by Aranyi et al. or those of the current study approach the classification accuracies seen in offline NIRS affective BCI studies, nor reach the 70% classification accuracy typically associated with ‘effective’ BCI use (127).

There have also been several studies with real-time emotion prediction paradigms using EEG as a neuroimaging modality. Daly et al. (128) sought to predict emotional valence and arousal in their participants in real-time as they listened to music. 7 of their 8 participants achieved greater-than-chance accuracies for their single online session. Iacoviello et al. (38) were able to discriminate a self-induced emotional state of disgust from neutral in real-time at an average rate greater than 90%. Ehrlich et al. (129) had participants modulate the valence of artificially generated music (auditory feedback) by recalling emotionally salient memories. Significant results were achieved for all but two of their 10 participants. Overall, these studies provide evidence that real-time emotion prediction is feasible, yet more work is needed to show reproducibility within population groups, neuroimaging modalities and experimental methodologies.

4.3 Interparticipant Variability

For the final aim of this thesis, a variety of factors assessing both interparticipant and intra-participant variability were investigated to understand their impact on BCI performance and help lay the foundation for future work in affective NIRS-BCI and paediatric BCI research. Interparticipant variability will be discussed in the remainder of this section. BCI performance varied considerably across participants, with some achieving online accuracies as high as 75%, and others struggling to reach the $p<0.05$ chance threshold of 58-60%. However, interparticipant variability in BCI performance is not uncommon; within BCI literature, there is consensus that there is no “universal BCI” system that works for every individual (126). Individual variations in brain structure, spontaneous physiological artefact production, and even variations in proficiency with different BCI training tasks or paradigms can all contribute to differences in BCI performance (126). Individuals who do not achieve sufficient accuracies with a BCI are often deemed “BCI illiterate”. However, this term can be misleading, as the issue has more to do with an incompatibility between an individual and a particular BCI system rather than a fundamental lack of ability to operate a BCI (130).

For the current study, a significant positive correlation was found between BCI performance and age. The three best performers in this study were all over 13 years of age, and comparably, the two lowest performers were 10 years of age or younger. The age of the participants in this study ranged from 8 to 14 years. A considerable amount of emotional, neurophysiological and cognitive development occurs across this age range that could all contribute to BCI ability. From the beginning to the end of grade school, a child's understanding and awareness of emotions becomes richer and more complex. They learn a wider range of emotional regulation strategies and are able to use them more successfully in everyday situations to adapt their behaviour according to their desired goals (72). It is possible that the older participants were better able to regulate their emotions in response to the visual neurofeedback than the younger participants, thus producing more reliable changes in hemodynamic activity for the BCI to classify. Also, the prefrontal cortex begins to develop early in childhood but continues to mature throughout adolescence and into young adulthood (76). It is possible that the more developed prefrontal cortices of the older participants evoked more reliable and distinct patterns of hemodynamic activity in response to affective stimuli. Executive attention networks and the ability to maintain attention over a sustained period of time also develop with age (131,132), meaning that the older participants were likely able to maintain their focus better over the course of the session than the younger participants, resulting in better performance with the BCI (34).

Sex was also investigated as a factor contributing to interparticipant variability in BCI performance. Although no significant differences were found in overall classification accuracy, sensitivity or specificity scores between female and male participants, females on average had higher sensitivity scores, which can be loosely interpreted as greater success at regulating positive emotions, and males had higher specificity scores, which could be interpreted as greater success at regulating negative emotions. There is evidence in the literature of sex differences in children in emotion expression, temperament and reactions to affective stimuli, lending to the hypothesis that in a study with a larger sample size, significant differences with regards to BCI performance and sex may have been seen. For example, McManis et al. (95) found that when viewing unpleasant images, girls reacted more defensively, with increased heart rate, skin conductance and startle reflexes than boys, who reacted with increased attention and interest. They also found that girls rated positive images as more pleasant than boys. In terms of emotion expression, girls have been found to express more positive emotions and internalizing negative emotions such as sadness or anxiety than boys, who tend to express more externalizing negative emotions such as anger and frustration (133). This aligns with sex differences in temperament, where girls tend to rate higher in traits of positive affect and sociability, while boys tend to rate higher in traits of negative affect, aggression and impulsivity (134). Overall, the literature suggests that males occupy a 'more negative emotional space', due to a combination of biological differences and the pressures of societal gender norms. This could translate to more pronounced changes in hemodynamic activity in response to negative emotional stimuli, or to an increased 'familiarity' with negative emotional responses, resulting in the higher specificity scores than females seen in this study.

We also looked to participants' temperament to investigate differences in BCI performance. Temperament refers to how an individual reacts to their environment, in terms of response latency, intensity, recovery and ability to self-regulate (102). Differences in temperament have been linked to differences in recruitment of the approach and withdrawal systems. The degree of recruitment of these opposing neural systems can impact the type and intensity of emotions evoked in response to a given situation or stimuli (125). It has been found that individuals with higher withdrawal or avoidant temperaments respond more strongly to negatively valenced words in an emotional Stroop task, while individuals with higher approach temperaments responded more strongly to positively valenced words in the same task (135), and that individuals with higher approach temperaments experience greater physiological changes in response to affective images than those with avoidant temperaments (136). In the context of this study, it was expected that differences in temperament in measures such as fear or surgency may reflect differences in emotional responses to negative or positive stimuli and lead to differences in performance with the BCI. Furthermore, effortful control, a dimension of temperament that refers to the ability to regulate and adjust the intensity of one's reactions, is linked to the ability to effectively regulate emotional responses (137). Children with low levels of effortful control have been shown to have reduced activation in the anterior cingulate cortex (ACC), a crucial component of the brain's emotional networks with extensive connections to the PFC, when viewing affective images (138). In the context of the current study, it was expected that participants with higher effortful control scores would perform better with the BCI. However, the only significant correlation between BCI performance and temperament was found between aggression and sensitivity and specificity, and between negative emotionality (combination of aggression, frustration and depressive mood) and sensitivity. The results of the correlation analysis suggest that participants who rated higher in aggression and negative emotionality were better at regulating/had a stronger response to positive emotions and were poorer at regulating/had a weaker response to negative emotions. It could be possible that participants who, by nature of their temperament tend to react with greater aggression and negative emotionality, have poorer control over these negative emotions and had greater difficulty regulating them for the online BCI sessions. However, these correlations were weak and should be considered tentatively. Unexpectedly, no significant correlation was found between effortful control and BCI performance.

The participants were also divided into two groups based on their overall level of performance to determine if there were any differences in temperament between 'high BCI performers' and 'low BCI performers. No significant differences were found between the two groups in any of the measures. Since a positive correlation was found between age and BCI performance, the high-performer and lowperformer groups were further divided by age to see if differences in temperament could provide insight on why two of the 'younger' participants were able to achieve 'high-performing' classification accuracies (P2 & P6), and alternatively, why two of the 'older' participants were in the 'low-performing' group (P3 & P4). No significant differences were found for any of the temperament measures between the older high-performing and low-performing groups, but the younger high-performing participants were found to rate higher in the temperamental measures of affiliation and depressive mood. It is possible that a high level of affiliation allowed these participants to empathize more and have a stronger response to the affective stimuli, many of which contained images of people either looking joyful and excited in the positive condition, or sad and scared in the negative condition. Further, their higher scores of depressive mood could suggest that negative affective stimuli evoked a stronger response in these participants than other children their age. However, once again due the small sample sizes these conclusions are only speculative.

Overall, the lack of significant correlations and weak correlations between temperament measures and BCI performance is likely due to the limited sample size. A greater number of participants is needed to identify patterns or trends in BCI performance with respect to temperament. It should also be noted that the temperament scores in this study were based on caregiver assessment. There can be a lack of convergence between a child's 'true' temperament and their caregiver's perception of their temperament (103), which could be influenced by their own expectations of their child's behaviour (134). In future studies, self-assessments of temperament, as well as a larger study population, would help to elucidate how temperament affects evoked emotional responses and the ability to self-regulate these responses in the context of an affective BCI.

4.4 Intersession (Intraparticipant) Variability

Considerable variability was seen within each participant's online BCI performance on different session days. For example, Participant 2 achieved classification accuracies of 68% in sessions 2 and 4, but only 55% in session 3, and Participant 3 achieved accuracies of 63.3% and 61.7% in sessions 2 and 4, but only 45% in session 3. Participants who achieved higher accuracies with the BCI also demonstrated more consistency across different session days in terms of which feature types were the most discriminatory features. Greater consistency in feature performance could imply that these participants produced more reliable, consistent hemodynamic signals in response to induced emotional states than the participants who showed inconsistency in feature performance. This variability can also be seen in the signal morphology of the trial-averaged hemodynamic response functions across different session days.

Furthermore, despite the fact that training a classifier on a larger data set should result in more accurate prediction models (15,111,139), offline classification accuracy decreased when training data from multiple sessions were combined. This also suggests that the hemodynamic response was not consistent across different session days.

Variability within the hemodynamic response is not uncommon in BCI literature. Holper et al. (140) found that a greater amount of intersession variability negatively impacted BCI performance in a motor imagery task. Power et al. (107) found that characteristics of the hemodynamic response varied across different session days in a mental arithmetic task, and Moghimi et al. (141) found inconsistencies in hemodynamic activity during repeated exposures to the same musical stimulus. Intersession variability has been attributed to participant-related factors such as changes in fatigue, attention, mood, motivation, and underlying spontaneous/baseline neural activity, to environmental factors such as auditory distractions, and to instrumentation-related factors such as deviations in optode placement and calibration (34,107,108,142–144).

To examine changes in fatigue, mood and attention on different session days, participant responses to an end-of-session mental state questionnaire were compared to BCI performance. At the group level, the questionnaire results suggest that participants performed more poorly on days where they felt the most fatigue during the session. Participants also performed more poorly on days where they experienced the greatest downward shift in mood throughout the session, and on days where they reported the greatest amount of frustration with the task. Observations at the individual participant level also provide evidence that participant mental state on different session days affected BCI performance. While these results align with what the literature says about changes in mental state impacting BCI performance (34,107,143), the questionnaire reports alone cannot claim causality. It cannot be concluded whether a decrease in mood and increased frustration was detrimental to task performance, or that the participants felt more frustrated and experienced a decrease in mood *because* they were struggling to obtain accurate classifications with the BCI. It is likely a combination of both; fatigue, a poor mood, or sense of frustration could decrease motivation and attention, leading to poorer task performance and lower classification accuracies, which in turn further worsened their mood and led to even more fatigue and frustration as they struggled to control the BCI. In studies of affective state during learning, it has been shown that while some level of frustration and confusion can lead to effective and beneficial problem solving, too much frustration can eventually lead to complete disengagement from the task (145).

Yanagisawa et al. (53) found that intensity of reaction to emotional stimuli explained some of the variability of BCI performance in their study; participants who rated the affective stimuli as highly salient were found to achieve higher classification accuracies in an emotion-induction based BCI task. For the current study, participants were asked to rate the perceived intensity of their emotions after each block of trials. It was expected that if participants rated a block of trials as highly intense, their evoked emotion would be stronger and thus more pronounced and easier for the BCI to distinguish, resulting in higher classification accuracies. However, no significant correlation was found between intensity rating and classification accuracy for any of the participants. This could be due to the fact that the same stimulus set was used for each participant for each session, so that by the end of the study the participants would have seen each image four times. Repeated exposure to an affective stimulus can eventually lead to habituation (146,147), so the increasing familiarity with the stimuli could have led the participants to perceive their evoked emotions as being less intense. Also, a single rating of intensity after every 12 trials with only a 5-point scale was collected. More detailed or more frequent ratings may be required to better assess trends between perceived emotional intensity and BCI performance. However, there is a trade-off between collecting more detailed information and minimizing session complexity; more frequent rating tasks or scale options would increase the cognitive burden on the participants, decreasing their ability to focus on the emotion regulation task (148).

4.5 Differences between Online & Offline Performance

Across all three online sessions, the offline cross-validated classification accuracies were consistently higher than the online, real-time classification accuracies. This is commonly seen in BCI literature (13), and indicates that the trained classifiers used for real-time prediction had difficulty generalizing to new, unlabeled data (111). There could be several reasons the classifiers exhibited poor generalizability. Firstly, the classifier used for a given session were trained on previous session data, but the study results have indicated that there is considerable variability in the hemodynamic response across different session days. This is supported by the fact that classifier performance tended to improve throughout the session as the classifiers were retrained to include more same-day session data. The last block of each session was on average, greater than the first block of each session, although this difference was not significant. However, each session was limited in length to 5 blocks to prevent participants from becoming too fatigued. It is likely that a further, significant improvement in classifier performance would be seen as more-same day data was collected, providing that participant attention could be maintained, as was seen with Rezazadeh et al.'s online covert speech BCI (85).

Secondly, the classifier training paradigm could have been prone to overfitting the data. As described in the methodology chapter of this thesis, dimensionality of the feature set is an important consideration in classifier model design. A larger number of features is not necessarily beneficial to classifier performance, especially when the training data set is small (109–111). Despite efforts to reduce dimensionality through feature selection and increase generalizability through regularization, it is possible that the dimension of the feature set (p=10 features) used in this study was still too large. Furthermore, the feature selection algorithm used in this study, sequential forward floating search, can be prone to overfitting (149). Using an embedded feature selection method such as the least absolute shrinkage and selection operator (LASSO) method could have allowed for feature set dimensionality reduction while being less likely to overfit the training data (150).

Finally, due to the nature of the experiment, there are likely samples within the collected data set that are *less than optimal*, possibly caused by shifts in attention, movement, distractions in the data collection environment, or fatigue. In general, the claim that more training data leads to more accurate classifiers is only true if the quality of the training data is high and truly representative of the phenomenon of interest. In the process of cross-validation, the data set is partitioned into subsets of training and test data. In repeated iterations of cross-validation, it can be seen that some of these subsets perform better, or are *more optimal,* than others. It is possible that higher online classification accuracies could be achieved if an *optimized subset* of the collected data is used for training the classifiers. Subset selection has already been investigated to minimize computational costs in machine learning problems where the training data set is extremely large (151–153). Future BCI work should focus on developing methods for optimizing the training data set to produce more robust and generalizable classifiers.

4.6 Limitations

Paediatric Population

Working with a paediatric population introduces several experimental limitations. Children have a significantly lower attention span and experience fatigue more quickly than adults (131), meaning the experimental sessions had to be kept as short as possible, limiting the amount of data that could be collected in a single session. Again, incorporating more same-day training data has been shown to improve BCI performance (107), as long as this data is high-quality, artefact-free and representative of the neurophysiological phenomenon of interest. Even within the shortened session, differences in attention could be seen between the younger and older participants. Younger participants had greater difficulty sitting still and required more frequent breaks than the older participants. This restlessness risks introducing artefacts in the acquired hemodynamic signal and producing unwanted shifts in the signal due to optode-scalp decoupling (108). Overall, a lack of attention or loss of focus can greatly detract from task performance (154) and is likely a significant limitation of working with a paediatric population for BCI research.

Furthermore, while school-aged children have been shown to understand and be capable of employing a variety of different emotional regulation strategies (72), emotion understanding and regulatory capabilities continue to develop along with their prefrontal cortical structures into adolescence and young adulthood (76). It is possible that the children in this study struggled with the task of regulating their emotions in response to the visual feedback. It is also possible that in this study, the children's emotional networks were not yet mature enough to evoke fully differentiable hemodynamic responses to positive and negative emotional stimuli. However, the study of emotions using NIRS is still in its infancy even in adult populations, and more work must be done to fully characterize the 'mature' hemodynamic response to evoked emotional states (47) before we can understand the limitations of immature emotional networks and the corresponding hemodynamic responses in children.

Saliency of Emotional Stimuli

In this study it was essential to ensure strong emotional responses were being evoked to produce reliable and differentiable hemodynamic signals. Although care was taken to personalize the stimulus set in this study for each participant to compensate for differences in affective stimulus saliency (57), and bimodal stimulation was used to enhance emotional stimulation (88), the emotional intensity ratings collected after each block suggested that the induced emotions were only of moderate intensity. Due to the young age of participants in the study, we were limited in the types of images that could be used as affective stimuli. The IAPS, GAPED and OASIS databases have much more salient images in their collections that could not be used for the current study for ethical reasons (e.g. dead bodies, extreme violence, erotica). It is likely that more pronounced brain activation would occur in response to more extreme stimuli, however, the emotional safety of the participants in the study was our priority.

The participants were exposed to each image/music excerpt in their stimulus set four times, once on each session day. It is possible that they experienced habituation to the affective qualities of the stimulus (146,147), resulting in weaker evoked emotional responses with consecutive sessions. However, if this were the case, the intensity ratings would likely decrease from session to session. No such trend was observed, rather, the intensity ratings were relatively consistent even on different session days. Also, although each additional session meant an additional exposure to the same stimulus set, it also meant additional time to practice emotional regulation strategies in response to the online feedback, mitigating habituation effects.

Instrumentation

There are also limitations related to the instrumentation used in this study. NIRS is relatively robust to motion artefacts compared to other neuroimaging modalities, provided that the optodes remain in good contact with the scalp. Changes in pressure between the optodes and the scalp can lead to changes in light coupling and can contaminate the signal (41,108). Good optode-scalp contact can be maintained by ensuring a tight fit with the headpiece holding the optodes and the scalp. However, there is a trade off between a tight fit and headpiece versatility and comfort. When dealing with a paediatric population with variable head sizes and a lower tolerance for discomfort, this design issue was of considerable importance. Hair colour, hair density and skin pigmentation can also impact the signal-to-noise ratio of the hemodynamic signal (41). Furthermore, while optodes are arranged on the forehead according to the 10-20 International system of electrode placement, slight deviations in optode set up on different session days, as well as differences in head shapes of participants could introduce variability due slightly different brain regions being monitored (108).

