Chapter 1

BCI for Mental Workload Assessment and Performance Evaluation in Space Teleoperations

Fani Deligianni[1] Daniel Freer[2] Yao Guo[2] and Guang-Zhong Yang[3]

Astronauts have to complete hundreds of hours of training with simulation systems that help them to improve their ability to operate robotic arms for docking operations. In docking tasks, performed for example on the CanadaArm2, the operator does not have direct view of the international space station but relies instead on visual feedback from multiple 2D camera views. Failure to accomplish the tasks on time costs millions of pounds and can potentially endanger the life of the crew members. Even in simulated tasks of the Soyuz-TMA approach and docking, tension and anxiety build up quickly as the precision required is high and virtual fuels are limited. In this chapter, we investigate how simulation systems can be used as a platform to enhance and measure an operator's performance, as well as to design and evaluate semi-autonomous modes of operation that facilitate effective human robot collaboration. Furthermore, we review how brain computer interfaces can monitor workload, attention and fatigue. These systems can be evolved to provide an intuitive human robot interaction experience that provides guidance and feedback as they are needed.

1.1 Human Robot Interaction in Space – What we learn from simulators

Astronauts spend hundreds of hours of training on simulators in order to acquire the necessary skills and abilities to follow procedures that are vital to their survival in space. Furthermore, they need to acquire expert knowledge to maintain and interact with complex electronic and robotic systems. In order for astronauts to retain the skills they learned from the simulator and to generalize or transfer to the operational domain, the training simulators used must be high fidelity [5]. This reflects the fact that crew needs to train to handle low likelihood events that are time critical. Several simulators mimic the extreme conditions of outer space, such as a lack of gravity, oxygen and pressure, as well as very low temperatures. Examples of specialized simulation facilities include the thermal vacuum chamber at the NASA

[1] School of Computing Science, Glasgow University, UK. (email: fani.deligianni@glasgow.ac.uk)
[2] Hamlyn Centre, Imperial College London, UK.
[3] Institute of Medical Robotics, Shanghai Jiao Tong University.

Goddard Space Flight Center and the Large Space Simulator at the European Space Agency. These systems are designed to physically and mentally prepare the crew as well as to examine material properties and tolerance in extreme conditions. On the other hand, several other types of simulators have been developed to train astronauts to use complex technology to perform surveillance, maintenance and assembly tasks, as well as how to dock modules, teleoperate robots and fly a spacecraft [1, 6]. Inherently, simulation systems provide a testbed in several aspects, which includes system testing and verification, ergonomics in human machine collaboration and neuroergonomics in closed-loop interactive interfaces.

System testing and verification are of paramount importance in both fully automated and semi-automated systems. Real-time, complex distributed systems that combine both hardware/sensing components along with software components are notoriously difficult to design and test. Engineers rely on simulation systems to provide realistic test cases of rare and dangerous scenarios that aim to highlight shortcomings and failures. Among the most well-known space flight simulators are the STS-133 simulator and the vertical motion simulator at Ames Research Centre [6].

The aviation industry is a representative example of the challenges to address. Although the industry complies to the highest standards and there is accumulated experience of several decades, catastrophic failures continue to highlight the downside of complex automated systems. In March 2019, Boeing 737 Max 8 and 9 jets were grounded following two deadly accidents. Apparently erroneous sensor readings triggered an automated system response, which could not be controlled by the crew [7]. From one perspective, automation has helped to improve safety, but the difficulty to comprehend the complexity of the systems along with an inherent lack of transparency of current machine learning algorithms hinders a widespread adoption of fully automated systems [8].

To this end, human machine interaction via intuitive designs [9] could empower humans and create augmented artificial intelligence frameworks that encompass expert knowledge along with data science. Simulators in these scenarios are important to perform the so called human-in-the-loop tests [10]. In particular, docking modules on the ISS are of particular significance because they require a good spatial understanding of the environment, even when presented with limited information.

Although human error is estimated to cause over 60% of fatal accidents in aviation, well trained crew is acknowledged to be critical to the overall system's safety [10]. There is extensive research in ergonomics to design human centered systems, and towards this end appropriate simulation environments are of paramount importance. Towards this end, cognitive workload data obtained in simulations provide valuable insight on how to design efficient interactive frameworks that minimize fatigue and improve performance.

In this chapter, we examine how realistic simulation environments used typically in teleoperated robots in ISS and docking tasks can facilitate the development of testbeds for augmented human and AI systems that sense human neurophysiology and allow them to react in real-time via intuitive designs, such as displays, haptic and auditory feedback.