Another significant limitation of NIRS instrumentation is that this neuroimaging method only allows for superficial cortical detection, with a maximum depth of about 2-3cm into the skull (80,81). Many of the brain regions implicated in emotion processing are located deeper inside the skull, such as the amygdala, hypothalamus, the anterior cingulate cortex, and even the ventromedial prefrontal cortex. A neuroimaging method providing greater depth of resolution, such as fMRI or MEG, may provide greater insight into the activation that occurs in the brain in response to positively and negatively induced emotions.

Cerebral Hemodynamic Response

The long latency of the cerebral hemodynamic responses is another limitation of using NIRS as an imaging modality for brain-computer interfaces. The slow hemodynamic response, the result of changes in cerebral blood flow and metabolism, has an onset of about 5-8 seconds (83). A time window of 20s was used in the current study to fully capture this response. However, such a long trial period considerably increased the length of the sessions and limited the amount of data that could be collected before participants grew fatigued. Studies using EEG, which has a higher temporal resolution, can collect much more data in a single session. To shorten NIRS data collection sessions, it would be worth investigating the *initial dip* phenomenon. The initial dip is the term given to the fast hemodynamic response (up to about 2 seconds), which refers to the initial drop in concentration of oxygenated hemoglobin that is observed as localized activated neurons rapidly metabolize oxygen before an increase in blood flow delivers more oxygenated hemoglobin to the region (155,156). The initial dip has been observed in a variety of brain regions including the visual, motor, auditory and prefrontal cortices, although to the author's knowledge no studies have investigated the initial dip in induced emotional states (156).

While efforts were made to suppress physiological noise in the hemodynamic response using a low-pass filter, there may be some systemic artefacts remaining in the collected signal. The hemodynamic response is especially sensitive to systemic fluctuations in blood flow within the scalp. Some algorithms have been investigated to separate superficial blood flow effects from the cortical hemodynamic signal, including multi-distance independent component analysis (MD-ICA). MD-ICA seeks to separate the hemodynamic signal into cortical and superficial components based on distance, using source-detector pairs that correspond to different measurement depths into the skull (157). However, other studies have suggested that these systemic artefacts have little effect on signal classification and BCI performance (158). Overall, more work is needed to investigate the effects of systemic physiological noise on BCI performance, especially in a paediatric context in an emotional induction task.

Statistical Power

This study was also limited by a small sample size, reducing the statistical power to draw conclusions on several aspects of BCI performance. Testing the developed BCI on a larger population would strengthen the validity of the results and allow for better investigation of trends in age, sex, temperament, fatigue, mood, perceived effort and frustration in relation to BCI performance.

5 Conclusion

5.1 Summary of Study Contributions

The main objective of this thesis was to determine the feasibility of a brain-computer interface that could detect emotional valence in children using near-infrared spectroscopy. This thesis demonstrated that it is possible to differentiate positive and negative emotional valence from changes in hemodynamic activity during an emotion-induction task, with an average accuracy of approximately 77%. This discrimination was possible with all ten participants and across four different session days, and the results from this study were comparable to existing affective NIRS-BCI studies conducted in adults. The ability to discriminate positive and negative emotional valence has been attributed to underlying differences in activation patterns that indicate the recruitment of so-called approach or withdrawal systems in the brain in response to rewarding or aversive stimuli.

This study also demonstrated the feasibility of real-time emotion prediction using the developed BCI. During the online sessions, participants were instructed to monitor the visual neurofeedback provided to them and regulate the intensity of their emotional responses to the affective stimuli accordingly. While above-chance level online classification accuracies were achieved for most of the participants in at least two of their three online sessions, none of the participants surpassed the 70% accuracy threshold considered ‘effective’ for BCI use. The differences between offline and online classification accuracies demonstrated in this study indicate that there is considerable room for improvement in the online prediction of emotional valence. Intersession variability in the hemodynamic response likely played a part in limiting the success of online prediction, as online classifiers were trained on data from the previous session days. Changes in mental state (such as fatigue, decrease in mood, and loss of attention) within and across different session days have been shown to impede BCI task performance. In the future, steps should be taken to improve the generalizability of the affective state classifiers, so they can correctly classify real-time, unlabeled data and better accommodate for intersession variability.

Considerable variability in BCI performance across participants was also found in this study. Age, sex, and temperament were all considered as factors contributing to interparticipant variability. BCI performance was found to increase with age, likely due to the increased attention span, maturation of the prefrontal cortex and a better ability to regulate emotions that develops with age. Limited conclusions could be drawn about BCI performance with respect to sex and temperament due to the small sample size of this study. Overall, more work is needed to elucidate trends in sex and temperament with respect to affective BCI performance.

This study was also susceptible to some methodological limitations, the first of which included conducting the study with a paediatric population. Children have greater difficulty maintaining attention in BCI tasks than adults, limiting the amount and quality of data that can be collected in one session. They also may have struggled with understanding or completing the task of emotional regulation. The study was also limited by the saliency of the emotional stimuli that could be used, for ethical reasons, to induce emotions in children, and exposure to the same set of affective stimuli in each session could have lowered emotional saliency due to habituation. In terms of instrumentation, there is a trade-off with NIRS between comfort and versatility of the headpiece used to mount the optodes and preventing scalp-optode decoupling. The imaging depth of NIRS into the skull, latency of the hemodynamic response and systemic artefacts in the collected signal were also areas of limitation, as was the small sample size of this study.

Overall, to the author’s knowledge the research described in this thesis was the first of its kind to investigate an affective near-infrared spectroscopy brain-computer interface for a paediatric population. It is also one of the first studies to conduct online prediction of affective state from hemodynamic activity. This study also demonstrated the impacts of interparticipant and intersession variability of the hemodynamic response on affective BCI performance.

5.2 Future Work

In moving towards an affective NIRS-based BCI to be used as an access pathway to communication, the following steps for future work are recommended:

1. The most immediate future work should focus on improving the online classification paradigm for real-time emotional valence prediction. Using the proposed BCI as a communication device would require better real-time prediction accuracy than what was demonstrated in this thesis. Improving generalizability of the classification paradigm for online classification could include developing methods to identify optimal subsets of training data. Seeing how changes in mental state, baseline hemodynamic activity and instrumentation factors can all introduce variability into the collected data, it is likely that certain training samples will better represent the underlying emotional neurophysiological response than others. Optimizing the training data set would ensure that only the highest quality data is being used to develop the classifier model.

2. Future work should also consider making BCI data collection sessions more engaging. More engaging sessions would ensure that younger children would be able to maintain focus for the

entire training session, despite their still-maturing attentional networks, allowing for higher-quality of collected data and likely yielding better classification performance. More engaging sessions would be necessary for BCI research to extend to even younger participants than the age range investigated for this study.

3. In a real communicative interaction, there would be no explicit stimulus evoking an emotional response. Instead, emotions would be self-generated, the combined result of a variety of external and internal emotionally salient sensory cues. The next step would therefore be to determine if real-time emotional valence prediction can be replicated using self-induced emotions rather than stimulus-evoked emotions.
4. Finally, the affective BCI would need to be tested with the target population, children with physical disabilities. In the current study, the BCI was tested with a typically developing population to control for the considerable amount of variability seen in conditions, disorders, diseases and traumatic injuries that can result in physical impairments. Children with physical disabilities may also have associated delays or impairments in cognitive or emotional processing that could affect their ability to operate an affective brain-computer interface.

Online affective NIRS-BCI studies should also be conducted in larger sample populations to better elucidate how age, sex and temperament and other interparticipant factors affect BCI performance. The literature would also benefit from a better understanding of the neural correlates of valence in emotion processing, and how these mature from childhood to adulthood.

5.3 Significance

A brain-computer interface that can differentiate between positive and negatively valenced emotional responses could be used to detect these emotional states in children with severe physical disabilities who have had limited success with existing assistive technologies for communication. The binary detection of emotional valence could be used to express feelings, preference, or even affirmative/negative responses to questions without the needs for words or other developed language abilities. This would provide children with severe physical disabilities with an access pathway to communication that circumvents the need for functional motor control and literacy skills. With access to communication, these children can engage within their communities, learn how to advocate for themselves, gain independence, and overall improve their quality of life.

6. Journal Article

The following section presents a journal article prepared for publication in a peer-reviewed journal such as Journal of Neural Engineering or NeuroImage, based on the work in this thesis.

6.1 Introduction

6.1.1 Motivation

Children with complex communication needs are at risk of experiencing significant challenges in their daily lives that can translate to delays in cognitive development, can detriment their formation of social relationships, their educational success and their ability to gain independence (2–4). Inadequate access to communication prevents these children from participating fully in their communities and from being included in society (4). Early intervention with alternative and augmentative communication (AAC) technologies is crucial for mitigating these challenges. AAC devices circumvent challenging physical or cognitive aspects of communication, allowing the individual to focus on the goal of communication itself (5). However, despite the importance of early intervention, much of the existing AAC technology is not designed with children in mind (6). A high rate of initial device abandonment (30%) and an even higher rate of failure to use devices long term (75%) with children suggests that AAC technologies are not appropriately addressing their communication needs (7). For children born with cerebral palsy, neurodegenerative disorders or who have had traumatic brainstem injury, severe motor impairments can affect their ability to operate an assistive communication device (8), further limiting their options for AAC technologies. Recently, brain-computer interfaces (BCIs) have attracted attention as an access technology given their ability to delineate communicative intent of the user through the monitoring and analysis of brain activity (12,13). Most BCI research to date, however, has focused on typically developing adults, with little work done to investigate BCIs for children (14). In this study, we explored a solution that could address the basic communication needs of children with severe physical disabilities through the development of a brain-computer interface to identify emotional states.

6.1.2 Background

6.1.2.1 Brain-Computer Interfaces

Brain-computer interfaces bypass the need for voluntary motor control by directly analyzing brain activity. To operate a BCI, the user will conduct a mental task or attend to a stimulus that has been chosen to represent a certain communicative or functional intent. Their neurophysiological signals are measured using an acquisition modality such as electroencephalography (EEG) or near-infrared spectroscopy (NIRS). The acquired signal is analyzed to identify the physiological phenomenon manifested in the signal that is evoked during the mental activity or in response to the stimulus (14). After the collected signal is processed, features characterizing the phenomenon of interest can be extracted and fed into a classification algorithm. Based on these features, the classification algorithm will assign the acquired signal to a corresponding control command (15) which can then be used to drive a communicative or other functional application. Many BCI systems designed for communication use spelling paradigms (e.g. the P300 speller) and have achieved acceptable accuracy levels, but have a limited information transfer rate (about two words per minute), rendering them fatiguing and frustrating (15,16). For children with physical disabilities who may already be facing associated delays in language and cognitive development, we may need to look beyond spelling paradigms and consider an alternative approach.

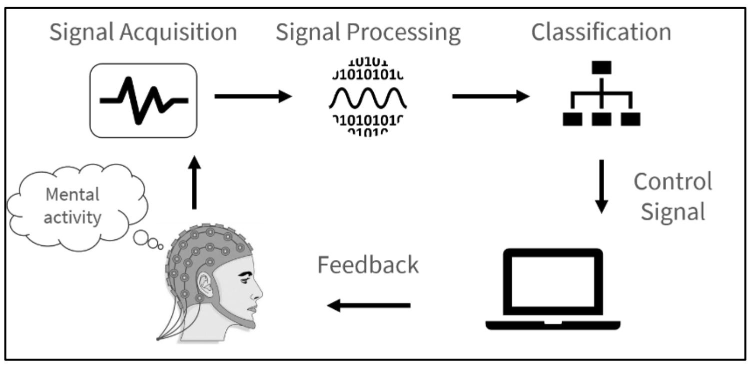


Figure 17: A schematic overview of a brain-computer interface. *The neurophysiological brain signal that occurs as a result of a mental task or activity is acquired, digitized and then processed and analyzed to produce a control signal that drives an application.*

Emotion as an Access Pathway

The inability to communicate basic needs, wants and emotions can incite intense feelings of fear, frustration, isolation and despair (18). Emotions underlie many of our basic needs, wants, preferences and opinions about others and our environments, and are closely linked to cognitive processes and memory (26). If emotional state can be accessed through a device like a BCI, it could circumvent the need for words to communicate.

Emotion Systems in the Brain

Traditionally, emotions were modeled discretely, categorized into separate entities (e.g. happiness, sadness, anger) that each have their own distinct physiological 'fingerprint' (20). However, the vast amount of variation found in emotional states across individuals and instances of these discrete categories (20,21) has steered researchers towards dimensional models of emotion such as the *circumplex model* (159). The circumplex model postulates that any emotional state falls somewhere along two fundamental dimensions of *valence*, the degree of pleasantness, and *arousal*, the degree of activation (159). For example, anger may be an emotional state with low valence and high arousal, while sadness would be of low valence and low arousal. There is emerging evidence of the underlying neurophysiological and behavioural correlates of valence and arousal, providing support to dimensional models of emotion in the literature (22,23). Fitting with the dimensional models, a meta-analysis of neuroimaging studies has demonstrated that large functional networks involving several different brain regions and structures are responsible for emotion processing, opposed to the simple 'one-to-one' mapping of brain structures typically associated with discrete models of emotions (21).

After exposure to an emotionally salient stimulus, limbic structures such as the amygdala, orbitofrontal cortex (OFC) and the anterior insula integrate the incoming sensory information with memories of the stimulus to evaluate its emotional value. From the amygdala, signals are distributed to the hypothalamus and the brainstem, where autonomic and endocrine responses are directed. There are also extensive connections between these limbic structures and the prefrontal cortex (PFC), where higher-level processes are activated in response to the contextualized emotional stimulus (25,27,28).

It is the activation of the prefrontal cortex in emotion processing that would allow emotions to be exploited for a brain-computer interface, as other limbic structures are located too deep inside the skull for superficial detection by portable neuro-imaging modalities (19). The PFC is essential for evaluating the emotional significance of a stimulus, interpreting and regulating emotional experience, and directing subsequent behaviours (29,30). The PFC can be understood to be evaluating the *core affect* of a stimulus – whether it is rewarding or threatening, and if it should therefore be approached or avoided, accepted or rejected (28–30). For an extensive review of the role of the PFC in emotion processing, please refer to Dixon *et al.* (29).

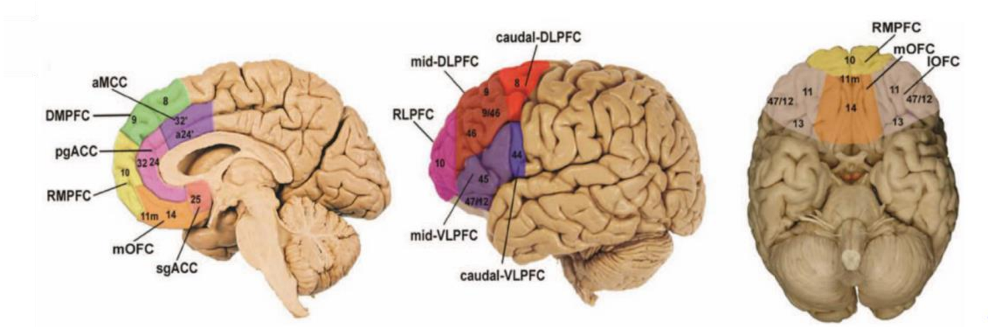


Figure 18: Parcellations of the prefrontal cortex*: aMCC = anterior mid cingulate cortex; DMPFC = dorsomedial prefrontal cortex; pgACC = pregenual anterior cingulate cortex; RMPFC = rostromedial prefrontal cortex; mOFC = medial orbitofrontal cortex; sgACC = subgenual anterior cingulate cortex; VLPFC = ventrolateral prefrontal cortex; RLPFC = rostrolateral prefrontal cortex; DLPFC = dorsolateral prefrontal cortex. Refer to Dixon et al. (29) for a complete review of the roles of these structures in emotion processing.*

Affective Brain-Computer Interfaces

Brain-computer interfaces that seek to detect and interpret emotional state from neurophysiological signals are known as affective brain-computer interfaces (aBCIs). Many of the existing aBCI studies have attempted to decode emotional states from electroencephalogram in typically developing adults. There is significant variability in the methodology of these studies, from the way emotions are defined and how emotional responses are elicited, to what features of the EEG are extracted and the algorithms used to classify these features into emotional states. Thus, unsurprisingly, these affective BCI studies report a wide range of results and levels of BCI performance. For thorough reviews of affective BCI research, please see Mühl et al. (19) and Liberati et al. (160).

6.1.2.3. Functional Near-Infrared Spectroscopy

Mechanisms of Near-Infrared Spectroscopy

Near-infrared spectroscopy uses light in the near-infrared range (~700-1200nm) to measure the hemodynamic activity of the brain. Near-infrared light is transmitted from a light source (e.g. LED, laser) through the tissues of the head and scalp and is absorbed by oxygenated and deoxygenated hemoglobin in cerebral blood. Oxygenated and deoxygenated hemoglobin (HbO & Hb) are biological chromophores whose spectra change with respect to their oxygenation state, which in turn fluctuates according to the brain's metabolic demands (40). Unabsorbed light is scattered throughout the tissue of the brain, some of which is eventually reflected out of the head and can be measured by detectors on the scalp. The change in concentration of oxygenated and deoxygenated hemoglobin is calculated using a modified version of the Beer-Lambert law, which is described in Scholkmann et al. (41) and Coyle et al. (42). This change in concentration of HbO and Hb can be directly related to brain activity. The most typical trend is that neuronal activity in a region of the brain increases the metabolic demands of that area, stimulating an increase of blood flow to the brain and resulting in an overall increase in [HbO] and a decrease in [Hb] (40–43).

NIRS & the Study of Emotions

Due to its relative low cost, ease of use and robustness to noise, near-infrared spectroscopy has been increasingly used for the functional mapping of brain activity (43,44). Despite the limited imaging depth of fNIRS (about 1-2 centimeters into the scalp), it remains a viable option for investigating the activation of the prefrontal cortex in studies of emotion processing (46). fNIRS offers a greater temporal resolution than fMRI, which is a considerable advantage when investigating responses as dynamic as emotion (47). For detailed syntheses on the use of fNIRS to study emotion, please refer to Bendall et al. (46), Balconi &

Molteni (47), and Doi et al. (48). fNIRS has also been extensively used to study emotion processing and the development of emotional networks in children (49).