1.1.1 Soyuz-TMA

The Soyuz-TMA is a spacecraft used by the Russian Federal Space Agency to launch missions from Earth to space. Currently, it is used to carry astronauts from and to International Space Station (ISS) and it is considered to be one of the safest and most cost-effective spacecrafts in operation. Astronauts are trained to operate the Soyuz-TMA in a number of modes, which include prelaunch preparation, insertion to the orbit, orbital maneuvering, approaching and docking at ISS, undocking, reentry to the atmosphere from orbit and landing. The onboard control system is based on a camera and a periscope view, whereas a KURS AM radio system provides information with regard to the relative velocity, attitude and distance of the spacecraft to the docking station. Some of the key data displayed along with the camera and periscope views include approach distance and velocity, rates of rotation for attitude stabilization and line of sight angle for alignment. During docking, it is important to maintain rotational and translational velocity within safe limits and the docking target should also be closely aligned with the spacecraft centerline. The Soyuz-TMA can operate either in automatic mode or can be switched to manual mode when the automatic system fails to dock. In this case, the spacecraft has to abort its approach, move backwards and try again. Due to the lack of gravity once a thruster exerts a translational or rotational force on the spacecraft, the spacecraft will continue moving or rotating unless the force is counterbalanced. Therefore, successfully docking the module to the ISS requires calculated movements and precision to approach the docking station without excessive forces and to avoid running out of fuel.

Figure 1 shows a simulator for the Soyuz-TMA approach and docking. The docking has two levels of difficulty: i) docking the spacecraft in the ISS hatch directly in front it, which does not require rotational maneuvering and ii) docking the spacecraft in one of the left/right/bottom ISS hatches, which involves activating first the translational and rotational thruster. Rotating the spacecraft is more challenging since excessive forces could result in spinning around in a difficult to control way. Another challenge related to the camera and periscope views is that the motion is a mirrored translation of the camera’s view; you move left whereas the docking target in the periscopic view is moving right, creating a perceptual conflict which can confuse the user if he is not adequately trained. Furthermore, there is no intelligent guidance to help the user understand what values could be the optimal for navigating.

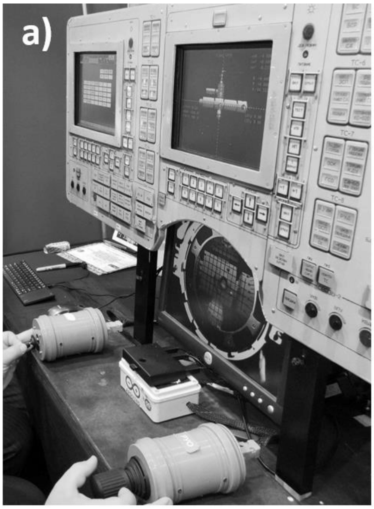


Figure 1: a) A Soyuz-TMA simulator at the UK Space Conference 2019 at Wales. [Photo credited to the British Interplanetary Society.]

For example, intuitive information regarding the 3D position and orientation of the Soyuz-TMA along with optimum navigation paths could allow users to learn faster.

1.1.2 Canadarm2 and Dextre

Perhaps the robots that have been most utilised in space have been the large robotic arms outside of the ISS, such as the Space Station Robotic Manipulator System (SSRMS), which is also known as the Canadarm2, and the Special Purpose Dexterous Manipulator (SPDM), also known as Dextre [11]. Together with the Mobile Base System, they form the Mobile Servicing System (MSS), the most used tool for On-Orbit Servicing (OOS) in space. The SSRMS, a 7 DoF, 17 meter long robot, was largely designed to perform ISS assembly tasks and to capture visiting vehicles and other similar large-scale tasks [12]. Dextre, as a 2-armed robot, was made to perform external maintenance on the ISS. Dextre has the ability to use many different tools to robustly perform smaller-scale and more dexterous tasks than the SSRMS. Each arm has 7 independently controllable DoF with an ORU/Tool Changeout Mechanism (OTCM) as their end-effector [13, 14].

One of the most effective aspects of these robots are the latching end-effectors, which were specifically designed to firmly latch onto grapple fixtures strategically placed in many locations around the ISS and on incoming cargo and spacecraft. There are several types of grapple fixtures which allow for different capabilities. For example, the Latchable Grapple Fixture (LGF) is intended for longer-term stowage on the POA, while the Power and Video Grapple Fixture (PVGF) additionally allows for access to data, video, and power. These innovations have given versatility to the robots, allowing the SSRMS to relocate its base by "walking" from fixture to fixture around the station. This concept has also given a mechanism for combining robots together in a macro-micro configuration, where for example the SSRMS

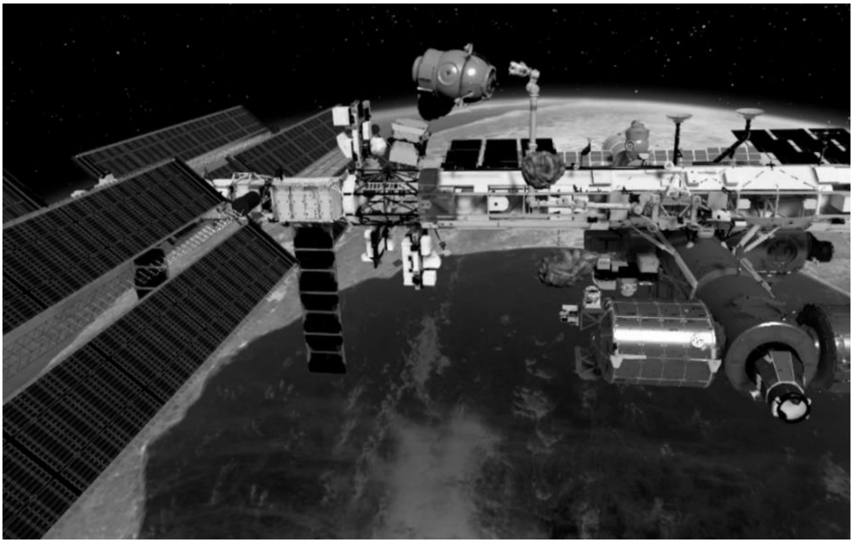


Figure 2: A photo-realistic 3D simulator of the International Space Station (ISS) developed at Imperial College London [1]. The simulator allows the user to interact with the CanadaArm2 robot based on four camera views as shown at Figure 3.

(Canadarm2) would position the SPDM (Dextre) in the best location while the SPDM performs the smaller scale tasks [12, 14].

The SSRMS can be controlled in multiple modes: joint control, end-point control, and automatic trajectory control. When using the different control modes the user must constantly consider several different frames of reference and coordinate systems [13], while looking at multiple camera angles to determine the best course of action to complete a given task. A certain level of autonomy has already been implemented into the SSRMS, such as the automatic trajectory control mode, which has allowed ground control to overcome the latency to the ISS and perform simpler tasks, including preparation and initial positioning of the robot arms [13]. The SPDM, on the other hand, is now mostly controlled from ground control on Earth despite the fact that it was initially designed to be controlled from within the ISS [15]. This is made possible through a series of tests called On-Orbit Checkout Requirements which ensure the safe operation of the SPDM even with significant amounts of latency.

Training to use the Canadarm2 and Dextre occurs at the Robotics Training Centre of the Canadian Space Agency in Saint-Hubert, Quebec. The training centre includes a replica of the Robotics Workstation that is on the ISS and sophisticated training and simulation software [16]. While it is difficult to find information about the specific software used in this training centre, research groups have developed their own similar software to investigate the best ways to train astronauts and flight controllers how to control these robots. Belghith et. al., for example, developed the Robot MANinpulation Tutor (RomanTutor), which, in addition to providing a general simulation platform for practicing robotic control, did automatic path planning for a particular task and considered strategies for camera and view selection [17]. These tools could then be used to show trainees what an ``optimal" solution to task performance may be, even in complex and ill-defined domains. Using a previous iteration of this simulator, Fournier-Viger developed a cognitive model of