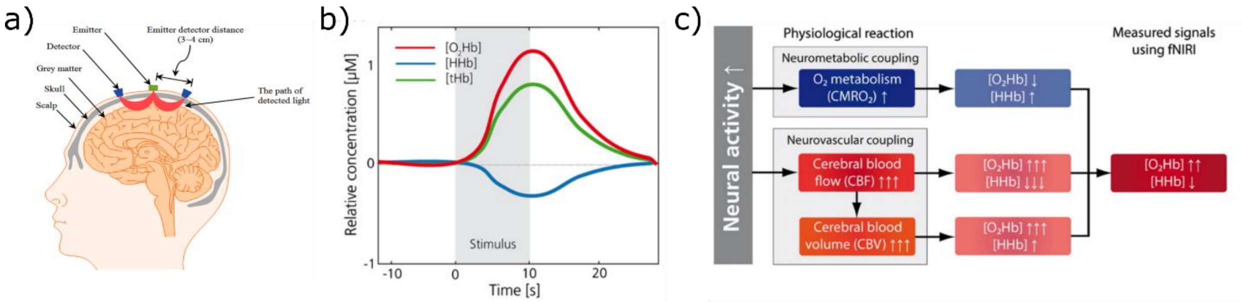


***Figure 19: a) An example of two near-infrared light emitter-detector pairs and the corresponding path of light** through the tissues of the head and into the gray matter of the cortical region of the brain. The light is hypothesized to follow a crescent-shaped path between emitter and detector, and can reach only a few centimeters into the skull, to the upper regions of the cerebral cortex. Modified from Naseer & Hong, (79); **b) Typical hemodynamic response due to increased brain activity;** a) typical signal of an evoked neuronal hemodynamic responses; and b) an overview of the hemodynamic response to neuronal activity, indicating the increased cerebral blood flow and increased rate of metabolism occurring in the brain during periods of activity, overall resulting in an increase in [HbO] and a decrease in [Hb]. Taken from Scholkmann et al. (41).*

NIRS Affective BCIs

There have already been several studies investigating fNIRS as a signal acquisition modality for an affective brain-computer interface. Tai et al. (50) appear to have been the first to attempt single-trial classification of emotional state from hemodynamic activity of the brain. They were able to discriminate emotion-induced brain activation from a baseline state with at least 75% accuracy for all their participants. Hosseini et al. (51) and Moghimi et al. (52) differentiated positive and negative emotional states using affective images and excerpts of music respectively. Heger et al. (33) used bimodal stimuli to induce emotions, and were able to discriminate both valence and arousal at rates significantly above chance. Yangisawa et al. (53) found that individual performance correlated with subjective ratings of the affective stimuli. Hu et al. (54) investigated three different subsets of positive emotions – encouragement, playfulness and harmony. Aranyi et al. designed an experiment where users could interact with a virtual character by expressing either positive emotions (55) or anger (56). It is important to note three key distinctions between these studies and the present work. First, most of these studies used unimodal stimuli (either affective images or music) to evoke emotional responses. Second, all these studies were conducted with a population of typically developed adults. Finally, all but two of these studies were conducted offline, meaning the data were collected, analyzed and BCI performance was evaluated post-hoc.

6.1.2.4 Self-Regulation & Neurofeedback

Evoking strong emotional responses and thus consistent brain signals from all participants in an affective BCI can be a challenge given the subjectivity of an individual's responses to affective stimuli (57). This variation could be accounted for by tailoring the stimuli material to each participant and collecting subjective ratings of their emotional responses to each stimulus. However, there can also be subjectivity in an individual's awareness and understanding of their own feelings, or their *emotional awareness* (58). The practice of *emotional regulation* can be used to heighten one's emotional awareness and provide a means of controlling or adjusting an emotional response in an adaptive way (59). Emotional regulation can affect the intensity, time course, quality and type of emotion experienced (60). There are many types of emotional regulation strategies, including *controlled generation*, where the individual uses mental imagery and reasoning to attribute emotional meaning to a stimulus and *controlled regulation*, where the individual actively reinterprets the meaning of the stimulus to alter its emotional saliency (59,61,62). These strategies have also been shown to activate regions of the prefrontal cortex involved in emotion processing (62). Using these strategies to intensify emotional responses could be a way to strengthen prefrontal cortex activation, producing more robust signals for an affective brain-computer interface. Emotional regulation could be incorporated in an affective BCI by providing the individual with neurofeedback, as in some sensory representation (e.g. visual or auditory) of their real-time emotion-based neural activity. Neurofeedback facilitates the conscious modulation of brain activity, and is based on the principles of operant conditioning (63). In clinical settings, neurofeedback has been used to train patients to self-regulate their own brain activity in a way that might ameliorate a particular behaviour or 'rewire' certain pathological networks, and has led to long-lasting functional reorganization of the brain (64,65).

6.1.2.5 Neurodevelopment of Emotion

Emotions are essentially adaptive behavioural mechanisms that that help us adjust our behaviour in response to incoming sensory information to achieve our goals. This adaptive role of emotional processing is present already early in infancy (e.g. babies crying in response to hunger, stimulation or changes in their environment) (72). During infancy and early childhood, emotions are regulated externally by caregivers and other adults. Caregivers soothe and manage emotional distress and arousal, monitor sensory input, and model socially appropriate expressions of emotion (72–74). Children attend to the emotional cues of their caregivers and begin to learn how to regulate their own emotional states. This increasing amount of self-control is facilitated by developing cognitive abilities, language skills, increasing emotional awareness and understanding, and a heightened sense of self (72–74). By the preschool age, children understand which basic emotions will be evoked in specific situations, know the consequences of different emotional responses, and can employ some simple emotional regulation strategies. By grade-school, children can rely on a variety of strategies to regulate their emotions successfully (72,73).

These developments in emotional awareness, understanding and regulation are paralleled by the functional neural development of the cortex throughout childhood. It is believed that in early infancy, emotions are directed mainly by *cortical excitatory processes* (e.g. the cause of sudden outbursts of crying) associated with sympathetic and parasympathetic regulation (72). These processes decline as *cortical inhibitory controls* emerge with the development of the frontal lobe and other cortical networks that facilitate more complex cognitive abilities (71,74,75). These areas associated with higher-order cognitive abilities and executive functioning, such as the prefrontal cortex, start to develop within the first year and continue to mature throughout childhood (71). This maturation process involves a reduction in neuronal density, synaptogenesis, branching of dendrites and increased myelination (71,76). Overall, these areas become more fractionated, developing into specialized functional networks that can carry out the complex cognitive tasks that are required for a comprehensive awareness and understanding of emotions and effective emotional regulation (76).

6.1.3 Study Overview

The primary goal of this research was to determine the feasibility of a brain-computer interface that could detect emotional valence from changes in hemodynamic activity in the prefrontal cortex of children. The BCI was developed and trained, and then tested in an online paradigm. Interparticipant and intersession variability was explored as well.

6.2 Methodology

6.2.1 Participants

10 typically developing children between 8 and 14 years old (mean age 11.5 ± 1.8 years, 3 males) participated in this study. Participants were screened for any neurological, psychological, cardiovascular or respiratory conditions, as well as for any history of brain injury or emotional trauma. Written consent was obtained from all participants prior to data collection.

6.2.2 Study Design

First, data was collected to train the BCI. The participants attended a single session where they underwent a series of emotion-induction trials. On each trial they were presented with a set of emotionally salient auditory and visual stimuli while their hemodynamic activity was measured using NIRS. The collected hemodynamic signals from the emotion-induction trials were labeled and used to train and test a classifier to recognize valence. After the BCI was trained, the participants were invited back for three more sessions to test the BCI online. These trials were identical to the first session, except for the presentation of visual feedback representing their emotional state as predicted by real-time classification results of the BCI. The participants were instructed to monitor the feedback and use emotional regulation strategies to strengthen their emotional response.

6.2.3 Instrumentation & Data Collection

Cerebral hemodynamic activity was measured using the Hitachi ETG-4000 NIRS system (Hitachi Medical Systems, Tokyo, Japan). A 3x5 grid of 8 light emitters and 7 light detectors was secured over the prefrontal cortex using a custom-made headpiece, with the bottom row of optodes sitting just above the eyebrows and centered at the nose. Detectors 11 and 12 (figure 4) were approximately aligned with the Fp1 and Fp2 sites of the 10-20 International System of electrode placement (77,78). Each emitter and detector were separated by 3cm, corresponding to a measurement depth of 2-3cm (79–81), reaching the cortical surface (82). This arrangement resulted in 22 integrated measurement sites, as indicated in figure 4. Data were sampled at 10Hz.

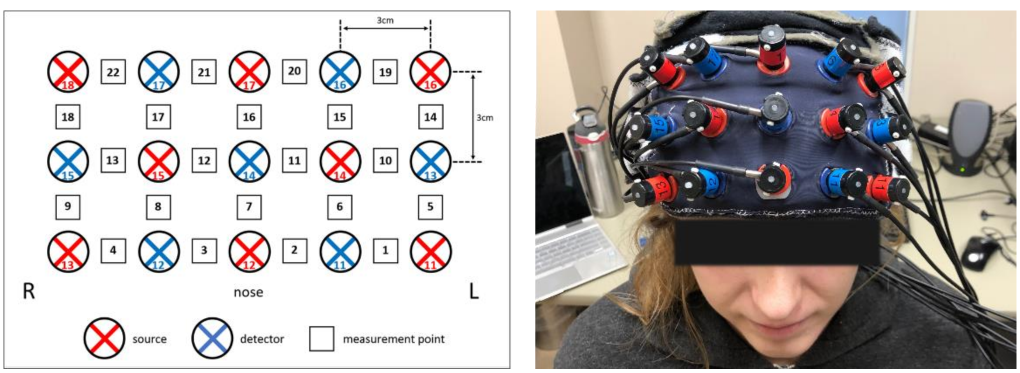


Figure 20: NIRS optode configuration over the prefrontal cortex. *The diagram on the left shows the configuration of the eight light sources (red) and eight light detectors (blue) arranged in a 3x5 grid over the forehead, resulting in 22 measurement sites, indicated by the black and white squares. L and R corresponding to the left and right sides of the forehead, respectively. The image on the right shows the optodes, mounted in the headpiece and placed over the forehead.*

6.2.4 Experimental Protocol

6.2.4.1 Session Structure

A single session was composed of five blocks of emotion-induction trials. Each block began with a 30-second baseline recording, during which the participants were instructed to relax and look at a neutral fixation point on the screen. Then, for each trial, a set of emotional stimuli was presented for a 20s response period (83). The set of stimuli used for a single trial was matched for valence and arousal. A prompt on the screen labeled each trial as 'positive' or 'negative', confirming for the participants the intended valence. For the first session, the participants were instructed to respond naturally to the stimuli. In sessions two through four, visual feedback was presented during the response period and the participants were instructed to use the feedback as a guide to regulate the strength of their emotional response. Each active response period was punctuated by a 20s rest period, allowing the hemodynamic activity to return to baseline levels (77,78,84–87). There were twelve trials within one block, for a total of 60 trials within one session. When a block was completed, participants would self-select when to proceed, allowing for a break. Each session took approximately 40 to 50 minutes to complete, and were approximately one month apart.

6.2.4.2 Affective Stimuli

The affective stimulus set included both visual and auditory stimuli presented simultaneously, as bimodal stimulation can enhance brain activation in emotion processing (88,89). The visual stimuli consisted of pictures drawn from three standardized databases: the International Affective Pictures System (IAPS), the Geneva Affective Pictures Database (GAPED) and the Open Affective Standardized Image Set (OASIS). These databases are collections of colour photographs from a wide range of semantic categories that have been reproducibly rated on their affective quality on scales of valence and arousal (90–92) and reliably evoke emotional responses in children (93–95). Prior to data collection, each participant rated the valence of each image in the selection to generate a personalized subset. The auditory stimuli consisted of 20s excerpts of music. Music has been shown to evoke strong emotional responses in listeners (96) and activate brain regions implicated in emotion processing (97). The excerpts of music were chosen from modern genres to ensure high saliency for the target population and were sampled without lyrics to reduce brain activation due to mental singing. The excerpts were rated by a music therapist for valence and arousal based on their tempo and mode (98,99). One 20s musical excerpt and 5 affective images, matched for valence and arousal, comprised a stimulus set for a single trial. For each trial, the 5 images would be displayed sequentially for 4s each, while the musical excerpt played. Either a positively- or negatively-valenced set of stimuli was presented for each trial, in a counter-balanced and pseudo-randomized order. For the 20s rest period, a clip of brown noise was played and 5 affectively-neutral images were displayed as a control stimulus (52).

6.2.4.3 Visual Feedback & Online Classification Paradigm

The visual feedback for the online sessions was in the form of a vertical bar that filled with colour according to the predicted valence of the response (see figure 5). Each trial started with the bar at a neutral middle position and would increase in height if the predicted emotional valence was positive and decrease in height if the predicted valence was negative. The participants were instructed to try to increase or decrease the height of this bar as much as possible over the 20s trial response period, according to the prompted valence of that trial. The BCI would process, analyze and classify 2-second segments of incoming data from the current trial response period and update the height of the feedback bar based on the classification output. The specific height of the feedback reflected the probability that the incoming segment of the hemodynamic signal belonged to either the positive or negative valence class. The BCI would analyze each 2-second segment cumulatively; that is, the first segment would be

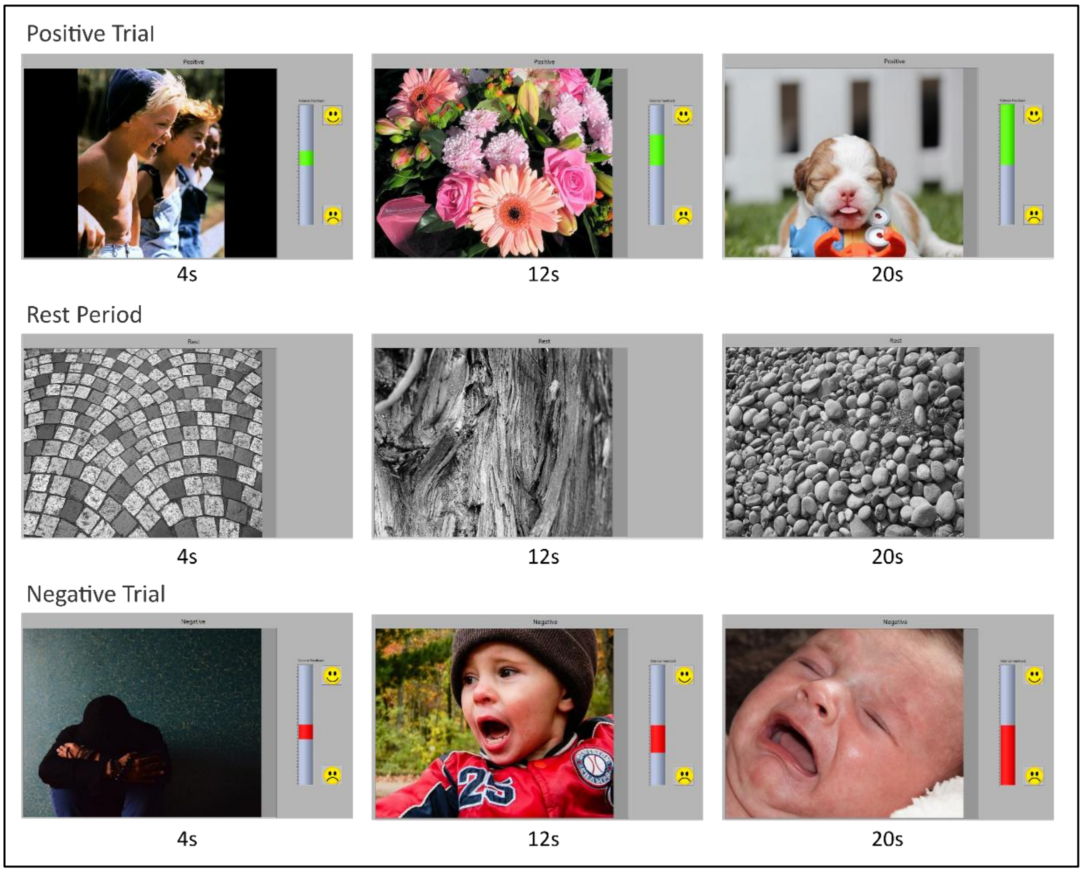


Figure 21: Screenshots of the experiment interface. *The top row shows three screenshots, taken at 4s, 12s and 20s respectively, for a positive trial; the middle row shows a neutral rest period; and the bottom row shows a negative trial. Each screenshot shows the displayed image at that time in the trial, the valence prompt above it, and the feedback bar on the right. The positive trial shows the feedback bar increasing in valence, and the negative trial shows the feedback bar decreasing in valence. The neutral rest period has no visual feedback.*

the first two seconds of the trial response period, the second segment would be the first four seconds, and so on until the classification of the entire 20s hemodynamic signal from that trial.

6.2.4.4 Questionnaires & Supplementary Data Collection

At the end of each block, the participants were asked to rate the intensity of the emotions they experienced for that set of trials using a modified version of the Self-Assessment Manikin (SAM), a visual tool for rating emotional experiences based on valence and arousal (100). At the end of each session, the participants were also asked to answer a short questionnaire about their subjective experience, which included questions on their mood, alertness, perceived effort and frustration experienced throughout the session. The parent-version of the Early Adolescent Temperament Questionnaire (EATQ), developed by Capaldi and Rothbart (103,104) and validated by Muris and Meesters (105), was used to assess the participants' temperament. Temperament, how an individual reacts to their environment within emotional, physical and attentional domains, can affect how they experience or regulate emotions (102).

6.2.3 Data Analysis

6.2.3.1 Signal Preprocessing

Within the acquired hemodynamic signal, there are several contaminating spontaneous oscillations of physiological origin that can produce variability that is not task or stimulus related. This physiological noise can usually be filtered by removing their respective frequency components from the acquired signal. Cardiac activity (0.8-1.2Hz), respiration (0.2-0.4Hz), and Mayer waves, or fluctuations due to arteriole pulsations (0.1Hz), are all physiological phenomena producing noise with frequency content higher than that of the hemodynamic response of interest (41,79,106). These effects were removed using a 3rd order type II Chebyshev low-pass infinite impulse response filter with a passband of 0-0.1Hz, transition band of 0.1-0.5Hz and a stopband cut off frequency of 0.5Hz, with a ripple of 0.1db and

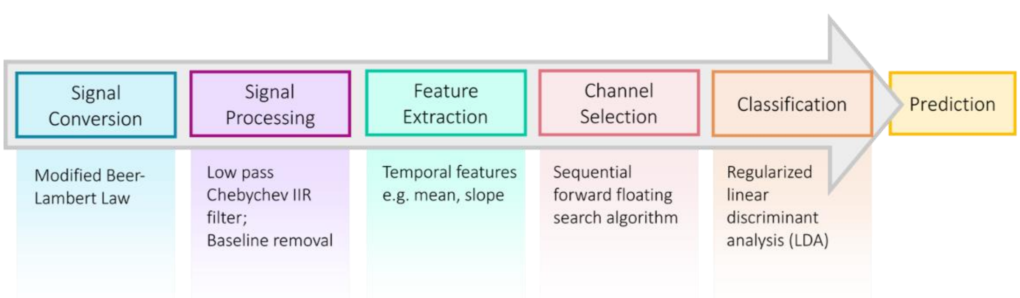


Figure 22: Data analysis pipeline *showing the processing steps for the incoming raw NIRS signal, ultimately yielding a prediction of the associated emotional valence after the classification step.*

minimum attenuation of 50dB (77,78,84–87). The mean of the 30s baseline recording for each block was calculated and subtracted from each subsequent trial in that block (85).

6.2.3.2 Feature Extraction & Selection

Temporal features that capture the morphology of the signal are commonly used in NIRS-BCI studies, with signal mean or signal slope being used in approximately half of the reported studies (79). These features can be calculated for each individual measurement channel, for each chromophore (HbO and Hb) and for any defined time interval. Based on preliminary analyses, seven temporal signal features were investigated – mean, slope, moving slope (4s window, 0.5s overlap), variance, root mean squared (RMS), skewness, and kurtosis, all calculated over the entire 20s trial response period. Multiple feature types, multiple chromophores and multiple measurement channels generates a high-dimensional feature set. It is likely that a high dimensional feature set contains redundant information, which can be detrimental to classifier performance (15,109–111). In this study, dimensionality reduction of the feature set was conducted in two steps. First, based on the offline data collected from session one, the feature type that yielded the highest classification accuracies was chosen as the most discriminatory feature for that participant. Secondly, a sequential forward floating search (SFFS) algorithm was used to determine the 5 best channels for each chromophore, resulting in a subset of 10 features. SFFS systematically searches through the available features to create an optimal subset, adding the best *l* features and removing the worst *r* features each iteration based on a fitness criterion (15,112). The Fisher criterion was used as the fitness criterion; refer to refs. (77,86,113) for more information on this method.

6.2.3.3 Classification

A linear discriminant analysis classifier (LDA) was used to discriminate the acquired hemodynamic signals based on their emotional valence. LDA is a commonly used classification algorithm in online BCI work due to its speed and low computational cost (15). LDA involves defining a linear decision boundary that separates the data into two classes, maximizing the distance between the class means while minimizing the variance within each class (15,79). To further mitigate the effects of a small training dataset, a regularized version of LDA was used (110,114). Regularized LDA involves defining a penalty parameter, gamma (γ), to reduce the high variance of models fitted to small training sets, thereby yielding a more generalizable model (114,115). Please refer to Rezazadeh et al. (85) for more details on the regularized LDA method. Potential values of gamma (from 0 to 1 in 0.05 increments) were assessed through 10-fold cross-validation, and the value yielding the highest accuracy was selected and used in the subsequent classifier model.

6.2.3.4 Online Classifier Retraining

A regularized LDA classifier was trained for each participant using their most discriminatory feature and 5 best channels of each chromophore based on the data collected from session 1. This classifier was then used to predict, in real time, the emotional valence of the incoming hemodynamic signals during the next online session. It has been shown that that same-day training data can improve classifier performance (85,107,108), so the classifier was retrained after each block to incorporate the new data from the current session. Retraining involved running all collected data through the data analysis pipeline, selecting 5 new 'best channels' for each chromophore, determining a new regularization parameter, and then training a new LDA classifier. This new classifier would then be used to predict the emotional valence of the incoming hemodynamic signals for the subsequent online block of trials. For each new online session, data from all previous sessions were combined to train the classifier.

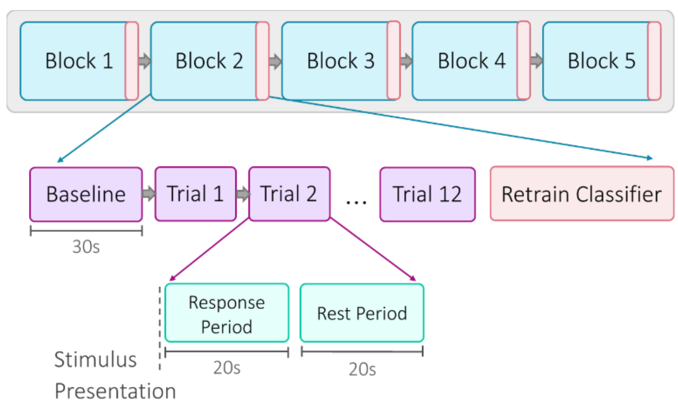


Figure 23: Overview of session structure for online sessions. *The classifier is retrained after each block of 12 trials. Classifier retraining involves selecting five new 'best channels' using SFFS for both chromophores, determining a new regularization parameter (γ) and training a new LDA classifier based on the accumulation of data from previous and current sessions. Session structure for the offline session was identical, only without the 'retrain classifier' step.*

6.3 Results

6.3.1 Session 1 Offline Results

The data collected from the first session was analyzed post-hoc using the pipeline described above, evaluating BCI performance for seven possible feature types. BCI performance was evaluated based on its classification accuracy, calculated using 10 iterations of 10-fold cross validation (78,116). The average cross-validated classification accuracy for each participant, using their most discriminatory feature, can be seen in figure 8. For most of the participants, slope or moving slope was their most discriminatory feature. Signal mean and slope are commonly reported as the most discriminatory features in NIRS-BCI studies (79). Interestingly, mean was the best feature for only 1 of the 10 participants. Variance, RMS, skewness and kurtosis consistently performed less robustly than mean, slope and moving slope as classification features. Using their best feature, every participant surpassed the chance level threshold of classification accuracy (65% for n=30 samples each class) (117).

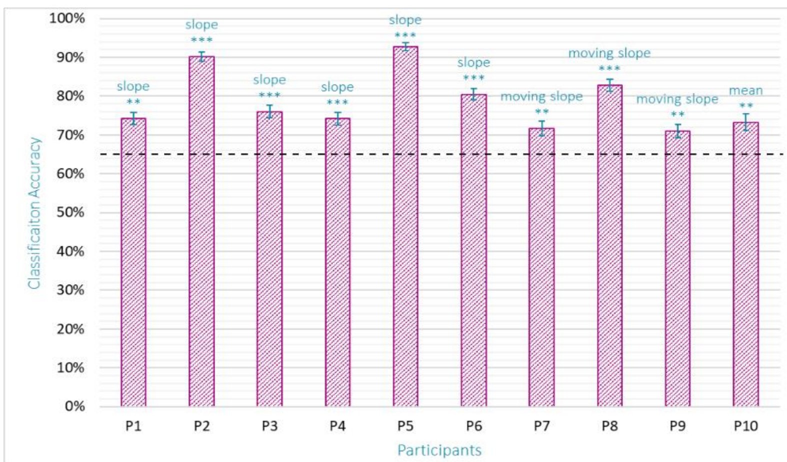


Figure 24: Cross-validated classification accuracies from session 1, using each participant's best feature. *Accuracies exceeding the upper limit of the 95%, 99% and 99.9% confidence intervals of chance are marked with *, ** and *** respectively, Combrisson & Jerbi (117).*

6.3.2 Sessions 2-4 Online Results

For the online sessions, BCI performance was evaluated based on the accuracy of the real-time predictions. A correct classification was tallied if the predicted classification matched the valence of the stimuli for that trial. The average online classification accuracies for each participant for each online session are displayed in table 1. All participants achieved above-chance accuracies for at least one of the online sessions, and 8 of the 10 participants achieved above-chance accuracies for two of the three online sessions. None of the participants achieved above-chance accuracy for all three online sessions.

BCI performance appeared to improve throughout a single session. On average, the last block of each session was more accurate than the first block of that session (e.g. $mean_{block\ 1,\ session\ 2}$ = 55.8% and $mean_{block\ 5,\ session\ 2}$ = 66.7%). However, these differences were not significant for any of the sessions ($p_{session2}$ = 0.28, $p_{session3}$= 0.12, $p_{session4}$ = 0.42). On a per-session basis, we see that only 3 of the 10 participants achieved above-chance online classification accuracies for the first online session, while 7 and 8 of the 10 participants achieved above-chance accuracies for the second and third online sessions respectively. On average, performance improved across the three online sessions (session 2 mean: 56.7%; session 3 mean: 60.7%; session 4 mean: 62.4%), which could be due to enhanced classifier performance from an increasing amount of training data and/or the participants becoming more familiar with the task of emotional regulation. However, a one-way ANOVA analysis of the average classification accuracies did not reveal a significant difference among the sessions (p = 0.38, Shapiro-Wilkes test for normality). This could be attributed to the fact that while we see an increase in online BCI performance for some participants (e.g. Participants 7 & 8), we see a decrease in performance for others (e.g. Participant 5), and others appear to have consistent performance, except for one anomalous session (e.g. Participants 2 & 3).

Table 8: Average online classification accuracies for each session, broken down by block*. The entire session average is bolded. Accuracies exceeding the upper limit of the 95%, 99% and 99.9% confidence intervals of chance are marked with *, ** and *** respectively.*

Session 2											
	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	Average
Block 1	41.7	41.7	66.7**	33.3	75.0***	58.3	58.3	33.3	50.0	41.7	50.0
Block 2	58.3	83.3***	66.7**	83.3***	83.3***	50.0	41.7	50.0	50.0	41.7	60.8*
Block 3	25.0	66.7**	66.7**	50.0	58.3	58.3	41.7	66.7**	50.0	75.0***	55.8
Block 4	75.0***	66.7**	66.7**	41.7	58.3	66.7**	25.0	66.7**	75.0***	41.7	58.3
Block 5	83.3***	83.3***	50.0	41.7	66.7**	50.0	33.3	50.0	41.7	83.3***	58.3
Average	**56.7**	**68.3****	**63.3***	**50.0**	**68.3****	**56.7**	**40.0**	**53.3**	**53.3**	**56.7**	**56.7**
Session 3											
	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	Average
Block 1	66.7**	58.3*	41.7	50.0	50.0	50.0	41.7	75.0***	75.0***	50.0	55.8
Block 2	75.0***	41.7	25.0	58.3*	58.3*	100.0***	75.0***	66.7**	75.0***	58.3*	63.3**
Block 3	41.7	75.0***	58.3*	50.0	58.3*	66.7**	83.3***	75.0***	41.7	66.7**	61.7**
Block 4	41.7	33.3	50.0	66.7**	58.3*	33.3	66.7**	58.3*	75.0***	75.0***	55.8
Block 5	58.3*	66.7**	50.0	83.3***	83.3***	75.0***	75.0***	83.3***	33.3	58.3**	66.7**
Average	**56.7**	**55.0**	**45.0**	**61.7****	**61.7****	**65.0****	**68.3*****	**71.7*****	**60.0***	**61.7****	**60.7***
Session 4											
	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	Average
Block 1	41.7	66.7**	75.0***	75.0***	41.7	33.3	58.3*	100.0***	58.3*	58.3*	61.1*
Block 2	66.7**	66.7**	75.0***	33.3	50.0	66.7**	83.3***	83.3***	66.7**	50.0	64.2**
Block 3	50.0	66.7**	91.7***	83.3***	66.7**	58.3*	58.3*	66.7**	33.3	50.0	62.5**
Block 4	66.7**	66.7**	16.7	50.0	41.7	66.7**	75.0***	58.3*	66.7**	58.3*	56.7
Block 5	66.7**	75.0***	50.0	50.0	83.3***	75.0***	75.0***	66.7**	66.7**	16.7	67.6***
Average	**58.3***	**68.3*****	**61.7****	**58.3***	**56.7**	**60.0***	**70.0*****	**75.0*****	**58.3***	**46.7**	**62.4****

6.3.3 Age

Figure 9 shows the average online classification accuracy for each participant, arranged by age from youngest to oldest. A significant positive correlation can be seen between online BCI performance and age (r=0.71, p=0.02). Three of the highest performers (P5, P7 and P8) were the three oldest participants, all at least 13 years old. The two lowest performers (P1 and P9) were also among the youngest participants, at ages 8 and 10 respectively.

6.3.4 Hemodynamic Response Functions

Figure 10 shows the trial-averaged hemodynamic response function (HRF) for Δ[HbO] and Δ[Hb] for Participant 4 for the first two sessions. The response signal from each measurement channel is shown mapped according to its position over the forehead. A clear distinction can be seen between the positive and negative response for both Δ[HbO] and Δ[Hb] in both sessions. However, both the positive and negative responses seem to vary across the different session days. This variability was seen for all participants.

6.3.5 Intensity Ratings

Participants were asked to rate the intensity of their emotions after each block of trials on a 5-point scale (from 1, representing 'not intense at all', to 5, representing 'extremely intense'). It was expected that a higher emotional intensity rating would indicate stronger evoked emotions for that block, resulting in a stronger hemodynamic response and yielding more accurate classification. However, no significant correlation between intensity rating and classification accuracy was found for any of the participants. In fact, the participants tended to rate each block consistently within and across sessions, with their ratings varying by only 1-2 points on the scale.

6.3.6 Session Experience Questionnaire

The responses to the subjective experience questionnaire on each participant's best and worst session days can be seen in Figure 11. Overall, participants were more tired at the end of the session on their worst session days and experienced a greater decrease in alertness from start to finish on their worst session days. On the day of their worst session, participants experienced a larger downward shift in their mood compared to the day of their best session. They used slightly more effort on their best session day compared to their worst session day and experienced greater frustration on their worst session days compared to their best session days.

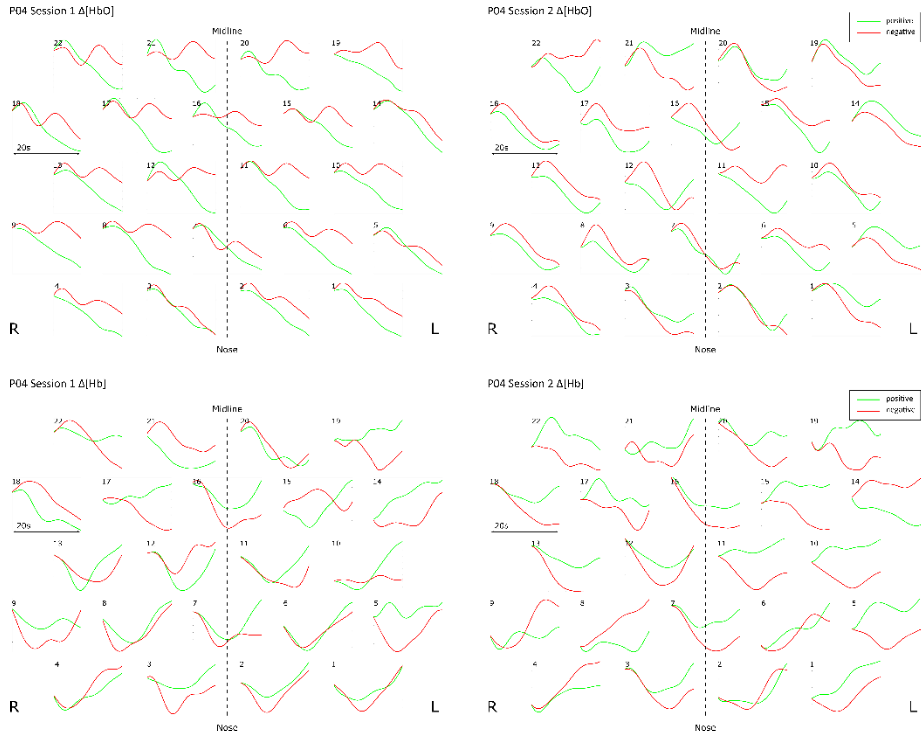


Figure 26: Trial-averaged hemodynamic response functions for Participant 4, Sessions 1 & 2. *The upper left shows Δ[HbO] for Session 1, the upper right shows Δ[HbO] for Session 2, the bottom left shows Δ[Hb] for Session 1, and the bottom right shows Δ[Hb] for Session 2. The response function for each measurement channel is shown mapped according to its position over the forehead. Left and right sides of the head as well as the nose are provided as landmarks. Green indicates the average of all 30 positive trials, and red indicates the average of all 30 negative trials from the respective session. Each response function is 20s in length.*

Figure 25: End-of-session subjective experience questionnaire responses. *Responses are shown for each participant's best (left) and worst (right) session days.*

6.3.7 Temperament Questionnaire

The parent-reported EATQ assessed the participants' temperament in ten different measures: activation control, affiliation, aggression, attention, depressive mood, fear, frustration, inhibitory control, shyness and surgency. Each question on the EATQ was answered with a five-point Likert scale (1 = "almost never true", 5 = "almost always true") and was categorized under one of the ten measures. Mean scores for each measure were calculated by averaging the responses for all the questions pertaining to that measure, resulting in a score from 1-5 for each of the ten measures. These scores were compared to online BCI performance (classification accuracy, sensitivity and specificity). Pearson's correlation coefficient comparing accuracy and score was calculated for each measure (118) and can be found in Table 2A. The only significant correlation found was between aggression and sensitivity (r = +0.63, p = 0.05) and specificity (r = -0.66, p = 0.04), suggesting that participants who were more aggressive in temperament were better at regulating positive emotional valence and worse at regulating negative emotional valence. The ten measures from the EATQ can also be combined into three super-measures; effortful control, an average of scores for attention, activation control and inhibitory control; negative emotionality, an average of scores for frustration, aggression and depressive mood; and extraversion, an average of scores for surgency and the reversed scores of shyness and fear. These scores were determined for each participant and compared with online BCI performance. Pearson's correlation coefficients for these comparisons can be found in Table R8B. The only significant correlation was found between emotionality and sensitivity (r = +0.64, p = 0.05), suggesting that participants with high levels of negative emotionality were better at regulating positive emotional states.

Table 9: Temperament measures - Pearson's correlation coefficient (r) and p-value (p) for: *A) Classification accuracy, sensitivity and specificity with respect to each of the ten measures of temperament; and B) Classification accuracy, sensitivity and specificity with respect to each of the three super-measures of temperament.*

A)

	Activation Control		Affiliation		Aggression		Attention		Depressive Mood	
	r	*p*	*r*	*p*	*r*	*p*	*r*	*p*	*r*	*p*
Accuracy	-0.05	0.89	0.06	0.86	0.23	0.52	-0.19	0.59	0.36	0.31
Sensitivity	-0.02	0.95	0.19	0.60	0.63	0.05*	-0.11	0.75	0.20	0.59
Specificity	0.23	0.53	0.28	0.43	-0.66	0.04*	0.07	0.84	0.22	0.55
	Fear		**Frustration**		**Inhibitory Control**		**Shyness**		**Surgency**	
	r	*p*	*r*	*p*	*r*	*p*	*r*	*p*	*r*	*p*
Accuracy	-0.31	0.39	0.26	0.47	-0.23	0.53	0.43	0.22	-0.07	0.85
Sensitivity	-0.27	0.46	0.33	0.35	-0.46	0.18	0.41	0.24	0.45	0.19
Specificity	0.22	0.54	0.00	1.00	0.35	0.32	-0.34	0.33	-0.28	0.43

B)

	Effortful Control		Emotionality/ Negative Affect		Extraversion/ Surgency	
	r	*p*	*r*	*p*	*r*	*p*
Accuracy	-0.11	0.77	0.41	0.24	-0.03	0.93
Sensitivity	-0.13	0.73	0.64	0.05*	0.09	0.80
Specificity	0.15	0.67	-0.35	0.32	0.07	0.84

6.3.8 Offline Classification Accuracies, All Sessions

In addition to the online classification accuracies, 10 iterations of 10-fold cross-validation were used to evaluate the BCI performance offline for sessions 2-4. The resulting average classification accuracies, along with those from session 1, can be found in table 3. Comparing the offline results of all four sessions, 38 of the 40 sessions exceeded the chance level of 65% (excluding P1 session 2 and P3 session 2). 35 of the 40 sessions exceeded the 99% confidence level, and 15 of the 30 sessions exceeded the 99.9% confidence interval.

Table 10: Offline average cross-validated classification accuracies for all four sessions individually.

Session	Participant										
	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	Average
1	74.3**	90.2***	73.3**	74.2**	92.7***	80.5***	71.7**	82.8***	71.0**	73.3**	78.4**
2	75.7**	80.5***	78.2**	71.7**	78.8**	73.2**	80.0***	74.3**	77.7**	83.8***	77.4**
3	62.3	80.0***	60.0	72.5**	91.3***	83.3***	78.5**	79.3**	81.2***	78.7**	76.7**
4	69.2*	85.5***	68.5*	78.5**	78.5**	79.2**	82.8***	88.2***	72.8**	65.3*	76.9**
Average	**70.4****	**84.0*****	**70.0****	**74.2****	**85.3*****	**79.0****	**78.3****	**81.2*****	**75.7****	**75.3****	**77.3****

6.3.8.1 Multiple Session Cross-Validation Results

To investigate how more training data impacted the BCI's ability to predict emotional valence, BCI performance was evaluated through 10-fold cross-validation using every possible grouping of two, three and all four sessions combined into one training data set. Because the chance-level threshold decreases as the classifier is trained on more samples (e.g. 65% for n=30 samples of each class in one session, and 60% for n=60 samples of each class in two sessions), the classification accuracies from one session, two sessions, three sessions and four sessions of combined data were normalized with respect to their chance-level threshold. Figure 12 shows these normalized accuracies compared to the number of sessions of training data for Participants 4&8. All participants exhibited similar trends; either a decrease

Figure 27: Average cross-validated classification accuracies, normalized against the chance-level threshold for the corresponding number of sessions of training data*, plotted against the number of sessions of training data for P4 and P8. All participants exhibited similar trends; a decrease in classification accuracy with more sessions of training data as with P8, or no change in accuracy as with P4.*

or no change in classification accuracy was seen as the number of sessions of training data increased. This means that classifier performance did not improve, despite being trained on more data.

6.4 Discussion

6.4.1 Feasibility of an NIRS-based Affective BCI for Children

Overall, the results of this study provide strong evidence that it is possible to discriminate emotional valence from hemodynamic activity in children during a bimodal emotion-induction task. The offline cross-validated classification accuracies for each of the four sessions, averaged across all participants (78.4%, 77.4%, 76.7%, and 76.9% for S1-S4 respectively), exceeded the theoretical levels of chance (117). Out of all 40 sessions, there were only two sessions that did not exceed the chance level. The results of the current study fit well with that of the existing affective NIRS-BCI literature. They are comparable to that of Hosseini et al. (range 58.0-83.0%, mean 70.6%) (51), and Moghimi et al. (range 62.0-86.9%, mean 71.9%) (52), who both sought to discriminate positive and negative valence using unimodal stimuli (affective images and music respectively). Tai et al. (50) achieved slightly higher accuracies, (range 75.0-96.7%, mean 84.6%) discriminating between the more distinct emotional vs neutral classes, and Heger et al. (33) achieved lower accuracies (range 44.0%-74.4%) in their multiclass task of discriminating both valence and arousal. Yanagisawa et al. (53) saw considerably more variability in performance, with some achieving as high as 96.7% accuracies but approximately one-third not achieving significant results.

The ability to differentiate positive and negative emotional states from hemodynamic activity in a brain-computer interface suggests underlying neurophysiological differences in the processing of emotional valence. In the literature, these differences have been attributed to two independent neural systems – the approach system, which primes an individual for approach, attachment or appetitive behaviours, and the withdrawal system, which primes an individual for avoidance, flight or defense (27). Stimuli that produce negative emotions such as fear or anxiety would activate brain regions involved with the withdrawal system, facilitating avoidance of the aversive stimulus. Stimuli that produce positive emotions such as happiness or excitement would recruit the approach system, facilitating advancement towards the rewarding stimulus (120). Differential activation with respect to emotional valence has been found in the amygdala (121), the orbito-frontal cortex (122), and the dorsolateral prefrontal cortex (120), as well as in the theta and delta frequency bands of the EEG (123). While distinct activation patterns have been identified, there is also evidence of overlap in the physiological responses of positive and negative valence (124). It should be noted that positive and negative valence are not directly analogous to approach and withdrawal; e.g. fear, a negatively valent emotion, would stimulate the withdrawal system, while anger, also a negative emotion, may instead stimulate the approach system, thus contributing to the overlap seen in positive and negative activation patterns.

6.4.2 Feasibility of Real-Time Emotion Prediction

The study demonstrated moderate success with real-time emotional prediction. Every participant was able to achieve above-chance accuracies ($p < 0.05$) with the online BCI at least once out of the three sessions, and eight of the ten participants achieved above-chance accuracies for two of the three sessions. In the existing literature, it appears that Aranyi et al. were the first to attempt to identify affective state in real time using near-infrared spectroscopy. Participants were shown a virtual character and were instructed to interact with it using either positive emotions (55) or anger (56). The real-time NIRS signal was monitored for hemispheric asymmetry, and an emotional response was 'predicted' if there was a statistically significant increase in average asymmetry when the participants were 'interacting' with the character compared to the preceding baseline trial. For positive emotions, they found that 11 out of 17 participants achieved a statistically significant increase for an average of 50% of their trials. For anger, 8 of 11 participants achieved this threshold for approximately 67% of their trials. These results are comparable to that of the current study, however, none of these reported results approach the classification accuracies seen in offline NIRS affective BCI studies, and fall short of the 70% classification accuracy typically associated with 'effective' BCI use (127).

6.4.3 Interparticipant Variability

BCI performance varied considerably across participants, with some achieving online accuracies as high as 75%, and others struggling to reach the $p<0.05$ chance threshold of 58-60%. However, interparticipant variability in BCI performance is not uncommon; within BCI literature, there is consensus that there is no "universal BCI" system that works for every individual (126). Individual variations in brain structure, spontaneous physiological artefact production, and even variations in proficiency with different BCI training tasks or paradigms can all contribute to differences in BCI performance (126).

A significant positive correlation was found in the current study between BCI performance and age. The three best performers in this study were all over 13 years of age, and comparably, the two lowest performers were 10 years of age or younger. The age of the participants in this study ranged from 8 to 14 years. A considerable amount of emotional, neurophysiological and cognitive development occurs across this age range that could all contribute to BCI ability. From the beginning to the end of grade school, a child's understanding of emotions becomes richer and more complex, and they learn to use a wider range of emotional regulation strategies to adapt their behaviour according to their goals (72). It is possible that the older participants were better able to regulate their emotions in response to the visual neurofeedback than the younger participants, thus producing stronger and more reliable changes in hemodynamic activity for the BCI to classify. Also, the prefrontal cortex begins to develop early in childhood but continues to mature throughout adolescence and into young adulthood (76). It is possible that the more developed prefrontal cortices of the older participants evoked more reliable and distinct patterns of hemodynamic activity in response to affective stimuli. Executive attention networks and the ability to maintain attention over a sustained period of time also develop with age (131,132), meaning that the older participants were likely able to maintain their focus better over the course of the session than the younger participants, resulting in better performance (34).

We also looked to participants' temperament to investigate differences in BCI performance. Differences in temperament have been linked to differences in recruitment of the approach and withdrawal systems, which can impact the type and intensity of emotions evoked in response to a given situation or stimuli (125). It has been found that individuals with higher withdrawal or avoidant temperaments respond more strongly to negatively valent words in an emotional Stroop task, while individuals with higher approach temperaments responded more strongly to positively valent words (135), and that individuals with higher approach temperaments experience greater physiological changes in response to affective images than those with avoidant temperaments (136). In the context of this study, it was expected that differences in temperament in measures such as fear or surgency may reflect differences in emotional responses to negative or positive stimuli. However, the only significant results of the correlation analysis suggest that participants with more aggressive temperaments or higher negative emotionality were better at regulating positive emotions and were poorer at regulating negative emotions. It could be possible that participants who, by nature of their temperament tend to react with greater aggression and negative emotionality, have poorer control over these negative emotions and had greater difficulty regulating them for the online BCI sessions. However, these correlations were weak ($p=0.05$) and should only be considered tentatively. The temperamental measure of effortful control is linked to the ability to effectively regulate emotional responses (137). Children with low levels of effortful control have been shown to have reduced activation in the anterior cingulate cortex (ACC), a crucial component of the brain's emotional networks, when viewing affective images (138). In the context of the current study, it was expected that participants with higher effortful control scores would perform better with the BCI, but unexpectedly, no significant correlation was found in the current study.

The lack of significant correlations between temperament measures and BCI performance is likely due to the limited sample size. Temperament in this study was also based on caregiver assessment, but there can be a lack of convergence between a child's 'true' temperament and their caregiver's perception of their temperament (103), as they can be influenced by their own expectations for their child's behaviour (134). In future studies, using self-assessments of temperament, as well as a larger study population, would help to elucidate how temperament affects evoked emotional responses and the ability to self-regulate these responses in the context of an affective BCI.

6.4.4 Intersession Variability

Considerable variability was seen within each participant's online BCI performance on different session days. For example, Participant 2 achieved classification accuracies of 68% in sessions 2 and 4, but only 55% in session 3. This variability can also be seen in the signal morphology of the trial-averaged hemodynamic response functions across different session days. Furthermore, despite the fact that training a classifier on a larger data set should result in more accurate prediction models (15,111,139), offline classification accuracy decreased when training data from multiple sessions was combined, contributing to the evidence the hemodynamic response was inconsistent across different session days. Hemodynamic response variability is not uncommon in BCI literature; Holper et al. (140) found that a greater amount of intersession variability negatively impacted BCI performance in a motor imagery task, Power et al. (107) found that characteristics of the hemodynamic response varied across days in a mental arithmetic task, and Moghimi et al. (141) found inconsistencies during repeated exposures to the same musical stimulus. Intersession variability has been attributed to participant-related factors such as changes in fatigue, attention, mood, motivation, and underlying spontaneous/baseline neural activity, to environmental factors such as auditory distractions, and to instrumentation-related factors such as deviations in optode placement and calibration (34,107,108,142–144).

To examine changes in fatigue, mood and attention on different session days, participant responses to an end-of-session mental state questionnaire were compared to BCI performance. At the group level, the questionnaire results suggest that participants performed more poorly on days where they felt the most fatigue during the session. Participants also performed more poorly on days where they experienced the greatest downward shift in mood throughout the session, and on days where they reported the greatest amount of frustration with the task. It is likely that fatigue, a poor mood, or sense of frustration decreased motivation and attention, leading to poorer task performance and lower classification accuracies, which in turn further worsened their mood and led to even more fatigue and frustration as they struggled to control the BCI. It has been shown that while some level of frustration and confusion can lead to effective and beneficial problem solving, too much frustration can lead to the complete loss of motivation (145).

It was expected that intensity ratings following each block of emotion induction trials would correlate with BCI performance. However, no significant correlation was found between intensity rating and classification accuracy for any of the participants. This could be due to the fact that the same stimulus set was used for each participant for each session, meaning that by the end of the study, the participants would have seen each image four times. Repeated exposure to an affective stimulus can eventually lead to habituation (146,147), resulting in the participants perceiving their emotions as being less intense. Also, a single rating of intensity after every 12 trials with only a 5-point scale was collected. More detailed or more frequent ratings may be required to better assess trends between emotional response intensity and BCI performance. However, there is a trade-off between collecting more detailed information and minimizing session complexity and cognitive load (148).

6.4.5 Differences between Online & Offline Performance

Across all three online sessions, the offline cross-validated classification accuracies were consistently higher than the online, real-time classification accuracies. This is commonly seen in BCI literature (13), and indicates that the trained classifiers used for real-time prediction had difficulty generalizing to new, unlabeled data (111). There could be several reasons the classifiers exhibited poor generalizability. Firstly, the classifier used for a given session were trained on previous session data, but the study results have indicated that there is considerable variability in the hemodynamic response across different session days. This is supported by the fact that classifier performance improved as more same-day session data was included. The last block of each session was on average, greater than the first block of each session, although this difference was not significant. However, each session was limited in length to prevent participants from becoming too fatigued. It is likely that a significant improvement in classifier performance would be seen as more-same day data was collected, as was seen with Rezazadeh et al.'s online covert speech BCI (85). Secondly, the classifier training paradigm could have been prone to overfitting the data. Despite efforts to reduce dimensionality through feature selection and increase generalizability through regularization, it is possible that the dimension of the feature set (p=10 features) used in this study was still too large (109,111). Furthermore, the feature selection algorithm used in this study, sequential forward floating search, can be prone to overfitting (149). Using a filter feature selection method such as the mutual information (MI) method or an embedded feature selection method such as the least absolute shrinkage and selection operator (LASSO) method could have allowed for feature set dimensionality reduction while being less likely to overfit the training data (150). Finally, due to the nature of the experiment, there are likely samples within the collected data set that are *less than optimal*, possibly caused by shifts in attention, movement, distractions in the data collection environment, or fatigue. Future BCI work should focus on developing methods for optimizing the training data set to produce more robust and generalizable classifiers.

6.4.6 Limitations

Paediatric Population

Working with a paediatric population introduces several experimental limitations. Children have a significantly lower attention span and experience fatigue more quickly than adults (131), meaning the experimental sessions had to be kept as short as possible, limiting the amount of data that could be collected in a single session. Even within the shortened session, differences in attention could be seen between the younger and older participants. Younger participants had greater difficulty sitting still and required more frequent breaks than the older participants. This restlessness risks introducing artefacts in the acquired hemodynamic signal and producing unwanted shifts in the signal due to optode-scalp decoupling (108). Overall, a lack of attention or loss of focus can greatly detract from task performance (154) and is likely a significant limitation of working with a paediatric population for BCI research.

Instrumentation

There are also limitations related to the instrumentation used in this study. NIRS is relatively robust to motion artefacts compared to other neuroimaging modalities, provided that the optodes remain in good contact with the scalp (41,108). Good optode-scalp contact can be maintained by ensuring a tight fit with the headpiece holding the optodes. However, there is a trade-off between a tight fit and headpiece versatility and comfort. When dealing with a paediatric population with variable head sizes and a lower tolerance for discomfort, this design issue was of considerable importance. Hair colour, hair density and skin pigmentation can also impact the signal-to-noise ratio of the hemodynamic signal (41). Another significant limitation of NIRS instrumentation is that this neuroimaging method only allows for superficial cortical detection, with a maximum depth of about 2-3cm into the skull (80,81). Many of the brain regions implicated in emotion processing are located deeper inside the skull, such as the amygdala, hypothalamus, the anterior cingulate cortex, and even the ventromedial prefrontal cortex. A neuroimaging method providing greater depth of resolution, such as fMRI, may provide greater insight into brain activation during emotional responses.

Cerebral Hemodynamic Response

The long latency of the cerebral hemodynamic responses is another limitation of using NIRS as an imaging modality for brain-computer interfaces. The slow hemodynamic response, the result of changes in cerebral blood flow and metabolism, has an onset of about 5-8 seconds (83). A time window of 20s was used in the current study to fully capture this response. However, such a long trial period considerably increased the length of the sessions and limited the amount of data that could be collected before participants grew fatigued. To shorten NIRS data collection sessions, it would be worth investigating the *initial dip* phenomenon, which refers to the initial drop in concentration of oxygenated hemoglobin that is observed as localized activated neurons rapidly metabolize oxygen before an increase in blood flow delivers more oxygenated hemoglobin to the region (155,156). The initial dip has been observed in a variety of brain regions including the visual, motor, auditory and prefrontal cortices, although to the author's knowledge no studies have investigated the initial dip in induced emotional states (156).

Statistical Power

This study was also limited by a small sample size, reducing the statistical power to draw conclusions on several aspects of BCI performance. Testing the developed BCI on a larger population would strengthen the validity of the results and allow for better investigation of trends in age, gender, temperament, fatigue, mood, perceived effort and frustration in relation to BCI performance.

6.5 Conclusion

6.5.1 Summary of Study Contributions

This study demonstrated that it is possible to differentiate emotional valence from changes in hemodynamic activity during an emotion-induction task in the prefrontal cortices of children with an average accuracy of approximately 77%. This discrimination was possible with all ten participants and across four different session days, and the results from this study were comparable to existing affective NIRS-BCI studies conducted in adults. The ability to discriminate positive and negative emotional valence has been attributed to underlying differences in activation patterns that indicate the recruitment of so-called approach or withdrawal systems in the brain in response to rewarding or aversive stimuli.

This study also demonstrated the feasibility of real-time emotion prediction using the developed BCI. While above-chance level online classification accuracies were achieved for most of the participants in at least two of their three online sessions, none of the participants surpassed the 70% accuracy threshold considered 'effective' for BCI use, and the differences between offline and online classification accuracies indicate that there is considerable room for improvement. Intersession variability in the hemodynamic response likely played a part in limiting the success of online prediction, as online classifiers were trained on data from the previous session days. Changes in mental state (such as fatigue, decrease in mood, and loss of attention) within and across different session days could have contributed to this variability. In the future, steps should be taken to improve the generalizability of the affective state classifiers, so they can better accommodate for intersession variability.

Considerable interparticipant was also found in this study. Age and temperament were identified as factors contributing to interparticipant variability. BCI performance was found to increase with age, likely due to the increased attention span, maturation of the prefrontal cortex and a better ability to regulate emotions that develops with age. Limited conclusions could be drawn about BCI performance with respect to temperament due to the small sample size of this study, and differences in gender could not be explored, as only 3 males participated in the study. More work is needed to elucidate trends in gender and temperament with respect to affective BCI performance.

Overall, to the author's knowledge the research described in this thesis was the first of its kind to investigate an affective near-infrared spectroscopy brain-computer interface for a paediatric population. It is also one of the first studies to conduct online prediction of affective state from hemodynamic activity. This study also demonstrated the impacts of interparticipant and intersession variability of the hemodynamic response on affective BCI performance.

6.5.2 Future Work

In moving towards an affective NIRS-based BCI to be used as a communication device, some future steps are recommended. The most immediate future work should focus on improving the generalizability of the online classification paradigm for real-time emotional valence prediction. Using the proposed BCI as a communication device would require better real-time prediction accuracy than what was demonstrated in this study. Improving the generalizability of the classification paradigm for online classification could include developing methods to identify optimal subsets of training data. The next step would be to determine if real-time emotional valence prediction can be replicated using self-induced emotions rather than stimulus-evoked emotion, as in a real communicative interaction, emotions would be self-generated. Finally, the affective BCI would need to be tested with the target population, children with physical disabilities. Children with physical disabilities may have associated delays or impairments in cognitive or emotional processing that could affect their ability to operate an affective brain-computer interface.

6.5.3 Significance

A brain-computer interface that can differentiate between positive and negatively valenced emotional responses could be used to detect these emotional states in children with severe physical disabilities who have had limited success with existing assistive technologies for communication. The binary detection of emotional valence could be used to express feelings, preference, or even affirmative/negative responses to questions without the needs for words or other developed language abilities. This would provide children with severe physical disabilities with an access pathway to communication that circumvents the need for functional motor control and literacy skills. With access to communication, these children can engage within their communities, learn how to advocate for themselves, gain independence, and overall improve their quality of life.

References

1. Crichton S. Understanding, identifying and supporting speech, language and communication needs in children. Community Pract. 2013;86(12):44–7.
2. Romski M, Sevcik RA, Barton-Hulsey A, Whitmore AS. Early Intervention and AAC: What a Difference 30 Years Makes. AAC Augment Altern Commun. 2015;31(3):181–202.
3. Light J, Drager K. AAC technologies for young children with complex communication needs: State of the science and future research directions. AAC Augment Altern Commun. 2007;23(3):204–16.
4. Sevcik RA, Romski MA, Adamson LB. Research directions in augmentative and alternative communication for preschool children. Disabil Rehabil. 2004;26(21–22):1323–9.
5. Millar DC, Light JC, Schlosser RW. The Impact of Augmentative and Alternative Communication Intervention on the Speech Production of Individuals With Developmental Disabilities: A Research Review. J Speech, Lang Hear Res. 2006;49:248–64.
6. Light JC, Drager KDR. Improving the Design of Augmentative and Alternative Technologies for Young Children. Assist Technol. 2002;14:17–32.
7. Desch LW, Gaebler-Spira D. Prescribing Assistive-Technology Systems: Focus on Children With Impaired Communication. Pediatrics [Internet]. 2008;121(6):1271–80. Available from: http://pediatrics.aappublications.org/cgi/doi/10.1542/peds.2008-0695
8. Tai K, Blain S, Chau T. A review of emerging access technologies for individuals with severe motor impairments. Assist Technol [Internet]. 2008;20(4):204–19. Available from: http://www.scopus.com/inward/record.url?eid=2-s2.0-58149512630&partnerID=40&md5=8bd0f64092443b32c00081a4b5147ef2
9. Fager S, Bardach L, Russell S, Higginbotham J. Access to augmentative and alternative communication: New technologies and clinical decision-making. J Pediatr Rehabil Med. 2012;5(1):53–61.
10. Myrden A, Schudlo L, Weyand S, Zeyl T, Chau T. Trends in Communicative Access Solutions for Children With Cerebral Palsy. J Child Neurol [Internet]. 2014;29(8):1108–18. Available from: http://journals.sagepub.com/doi/10.1177/0883073814534320
11. Bruno M-A, Schnakers C, Damas F, Pellas F, Lutte I, Bernheim J, et al. Locked-In Syndrome in Children: Report of Five Cases and Review of the Literature. Pediatr Neurol [Internet]. 2009;41(4):237–46. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0887899409002665
12. Tabar YR, Halici U. Brain Computer Interfaces for Silent Speech. Eur Rev. 2016;25(2):208–30.
13. McFarland DJ, Wolpaw JR. Brain-Computer Interfaces for Communication and Control. Commun ACM. 2011;54(5):60–6.
14. Moghimi S, Kushki A, Marie Guerguerian A, Chau T. A Review of EEG-Based Brain-Computer Interfaces as Access Pathways for Individuals with Severe Disabilities. Assist Technol [Internet]. 2013;25(2):99–110. Available from: http://www.tandfonline.com/doi/abs/10.1080/10400435.2012.723298
15. Nicolas-Alonso LF, Gomez-Gil J. Brain Computer Interfaces, a Review. Sensors. 2012;12:1211–79.
16. Lazarou I, Nikolopoulos S, Petrantonakis PC, Kompatsiaris I, Tsolaki M. EEG-Based Brain–Computer Interfaces for Communication and Rehabilitation of People with Motor Impairment: A Novel Approach of the 21st Century. Front Hum Neurosci [Internet]. 2018;12(January):1–18. Available from: http://journal.frontiersin.org/article/10.3389/fnhum.2018.00014/full
17. Cecotti H. Spelling with non-invasive Brain-Computer Interfaces - Current and future trends. J Physiol Paris [Internet]. 2011;105:106–14. Available from: http://dx.doi.org/10.1016/j.jphysparis.2011.08.003
18. Happ MB. Interpretation of nonvocal behavior and the meaning of voicelessness in critical care. Soc Sci Med. 2000;50(9):1247–55.
19. Mühl C, Allison B, Nijholt A, Chanel G. A survey of affective brain computer interfaces: principles, state-of-the-art, and challenges. Brain-Computer Interfaces [Internet]. 2014;1(2):66–84. Available from:

http://www.tandfonline.com/doi/abs/10.1080/2326263X.2014.912881

20. Feldman Barrett L. Are Emotions Natural Kinds ? Perspect Psychol Sci [Internet]. 2006 [cited 2018 Apr 12];1(1):28–58. Available from: http://journals.sagepub.com.myaccess.library.utoronto.ca/doi/pdf/10.1111/j.1745-6916.2006.00003.x
21. Hamann S. Mapping discrete and dimensional emotions onto the brain: Controversies and consensus. Trends Cogn Sci [Internet]. 2012;16(9):458–66. Available from: http://dx.doi.org/10.1016/j.tics.2012.07.006
22. Kuppens P, Tuerlinckx F, Russell JA, Barrett LF. The relation between valence and arousal in subjective experience. Psychol Bull. 2013;139(4):917–40.
23. Feldman Barrett L. Valence is a basic building block of emotional life. J Res Pers. 2006;40(1):35–55.
24. Scherer KR. What are emotions? and how can they be measured? Soc Sci Inf. 2005;44(4):695–729.
25. Kandel ER, Schwartz JH, Jessell TM, Siegelbaum SA, Hudspeth AJ. Principles of Neural Science. 5th ed. McGraw-Hill; 2013.
26. LeDoux JE. Emotion Circuits in the Brain. Annu Rev Neurosci [Internet]. 2000;23:155–84. Available from: http://psychiatryonline.org/doi/abs/10.1176/foc.7.2.foc274
27. Davidson RJ. Anxiety and affective style: Role of prefrontal cortex and amygdala. Biol Psychiatry. 2002;51(1):68–80.
28. Feldman Barrett L, Mesquita B, Ochsner KN, Gross JJ. The Experience of Emotions. Annu Rev Psychol. 2007;58(1):373–403.
29. Dixon ML, Thiruchselvam R, Todd R, Christoff K. Psychological Bulletin Review Emotion and the Prefrontal Cortex : An Integrative Review. Psychol Bull. 2017;143(10):1033–81.
30. Hiser J, Koenigs M. The Multifaceted Role of the Ventromedial Prefrontal Cortex in Emotion, Decision Making, Social Cognition, and Psychopathology. Biol Psychiatry [Internet]. 2018;83(8):638–47. Available from: https://doi.org/10.1016/j.biopsych.2017.10.030
31. Phillips ML, Ladouceur CD, Drevets WC. A neural model of voluntary and automatic emotion regulation: Implications for understanding the pathophysiology and neurodevelopment of bipolar disorder. Mol Psychiatry. 2008;13(9):833–57.
32. Ochsner KN, Ray RR, Hughes B, Mcrae K, Cooper JC, Weber J, et al. Bottom up and top down processes in emotion generation: common and distinct neural mechanisms. Psychol Sci. 2009;20(11):1322–31.
33. Heger D, Mutter R, Herff C, Putze F, Schultz T. Continuous recognition of affective states by functional near infrared spectroscopy signals. Brain-Computer Interfaces [Internet]. 2014;1(2):113–25. Available from: http://dx.doi.org/10.1080/2326263X.2014.912884
34. Myrden A, Chau T. Effects of user mental state on EEG-BCI performance. Front Hum Neurosci [Internet]. 2015;9:1–13. Available from: http://journal.frontiersin.org/Article/10.3389/fnhum.2015.00308/abstract
35. Mehmood RM, Du R, Lee HYOJ. Optimal Feature Selection and Deep Learning Ensembles Method for Emotion Recognition From Human Brain EEG Sensors. IEEE Access. 2017;5:14797–806.
36. Yano K, Suyama T. A Novel Fixed Low-Rank Constrained EEG Spatial Filter Estimation with Application to Movie-Induced Emotion Recognition. Comput Intell Neurosci. 2016;2016:6734720:1-12.
37. Atkinson J, Campos D. Improving BCI-based emotion recognition by combining EEG feature selection and kernel classifiers. Expert Syst Appl [Internet]. 2016;47:35–41. Available from: http://dx.doi.org/10.1016/j.eswa.2015.10.049
38. Iacoviello D, Petracca A, Spezialetti M, Placidi G. A real-time classification algorithm for EEG-based BCI driven by self-induced emotions. Comput Methods Programs Biomed [Internet]. 2015;122(3):293–303. Available from: http://dx.doi.org/10.1016/j.cmpb.2015.08.011
39. McFarland DJ, Parvaz MA, Sarnacki WA, Goldstein RZ, Wolpaw JR. Prediction of subjective ratings of emotional pictures by EEG features. J Neural Eng [Internet]. 2017;14(1):016009. Available from:

http://iopscience.iop.org/article/10.1088/1741-2552/14/1/016009

40. Gratton G, Fabiani M, Elbert T, Rockstroh B. Seeing right through you: Applications of optical imaging to the study of the human brain. Psychophysiology [Internet]. 2003;40(4):487–91. Available from: http://onlinelibrary.wiley.com/doi/10.1111/1469-8986.00050/full

41. Scholkmann F, Kleiser S, Metz AJ, Zimmermann R, Mata Pavia J, Wolf U, et al. A review on continuous wave functional near-infrared spectroscopy and imaging instrumentation and methodology. Neuroimage [Internet]. 2014;85:6–27. Available from: http://dx.doi.org/10.1016/j.neuroimage.2013.05.004

42. Coyle SM, Ward TE, Markham CM. Brain-computer interface using a simplified functional near-infrared spectroscopy system. J Neural Eng. 2007;4:219–26.

43. Hoshi Y. Functional near-infrared optical imaging: utility and limitations in human brain mapping. Psychophysiology [Internet]. 2003;40(4):511–20. Available from: http://www.ncbi.nlm.nih.gov/pubmed/14570159

44. Plichta MM, Herrmann MJ, Baehne CG, Ehlis AC, Richter MM, Pauli P, et al. Event-related functional near-infrared spectroscopy (fNIRS): Are the measurements reliable? Neuroimage. 2006;31(1):116–24.

45. Steinbrink J, Villringer A, Kempf F, Haux D, Boden S, Obrig H. Illuminating the BOLD signal: combined fMRI-fNIRS studies. Magn Reson Imaging. 2006;24(4):495–505.

46. Bendall RCA, Eachus P, Thompson C. A Brief Review of Research Using Near-Infrared Spectroscopy to Measure Activation of the Prefrontal Cortex during Emotional Processing: The Importance of Experimental Design. Front Hum Neurosci. 2016;10(October):1–7.

47. Balconi M, Molteni E. Past and future of near-infrared spectroscopy in studies of emotion and social neuroscience. J Cogn Psychol. 2016;28(2):129–46.

48. Doi H, Nishitani S, Shinohara K. NIRS as a tool for assaying emotional function in the prefrontal cortex. Front Hum Neurosci [Internet]. 2013;7(November):1–6. Available from: http://journal.frontiersin.org/article/10.3389/fnhum.2013.00770/abstract

49. Maria A, Shekhar S, Nissilä I, Kotilahti K, Huotilainen M, Karlsson L, et al. Emotional Processing in the First 2 Years of Life: A Review of Near-Infrared Spectroscopy Studies. J Neuroimaging. 2018;28(5):441–54.

50. Tai K, Chau T. Single-trial classification of NIRS signals during emotional induction tasks: Towards a corporeal machine interface. J Neuroeng Rehabil. 2009;6(1):1–14.

51. Hosseini SMH, Mano Y, Rostami M, Takahashi M, Sugiura M, Kawashima R. Decoding what one likes or dislikes from single-trial fNIRS measurements. Neuroreport. 2011;22(6):269–73.

52. Moghimi S, Kushki A, Power S, Guerguerian AM, Chau T. Automatic detection of a prefrontal cortical response to emotionally rated music using multi-channel near-infrared spectroscopy. J Neural Eng. 2012;9(2).

53. Yanagisawa K, Tsunashima H. Evaluation of pleasant and unpleasant emotions evoked by visual stimuli using NIRS. 15th Int Conf Control Autom Syst Proc. 2015;383–8.

54. Hu X, Zhuang C, Wang F, Liu Y-J, Im C-H, Zhang D. fNIRS Evidence for Recognizably Different Positive Emotions. Front Hum Neurosci. 2019;13(April):1–11.

55. Aranyi G, Pecune F, Charles F, Pelachaud C, Cavazza M. Affective interaction with a virtual character through an fNIRS brain-computer interface. Front Comput Neurosci. 2016;10:1–14.

56. Aranyi G, Charles F, Cavazza M. Anger-based BCI using fNIRS neurofeedback. UIST 2015 - Proc 28th Annu ACM Symp User Interface Softw Technol. 2015;511–21.

57. Hamann S, Canli T. Individual differences in emotion processing. Curr Opin Neurobiol. 2004;14(2):233–8.

58. Lichev V, Sacher J, Ihme K, Rosenberg N, Quirin M, Lepsien J, et al. Automatic emotion processing as a function of trait emotional awareness: An fMRI study. Soc Cogn Affect Neurosci. 2015;10(5):680–9.

59. Gross JJ. Emotion regulation: affective, cognitive, and social consequences. Psychophysiology [Internet]. 2002;39(3):281–91. Available from:

http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=12212647

60. Peña-Sarrionandia A, Mikolajczak M, Gross JJ. Integrating emotion regulation and emotional intelligence traditions: A meta-analysis. Front Psychol. 2015;6:1–27.

61. Peña-Sarrionandia A, Mikolajczak M, Gross JJ. Integrating emotion regulation and emotional intelligence traditions: A meta-analysis. Front Psychol. 2015;6(FEB):1–27.

62. Ochsner KN, Gross JJ. The cognitive control of emotion. Trends Cogn Sci. 2005;9(5):242–9.

63. Sherlin LH, Arns M, Lubar J, Heinrich H, Kerson C, Strehl U, et al. Neurofeedback and Basic Learning Theory: Implications for Research and Practice. J Neurother. 2011;15:292–304.

64. Sitaram R, Ros T, Stoeckel L, Haller S, Scharnowski F, Lewis-Peacock J, et al. Closed-loop brain training: The science of neurofeedback. Nat Rev Neurosci. 2017;18(2):86–100.

65. Ninaus M, Kober SE, Witte M, Koschutnig K, Neuper C, Wood G. Brain volumetry and self-regulation of brain activity relevant for neurofeedback. Biol Psychol [Internet]. 2015;110:126–33. Available from: http://dx.doi.org/10.1016/j.biopsycho.2015.07.009

66. Mayer K, Blume F, Wyckoff SN, Brokmeier LL, Strehl U. Neurofeedback of slow cortical potentials as a treatment for adults with Attention Deficit-/Hyperactivity Disorder. Clin Neurophysiol [Internet]. 2016;127(2):1374–86. Available from: http://dx.doi.org/10.1016/j.clinph.2015.11.013

67. Birbaumer N, Kübler A, Ghanayim N, Hinterberger T, Perelmouter J, Kaiser J, et al. The thought translation device (TTD) for completely paralyzed patients. IEEE Trans Rehabil Eng. 2000;8(2):190–3.

68. Hinterberger T, Schmidt S, Neumann N, Mellinger J, Blankertz B, Curio G, et al. Brain-computer communication and slow cortical potentials. IEEE Trans Biomed Eng. 2004;51(6):1011–8.

69. Grosse-Wentrup M, Schölkopf B. A brain-computer interface based on self-regulation of gamma-oscillations in the superior parietal cortex. J Neural Eng. 2014;11(5).

70. Weyand S, Takehara-Nishiuchi K, Chau T. Weaning Off Mental Tasks to Achieve Voluntary Self-Regulatory Control of a Near-Infrared Spectroscopy Brain-Computer Interface. IEEE Trans Neural Syst Rehabil Eng. 2015;23(4):548–61.

71. Casey BJ, Tottenham N, Liston C, Durston S. Imaging the developing brain: What have we learned about cognitive development? Trends Cogn Sci. 2005;9(3 SPEC. ISS.):104–10.

72. Thompson RA. Emotional Regulation and Emotional Development. Educ Psychol Rev. 1991;3(4):269–307.

73. Denham S, von Salisch M, Olthof T, Kochanoff A, Caverly S. Emotional and Social Development in Childhood. In: Smith PK, Hart CH, editors. Blackwell Handbook of Childhood Social Development. Oxford, UK: Blackwell Publishers; 2002. p. 307–28.

74. Posner MI, Rothbart MK, Sheese BE, Voelker P. Developing Attention: Behavioral and Brain Mechanisms. Adv Neurosci. 2014;2014:405094:1-9.

75. Perlman SB, Pelphrey KA. Regulatory brain development: Balancing emotion and cognitiont. Soc Neurosci. 2010;5(5–6):533–42.

76. Tsujimoto S. Review: The prefrontal cortex: Functional neural development during early childhood. Neuroscientist. 2008;14(4):345–58.

77. Schudlo LC, Chau T. Towards a ternary NIRS-BCI: Single-trial classification of verbal fluency task, Stroop task and unconstrained rest. J Neural Eng [Internet]. 2015;12(6):66008: 1-12. Available from: http://dx.doi.org/10.1088/1741-2560/12/6/066008

78. Weyand S, Schudlo L, Takehara-Nishiuchi K, Chau T. Usability and performance-informed selection of personalized mental tasks for an online near-infrared spectroscopy brain-computer interface. Neurophotonics [Internet]. 2015;2(2):025001. Available from: http://neurophotonics.spiedigitallibrary.org/article.aspx?doi=10.1117/1.NPh.2.2.025001

79. Naseer N, Hong K-S. fNIRS-based brain-computer interfaces: a review. Front Hum Neurosci [Internet].

2015;9(January):1–15. Available from: http://journal.frontiersin.org/article/10.3389/fnhum.2015.00003/abstract

80. Okada E, Firbank M, Schweiger M, Arridge, Cope M, Delpy D. Theoretical and experimental investigation of near-infrared light propagation in a model of the adult head. Appl Opt [Internet]. 1997;36(1):21–31. Available from: http://discovery.ucl.ac.uk/179663/

81. Haeussinger FB, Heinzel S, Hahn T, Schecklmann M, Ehlis AC, Fallgatter AJ. Simulation of near-infrared light absorption considering individual head and prefrontal cortex anatomy: Implications for optical neuroimaging. PLoS One. 2011;6(10).

82. Okamoto M, Dan H, Sakamoto K, Takeo K, Shimizu K, Kohno S, et al. Three-dimensional probabilistic anatomical cranio-cerebral correlation via the international 10-20 system oriented for transcranial functional brain mapping. Neuroimage. 2004;21(1):99–111.

83. Coyle S, Ward T, Markham C, McDarby G. On the suitability of near-infrared (NIR) systems for next-generation brain – computer interfaces. Physiol Meas. 2004;25:815–22.

84. Power SD, Kushki A, Chau T. Automatic single-trial discrimination of mental arithmetic, mental singing and the no-control state from prefrontal activity: Toward a three-state NIRS-BCI. BMC Res Notes [Internet]. 2012;5(1):141:1-10. Available from: http://www.biomedcentral.com/1756-0500/5/141

85. Rezazadeh Sereshkeh A, Yousefi R, Wong AT, Chau T. Online classification of imagined speech using functional near-infrared spectroscopy signals. J Neural Eng [Internet]. 2019;16(1):016005:1-13. Available from: http://www.ncbi.nlm.nih.gov/pubmed/30260320

86. Schudlo LC, Chau T. Dynamic topographical pattern classification of multichannel prefrontal NIRS signals: II. Online differentiation of mental arithmetic and rest. J Neural Eng. 2014;11(1).

87. Weyand S, Takehara-Nishiuchi K, Chau T. Exploring methodological frameworks for a mental task-based near-infrared spectroscopy brain-computer interface. J Neurosci Methods [Internet]. 2015;254:36–45. Available from: http://dx.doi.org/10.1016/j.jneumeth.2015.07.007

88. Baumgartner T, Lutz K, Schmidt CF, Jäncke L. The emotional power of music: How music enhances the feeling of affective pictures. Brain Res. 2006;1075(1):151–64.

89. Baumgartner T, Esslen M, Jäncke L. From emotion perception to emotion experience: Emotions evoked by pictures and classical music. Int J Psychophysiol. 2006;60(1):34–43.

90. Lang PJ, Bradley MM, Cuthbert BN. International affective picture system (IAPS): Affective ratings of pictures and instruction manual. Gainesville, FL.; 2008.

91. Dan-Glauser ES, Scherer KR. The Geneva affective picture database (GAPED): A new 730-picture database focusing on valence and normative significance. Behav Res Methods. 2011;43(2):468–77.

92. Kurdi B, Lozano S, Banaji MR. Introducing the Open Affective Standardized Image Set (OASIS). Behav Res Methods. 2017;49(2):457–70.

93. Sharp C, van Goozen S, Goodyer I. Children's subjective emotional reactivity to affective pictures: Gender differences and their antisocial correlates in an unselected sample of 7-11-year-olds. J Child Psychol Psychiatry Allied Discip. 2006;47(2):143–50.

94. Hajcak G, Dennis T. Brain Potentials During Affective Picture Processing in Children. Biol Psychol [Internet]. 2009;80(3):333–8. Available from: http://www.sciencedirect.com/science/article/pii/S0301051108002329

95. Mcmanis MH, Bradley MM, Berg WK, Cuthbert BN, Lang PJ. Emotional reactions in children: Verbal, physiological, and behavioral responses to affective pictures. Psychophysiology [Internet]. 2001 [cited 2018 Mar 21];38:222–31. Available from: https://journals-scholarsportal-info.myaccess.library.utoronto.ca/pdf/00485772/v38i0002/222_ericvpabrtap.xml

96. Kreutz G, Ott U, Teichmann D, Osawa P, Vaitl D. Using music to induce emotions: influences of musical preference and absorption. Psychol Music [Internet]. 2008 [cited 2018 Mar 21];36(1):101–26. Available from: https://journals-scholarsportal-info.myaccess.library.utoronto.ca/pdf/03057356/v36i0001/101_umtieiompaa.xml

97. Koelsch S. Investigating emotion with music: neuroscientific approaches. Ann N Y Acad Sci. 2005;1060:412–8.

98. Dalla Bella S, Peretz I, Rousseau L, Gosselin N. A developmental study of the affective value of tempo and mode in music [Internet]. Vol. 80, Cognition. 2001 [cited 2018 Mar 21]. p. B1–10. Available from: https://journals-scholarsportal-info.myaccess.library.utoronto.ca/pdf/00100277/v80i0003/b1_adsotaotamim.xml

99. Nieminen S, Istók E, Brattico E, Tervaniemi M, Huotilainen M. The development of aesthetic responses to music and their underlying neural and psychological mechanisms. CORTEX [Internet]. 2011 [cited 2018 Mar 21];47:1138–46. Available from: https://journals-scholarsportal-info.myaccess.library.utoronto.ca/pdf/00109452/v47i0009/1138_tdoarttunapm.xml

100. Bradley MM, Lang PJ. Measuring emotion: The self-assessment manikin and the semantic differential. J Behav Ther Exp Psychiatry. 1994;25(1):49–59.

101. Hayashi ECS, Posada JEG, Maike VRML, Baranauskas MCC. Exploring new formats of the Self-Assessment Manikin in the design with children. Proc 15th Brazilian Symp Hum Factors Comput Syst - IHC '16 [Internet]. 2016;1–10. Available from: http://dl.acm.org/citation.cfm?doid=3033701.3033728

102. Rothbart MK. Temperament , Development , and Personality. Curr Dir Psychol Sci. 2007;16(4):207–12.

103. Capaldi D, Rothbart M. Development and Validation of an Early Adolescent Temperament Measure. J Early Adolesc. 1992;12(2):153–73.

104. Ellis L, Rothbart M. Revision of the early adolescent temperament questionnaire. Soc Res Child Dev [Internet]. 2001;81:1–5. Available from: http://www.bowdoin.edu/~sputnam/rothbart-temperament-questionnaires/pdf/lesa-ellis-srcd-poster-reprint.pdf

105. Muris P, Meesters C. Reactive and regulative temperament in youths: Psychometric evaluation of the early adolescent temperament questionnaire-revised. J Psychopathol Behav Assess. 2009;31(1):7–19.

106. Coyle S, Ward T, Markham C. Physiological noise in near-infrared spectroscopy: implications for optical brain computer interfacing. 26th Annu Int Conf IEEE Eng Med Biol Soc [Internet]. 2004;4:4540–3. Available from: http://ieeexplore.ieee.org/document/1404260/

107. Power SD, Kushki A, Chau T. Intersession consistency of single-trial classification of the prefrontal response to mental arithmetic and the no-control state by NIRS. PLoS One. 2012;7(7):e37791:1-12.

108. Orihuela-Espina F, Leff DR, James DR, Darzi AW, Yang GZ. Quality control and assurance in functional near infrared spectroscopy (fNIRS) experimentation. Phys Med Biol. 2010;55:3701–24.

109. Kudo M, Sklansky J. Comparison of algorithms that select features for pattern classifiers. Pattern Recognit. 2000;33(1):25–41.

110. Gerven M Van, Farquhar J, Schaefer R, Vlek R, Geuze J, Nijholt A, et al. The brain-computer interface cycle. J Neural Eng. 2009;6(4).

111. Jain AK, Duin RP, Mao J. Statistical Pattern Recognition: A Review. IEEE Trans Pattern Anal Mach Intell. 2000;22(1):4–37.

112. Pudil P, Novovičová J, Kittler J. Floating search methods in feature selection. Pattern Recognit Lett. 1994;15:1119–25.

113. Power SD, Kushki A, Chau T. Towards a system-paced near-infrared spectroscopy brain-computer interface: Differentiating prefrontal activity due to mental arithmetic and mental singing from the no-control state. J Neural Eng. 2011;8:066004:1-14.

114. Guo Y, Hastie T, Tibshirani R. Regularized linear discriminant analysis and its application in microarrays. Biostatistics. 2007;8(1):86–100.

115. Friedman JH. Regularized Discriminant Analysis. J Am Stat Assoc. 1988;1988(July).

116. Naseer N, Hong MJ, Hong KS. Online binary decision decoding using functional near-infrared spectroscopy for the development of brain-computer interface. Exp Brain Res. 2014;232(2):555–64.

117. Combrisson E, Jerbi K. Exceeding chance level by chance: The caveat of theoretical chance levels in brain signal classification and statistical assessment of decoding accuracy. J Neurosci Methods [Internet]. 2015;250:126–36. Available from: http://dx.doi.org/10.1016/j.jneumeth.2015.01.010

118. Rimbert S, Gayraud N, Bougrain L, Clerc M, Fleck S. Can a Subjective Questionnaire Be Used as Brain-Computer Interface Performance Predictor? Front Hum Neurosci. 2019;12(January):529: 1-11.

119. Lang PJ, Bradley MM, Cuthbert BN. Emotion, Attention, and the Startle Reflex. Psychol Rev. 1990;97(3):377–95.

120. Colibazzi T, Posner J, Wang Z, Gorman D, Gerber A, Yu S, et al. Neural Systems Subserving Valence and Arousal During the Experience of Induced Emotions. Emotion. 2010;10(3):377–89.

121. Davidson RJ. Anterior Cerebral Asymmetry and the Nature of Emotion. Brain Cogn. 1992;151:125–51.

122. Lewis PA, Critchley HD, Rotshtein P, Dolan RJ. Neural correlates of processing valence and arousal in affective words. Cereb Cortex. 2007;17(3):742–8.

123. Balconi M, Grippa E, Vanutelli ME. What hemodynamic (fNIRS), electrophysiological (EEG) and autonomic integrated measures can tell us about emotional processing. Brain Cogn [Internet]. 2015;95:67–76. Available from: http://dx.doi.org/10.1016/j.bandc.2015.02.001

124. Lane RD, Reiman EM, Bradley MM, Lang PJ, Ahern GL, Davidson RJ, et al. Neuroanatomical correlates of pleasant and unpleasant emotion. Neuropsychologia. 1997;35(11):1437–44.

125. Bates J, Goodnight J, Fite J. Temperament and Emotion. In: Lewis M, Haviland-Jones J, Feldman Barret L, editors. Handbook of Emotions. 3rd ed. New York, NY, US: The Guilford Press; 2008. p. 485–96.

126. Allison BZ, Neuper C. Brain-Computer Interfaces [Internet]. Tan D, Nijholt A, editors. New York, NY, US; 2010. Available from: http://link.springer.com/10.1007/978-3-642-02091-9

127. Kubler A, Mushahwar V, Hochberg L, Donoghue J. BCI Meeting 2005 - Workshop on clinical issues and applications. IEEE Trans Neural Syst Rehabil Eng. 2006;14(2):131–4.

128. Daly I, Williams D, Kirke A, Weaver J, Malik A, Hwang F, et al. Affective brain-computer music interfacing. J Neural Eng. 2016 Jul 11;13(4).

129. Ehrlich SK, Agres KR, Guan C, Cheng G. A closed-loop, music-based brain-computer interface for emotion mediation. PLoS One. 2019 Mar 1;14(3):1–24.

130. Thompson MC. Critiquing the Concept of BCI Illiteracy. Sci Eng Ethics [Internet]. 2018;25(4):1217–33. Available from: https://doi.org/10.1007/s11948-018-0061-1

131. Segalowitz SJ, Santesso DL, Jetha MK. Electrophysiological changes during adolescence: A review. Brain Cogn [Internet]. 2010;72(1):86–100. Available from: http://dx.doi.org/10.1016/j.bandc.2009.10.003

132. Betts J, Mckay J, Maruff P, Anderson V. The development of sustained attention in children: The effect of age and task load. Child Neuropsychol. 2006;12(3):205–21.

133. Chaplin TM, Aldao A. Gender differences in emotion expression in children: A meta-analytic review. Psychol Bull. 2013;139(4):735–65.

134. Olino TM, Durbin CE, Klein DN, Hayden EP, Dyson MW. Gender differences in young children's temperament traits: Comparisons across observational and parent-report methods. J Pers. 2013;81(2):119–29.

135. Mauer N, Borkenau P. Temperament and early information processing: Temperament-related attentional bias in emotional Stroop tasks. Pers Individ Dif. 2007;43(5):1063–73.

136. Yoshino A, Kimura Y, Yoshida T, Takahashi Y, Nomura S. Relationships between temperament dimensions in personality and unconscious emotional responses. Biol Psychiatry. 2005;57(1):1–6.

137. Rothbart MK, Ellis LK, Rueda MR, Posner MI. Developing Mechanisms of Temperamental Effortful Control. J Pers. 2003;71(6):1113–43.

138. Stadler C, Sterzer P, Schmeck K, Krebs A, Kleinschmidt A, Poustka F. Reduced anterior cingulate activation in aggressive children and adolescents during affective stimulation: Association with temperament traits. J

Psychiatr Res. 2007;41:410–7.

139. Banko M, Brill E. Scaling to very very large corpora for natural language disambiguation. Microsoft Res. 2001;26–33.

140. Holper L, Kobashi N, Kiper D, Scholkmann F, Wolf M, Eng K. Trial-to-trial variability differentiates motor imagery during observation between low versus high responders: A functional near-infrared spectroscopy study. Behav Brain Res [Internet]. 2012;229(1):29–40. Available from: http://dx.doi.org/10.1016/j.bbr.2011.12.038

141. Moghimi S, Schudlo L, Chau T, Guerguerian AM. Variability in prefrontal hemodynamic response during exposure to repeated self-selected music excerpts, a near-infrared spectroscopy study. PLoS One [Internet]. 2015;10(4):1–17. Available from: http://dx.doi.org/10.1371/journal.pone.0122148

142. Hu X-S, Hong K-S, Ge SS. Reduction of trial-to-trial variability in functional near-infrared spectroscopy signals by accounting for resting-state functional connectivity. J Biomed Opt. 2013;18(1):017003.

143. Kleih SC, Kubler A. Psychological Factors Influencing Brain-Computer Interface (BCI) Performance. Proc - 2015 IEEE Int Conf Syst Man, Cybern SMC 2015. 2016;3192–6.

144. Lakey CE, Berry DR, Sellers EW. Manipulating attention via mindfulness induction improves P300-based brain – computer interface performance. J Neural Eng. 2011;8:025019 1-7.

145. D'Mello S, Graesser A. Dynamics of affective states during complex learning. Learn Instr [Internet]. 2012;22(2):145–57. Available from: http://dx.doi.org/10.1016/j.learninstruc.2011.10.001

146. Dijksterhuis A, Smith PK. Affective Habituation: Subliminal Exposure to Extreme Stimuli Decreases Their Extremity. Emotion. 2002;2(3):203–14.

147. Leventhal AM, Martin RL, Seals RW, Tapia E, Rehm LP. Investigating the dynamics of affect: Psychological mechanisms of affective habituation to pleasurable stimuli. Motiv Emot. 2007;31(2):145–57.

148. Sprague SA, McBee MT, Sellers EW. The effects of working memory on brain-computer interface performance. Clin Neurophysiol [Internet]. 2016;127(2):1331–41. Available from: http://dx.doi.org/10.1016/j.clinph.2015.10.038

149. Loughrey J, Cunningham P. Overfitting in Wrapper-Based Feature Subset Selection: The Harder You Try the Worse it Gets. In: Bramer M, Coenen F, Allen T, editors. Research and Development in Intelligent Systems. 21st ed. London, UK: Springer-Verlag London Limited; 2005. p. 33–43.

150. Chandrashekar G, Sahin F. A survey on feature selection methods. Comput Electr Eng [Internet]. 2014;40(1):16–28. Available from: http://arxiv.org/abs/1808.04468

151. Mourad S, Tewfik A, Vikalo H. Data subset selection for efficient SVM training. 25th Eur Signal Process Conf EUSIPCO 2017. 2017;25:863–7.

152. Kaushal V, Sahoo A, Doctor K, Raju N, Shetty S, Singh P, et al. Learning From Less Data: Diversified Subset Selection and Active Learning in Image Classification Tasks. 2018;1–15. Available from: http://arxiv.org/abs/1805.11191

153. McKenzie M, Wong SC. Subset selection of training data for machine learning: a situational awareness system case study. Next-Generation Robot II; Mach Intell Bio-inspired Comput Theory Appl IX. 2015;9494:1–8.

154. Smallwood J, Schooler JW. The Restless Mind. Psychol Conscious Theory, Res Pract. 2013;1(S):130–49.

155. Zafar A, Hong K-S. Detection and classification of three-class initial dips from prefrontal cortex. Biomed Opt Express [Internet]. 2017;8(1):367–83. Available from: https://www.osapublishing.org/abstract.cfm?URI=boe-8-1-367

156. Hong K-S, Zafar A. Existence of Initial Dip for BCI: An Illusion or Reality. Front Neurorobot [Internet]. 2018;12(October):1–21. Available from: https://www.frontiersin.org/article/10.3389/fnbot.2018.00069/full

157. Funane T, Atsumori H, Katura T, Obata AN, Sato H, Tanikawa Y, et al. Quantitative evaluation of deep and

shallow tissue layers' contribution to fNIRS signal using multi-distance optodes and independent component analysis. Neuroimage [Internet]. 2014;85:150–65. Available from: http://dx.doi.org/10.1016/j.neuroimage.2013.02.026

158. Chan J, Power S, Chau T. Investigating the need for modelling temporal dependencies in a brain-computer interface with real-time feedback based on near infrared spectra. J Near Infrared Spectrosc. 2012;20(1):107–16.

159. Russell JA. A circumplex model of affect. J Pers Soc Psychol. 1980;39(6):1161–78.

160. Liberati G, Federici S, Pasqualotto E. Extracting neurophysiological signals reflecting users' emotional and affective responses to BCI use: A systematic literature review. NeuroRehabilitation. 2015;37(3):341–58.

Appendices

Contents

Appendix A: NIRS Affective BCI Studies

Study	Participants	Elicitation	Protocol	Classification Task	Channels/Optode placement	Features	Classification	Results
Tai et al, 2009	10 TD adults	1 pos (+) IAPS image each participant, 1 neg (-) image.	*1 session,* 60 trials (30 each valence). *Each trial:* 30s baseline, 10s activation, 20s rest.	Emotion vs baseline	8 channels over PFC, four on each side.	6 temporal features, 7 time-frequency. Genetic Algorithm to select best 2 features. Different features selection 'problems' – valence, recording site, time interval.	LDA & SVM (no difference found)	*Ranges:* 75-96.7%
Hosseini et al, 2011	5 TD adults	20 (+) images, 20 (-) images, 20 neutral images.	*1 session,* 60 trials (20 each valence). *Each trial:* 3s stimulus presentation, 7s rest.	Positive vs other, negative vs other	52 channels over frontal and fronto-temporal lobes.	Mean (of a 4s interval, starting 1s after stimulus presentation)	Linear SVM	*Averages:* 72.9% for positive, 68.3% for negative *Ranges:* 60.7-83% for pos, 58-79.7% for neg
Moghimi et al, 2012	10 TD adults	78 researcher-selected, 6 participant-selected pieces of music (45s, (+) and (-)). Brown noise (BN) used as a control stimulus.	*4 sessions,* 3 blocks, 12 trials (4 (+), 4 (-), 4 control). *Each trial:* 10s of BN, 45s of music, 5s of BN. Participants rated valence and arousal after each trial.	High Arousal vs neutral and Positive vs Negative	9 channels over forehead, trapezoidal arrangement.	4 single-channel: Mean, slope, variance, change in avg from baseline. 2 laterality: Ratio of signal slope, difference in avg. Selected 48 highest-rated arousal trials, and 24 highest pos and 24 highest neg trials.	LDA	*Averages:* 71.93% for arousal vs neutral, 71.94% for (+) vs (-) *Ranges:* 58.1-90.2% for arousal vs neutral, 62-86.9% for pos vs neg

						Fisher score calculated and used to select top 2 features.		
Heger et al., 2013	8 TD adults	Images from IAPS, sounds form IADS	*1 session,* 30 trials (10 each for VA, Va, and vA) *Each trial:* 35s, each contained 4-6 images, 4-8 sounds. Neutral period after each trial for 30-40s.	3 sets of tasks: - VA, Va, vA vs neutral; - VA vs other, Va vs other, vA vs other; and - VA vs Va, VA vs vA, Va vs vA	8 channels, two diamonds on either side of forehead.	Mean for each channel/chromophore, 4 wavelet transformation coefficients. Mutual information feature selection to remove features that contributed to less than 10% of the total mutual information.	LDA	Task 1: significant for all *best* Task 2: significant for all Task 3: not significant for all
Yanagisawa et al. (2015)	21 TD adults	Images from IAPS	*1 session,* 2 sets of 8 trials *Each trial:* 25s rest preceding and following 25s trial period, 25s trial period had 5 images for 5s each.	Positive vs negative	14 sources, 13 detectors for 42 channels in a grid over the forehead. Narrowed down to 18 channels over the PFC for BCI study.	Mean of the last 20s of the signal.	Hierarchical Neural Network	As high as 96.7%, but 8/21 were lower than 60%
Aranyi et al. (2015)	11 TD adults	Virtual character (real-time feedback)	*1 session,* 6 trials *Each trial:* 20s rest 22s mental arithmetic (view) 22s emotional neurofeedback	Anger vs baseline	8 channels, 4 on either side of the dorsal-lateral prefrontal cortex	Increase in average hemispherical asymmetry during ‘neurofeedback’ part of the trial, compared to baseline during ‘view’ part of the trial	Statistically significant increase in average asymmetry compared to baseline	38 of the 66 analyzed trials (from all participants together) were successful. 8 of 11 participants completed at least 50% of the trials successfully

Aranyi et al. (2016)	18 TD adults	Virtual character (real-time feedback)	*1 session,* 8 trials *Each trial:* 15s rest 40s mental arithmetic (view) 40s emotional neurofeedback	Positive emotion vs baseline	8 channels, 4 on either side of the dorsal-lateral prefrontal cortex	Increase in average hemispherical asymmetry during 'neurofeedback' part of the trial, compared to baseline during 'view' part of the trial	Statistically significant increase in average asymmetry compared to baseline	70 out of the total analyzed 136 trials (from all participants) were successful. 11 of 17 participants completed at least 50% of the trials successfully
Hu et al. (2019)	15 TD adults	30 movie clips, 3 of each emotion type: Joy, gratitude, serenity, interest, hope, pride, amusement, inspiration, awe, love 1 neutral 6 negative 70s avg duration.	*1 session,* Showed all 30 film clips - Start with neutral, then 6 (-), then 30 (+), randomized for type of emotion. 45s rest in between clips.	10 types of positive emotions were collapsed into 3 positive emotion clusters – encouragement, playfulness, harmony.	6 sources, 14 detectors, each in rows across the forehead, for 24 channels.	Mean of last 30s of each clip, broken into 3 10s segments. Used Hb and HbO as separate feature sets.	SVM	For HbO, achieved avg accuracy of 73.79% discriminating between the three positive emotion clusters

Appendix B: End-of-Session Subject Experience Questionnaire

Participant #: _______________ **Session #: _______________**

Please answer the following questions about how you were feeling before, after and during the session.

1. How tired were you at the beginning of the session?

☐	☐	☐	☐	☐
Very tired	Tired	Neutral	Awake	Very awake

How tired are you at the end of the session?

☐	☐	☐	☐	☐
Very tired	Tired	Neutral	Awake	Very awake

2. How were you feeling at the beginning of the session?

☐	☐	☐	☐	☐
Very upset	Upset	Neutral	Happy	Very happy

How are you feeling at the end of the session?

☐	☐	☐	☐	☐
Very upset	Upset	Neutral	Happy	Very happy

3. Did it get harder to focus on the task as the session went on?

☐	☐
Yes	No

If yes, when did it become harder to focus?

☐	☐	☐	☐	☐
After block 1	After block 2	After block 3	After block 4	It was hard to focus the entire time

4. How much effort did you have to use to perform the tasks in this session? (How hard did you have to work?)

☐	☐	☐	☐	☐
Very low	Low	Neutral	High	Very high

5. How much frustration and/or discouragement did you feel while you were completing the tasks in this session?

☐	☐	☐	☐	☐
Very low	Low	Neutral	High	Very high

6. Compared to the last session(s), how much effort did you have to use in this session?

☐	☐	☐
Less	Same	More

7. Compared to the last session(s), how much frustration and/or discouragement did you feel in this session?

☐	☐	☐
Less	Same	More

Appendix C: Early Adolescent Temperament Questionnaire

Early Adolescent Temperament Questionnaire - Revised
Parent Report

Study Title: A novel near-infrared spectroscopy brain-computer interface for the detection of emotional valence in children

Principal Investigator: Dr. Tom Chau, PhD, PEng
Co-Investigator: Erica Floreani, MASc Candidate

Participant Number: ______

Directions

On the following pages you will find a series of statements that people might use to describe their child. The statements refer to a wide number of activities and attitudes.

For each statement, please circle the answer which best describes how true each statement is for your child. There are no best answers. People are very different in how they feel about these statements. Please circle the first answer that comes to you.

You will use the following scale to describe how true or false a statement is about your child:

Circle number:	If the statement is:
1	Almost always untrue of your child
2	Usually untrue of your child
3	Sometimes true, sometimes untrue of your child
4	Usually true of your child
5	Almost always true of your child

NOTE: Please make certain to answer all questions on BOTH SIDES of the pages.

Your son or daughter:	Almost always untrue	Usually untrue	Sometimes true, sometimes untrue	Usually true	Almost always true
1) Worries about getting into trouble.	1	2	3	4	5
2) When angry at someone, says thing s/he knows will hurt that person's feelings.	1	2	3	4	5
3) Has a hard time finishing things on time.	1	2	3	4	5
4) Thinks traveling to Africa or India would be exciting and fun.	1	2	3	4	5
5) If having a problem with someone, usually tries to deal with it right away.	1	2	3	4	5
6) Has a hard time waiting his/her turn to speak when excited.	1	2	3	4	5
7) Often does not seem to enjoy things as much as his/her friends.	1	2	3	4	5
8) Opens presents before s/he is supposed to.	1	2	3	4	5
9) Would be frightened by the thought of skiing fast down a steep slope.	1	2	3	4	5
10) Feels like crying over very little on some days.	1	2	3	4	5
11) If very angry, might hit someone.	1	2	3	4	5
12) Likes taking care of other people.	1	2	3	4	5
13) Likes to be able to share his/her private thoughts with someone else.	1	2	3	4	5
14) Usually does something fun for awhile before starting her/his homework, even though s/he is not supposed to.	1	2	3	4	5
15) Finds it easy to really concentrate on a problem.	1	2	3	4	5
16) Thinks it would be exciting to move to a new city.	1	2	3	4	5
17) When asked to do something, does it right away, even if s/he doesn't want to.	1	2	3	4	5
18) Would like to be able to spend time with a good friend every day.	1	2	3	4	5
19) Tends to be rude to people s/he doesn't like.	1	2	3	4	5
20) Is annoyed by little things other kids do.	1	2	3	4	5
21) Gets very irritated when someone criticizes her/him.	1	2	3	4	5
22) When interrupted or distracted, forgets what s/he was about to say.	1	2	3	4	5

	Almost always untrue	Usually untrue	Sometimes true, sometimes untrue	Usually true	Almost always true
23) Is more likely to do something s/he shouldn't do the more s/he tries to stop her/himself.	1	2	3	4	5
24) Enjoys exchanging hugs with people s/he likes.	1	2	3	4	5
25) Tends to try to blame mistakes on someone else.	1	2	3	4	5
26) Is sad more often than other people realize.	1	2	3	4	5
27) Can generally think of something to say, even with strangers.	1	2	3	4	5
28) Wouldn't be afraid to try a risky sport like deep sea diving.	1	2	3	4	5
29) Expresses a desire to travel to exotic places when s/he hears about them.	1	2	3	4	5
30) Worries about our family when s/he is not with us.	1	2	3	4	5
31) Gets irritated when I will not take her/him someplace s/he wants to go.	1	2	3	4	5
32) Slams doors when angry.	1	2	3	4	5
33) Is hardly ever sad, even when lots of things are going wrong.	1	2	3	4	5
34) Would like driving a racing car.	1	2	3	4	5
35) Has a difficult time tuning out background noise and concentrating when trying to study.	1	2	3	4	5
36) Usually finishes her/his homework before it's due.	1	2	3	4	5
37) Likes it when something exciting and different happens at school.	1	2	3	4	5
38) Usually gets started right away on difficult assignments.	1	2	3	4	5
39) Is good at keeping track of several different things that are happening around her/him.	1	2	3	4	5
40) Is energized by being in large crowds of people.	1	2	3	4	5
41) Makes fun of how other people look.	1	2	3	4	5
42) Doesn't criticize others.	1	2	3	4	5
43) Wants to have close relationships with other people.	1	2	3	4	5
44) Is shy.	1	2	3	4	5

	Almost always untrue	Usually untrue	Sometimes true, sometimes untrue	Usually true	Almost always true
45) Gets irritated when s/he has to stop doing something s/he is enjoying.	1	2	3	4	5
46) Usually puts off working on a project until it is due.	1	2	3	4	5
47) Is able to stop him/herself from laughing at inappropriate times.	1	2	3	4	5
48) Is afraid of the idea of me dying or leaving her/him.	1	2	3	4	5
49) Is often in the middle of doing one thing and then goes off to do something else without finishing it.	1	2	3	4	5
50) Is not shy.	1	2	3	4	5
51) Is quite a warm and friendly person.	1	2	3	4	5
52) Sometimes seems sad even when s/he should be enjoying her/himself like at Christmas, or on a trip.	1	2	3	4	5
53) Doesn't enjoy playing softball or baseball because s/he is afraid of the ball.	1	2	3	4	5
54) Likes meeting new people.	1	2	3	4	5
55) Feels scared when entering a darkened room at night.	1	2	3	4	5
56) Wouldn't want to go on the frightening rides at the fair.	1	2	3	4	5
57) Hates it when people don't agree with him/her.	1	2	3	4	5
58) Gets very frustrated when s/he makes a mistake in her/his school work.	1	2	3	4	5
59) Is usually able to stick with his/her plans and goals.	1	2	3	4	5
60) Pays close attention when someone tells her/him how to do something.	1	2	3	4	5
61) Is nervous being home alone.	1	2	3	4	5
62) Feels shy about meeting new people.	1	2	3	4	5

Appendix D: Formulae for Accuracy, Sensitivity & Specificity

$$Classification\ Accuracy = \frac{(TP + TN)}{(TP + TN + FN + FP)}$$

$$Classification\ Sensitivity = \frac{TP}{(TP + FN)}$$

$$Classifciation\ Specificity = \frac{TN}{(TN + FP)}$$

where

$TP = true\ positives,\quad TN = true\ negatives,\quad FP = false\ positives,\quad FN = false\ negatives$