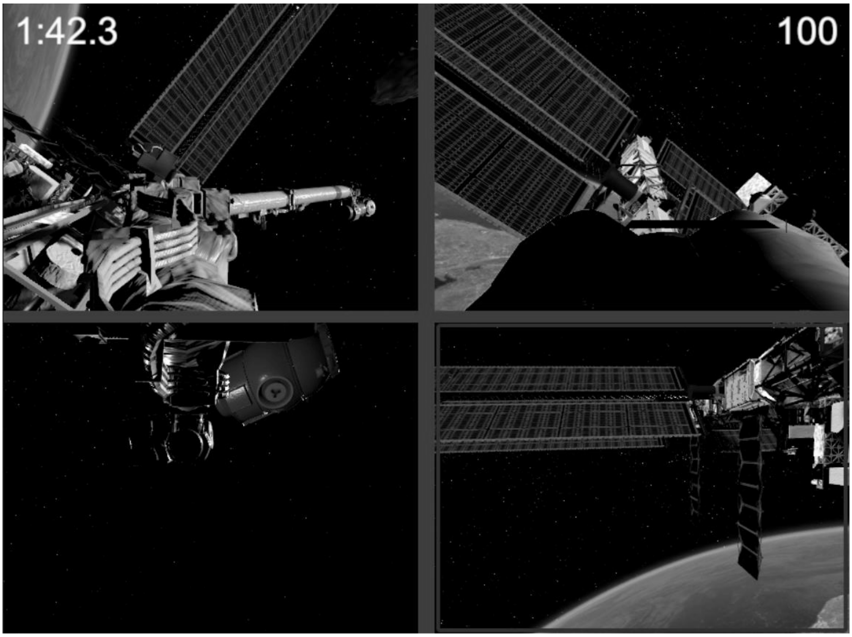


Figure 3: In order to control the CanadaArm2, users rely on 2D views of cameras located at close proximity to the joints of the CanadaArm2 and other key locations at ISS.

Canadarm2 task performance to break down a complex and ill-defined task into more understandable steps, evaluating spatial representations [18].

Another simulator for the Canadarm2 is available on the CSA website [19]. This simulator plays more like a game, teaching the different aspects of Canadarm2 control before allowing you to attempt a ``mission" utilising the skills that were learned in the tutorials. Control is carried out with a keyboard and mouse, allowing the player to move and turn the Canadarm2 to match various given visual cues. Training tasks within this simulator include following a circular trajectory with the end-effector using the camera placed on the end-effector, rotational control for precise docking, and finally a complete task in which the Canadarm2 is controlled to perform a task aboard the ISS involving the replacement of a component, and carrying an astronaut to perform a final task at this component. While this simulator provides insight into many of the necessary skills for Canadarm2 operation, it also provides real-time suggestions about which keys to press to successfully achieve a given outcome. This results in minimal cognitive load, and is likely be difficult to translate to real robotic control.

Most recently, a Canadarm simulator was developed at Imperial College London, Figure 3, which aimed to provide both photorealism and increased cognitive load [1]. This simulator was built to allow for different grades of cognitive load through the addition of confounding factors such as latency, time pressure, and pieces of space debris which acted as obstacles to avoid. The simulator was also developed to be compatible with physiological data collection, and thus was paired with modules to collect EEG and eye-tracking data, as well as information about heartrate, body temperature, and Galvanic Skin Response (GSR). The collected data was synchronized via LSL and analysed to determine the effect of the added workload from each of the proposed confounding factors. Measures from these sensors were additionally compared to a task performance score, which considered the user's precision at each stage of the task, time to complete the task, and any errors or collisions that occurred.

1.2 Cognitive Models Underlying NeuroErgonomics in Space Flight

In the field of human factors and ergonomics, there is extensive research literature on how to develop human-centred designs of technology that aim to minimise errors, enhance performance and enable effective human-machine interaction [20]. The recent expansion in artificial intelligence (AI) and the success of these systems in information retrieval, robot vision and language processing automate low-level applications [21, 22]. These systems are gradually adapted to everyday life and have already automated several manual tasks. However, how these systems can interact for higher-level decision making and whether we can trust them remains a challenge. In addition, safety concerns and ethical considerations with relation to the underlying responsibility are profound. These factors imply that recent progress will translate several applications from manual to semi-automatic, and thus designing supervisor control mechanisms that take into consideration human factors and ergonomics are in high demand.

Neuroergonomics are concerned with human brain function and performance in a number of critical applications that range from medical interventions, aviation, driving and so on [23-25]. Experts in the field predict that within the next 20 years, neuroimaging technologies involved in human cognitive augmentation would mature to seamlessly allow monitoring and enhancement of brain processes [23]. Brain functions, such as decision making, cognition, attention, vigilance and situation awareness are important to complete a task successfully. In neuroergonomics, there is a distinction between vigilance, also referred as sustained attention, and attention under workload. Vigilance is usually tested under low workload and it involves the ability to detect a stimulus, which is important in applications such as air traffic control, surveillance and inspection tasks. Usually, it is evaluated based on the reaction time, which is the time from the stimulus presentation to a simple motor response of the subject. On the other hand, maintaining attention and situation awareness under workload involves shifts of attention via higher cognitive functions and executive control. In these scenarios, spatial attention, which refers to the ability of orienting attention to a particular

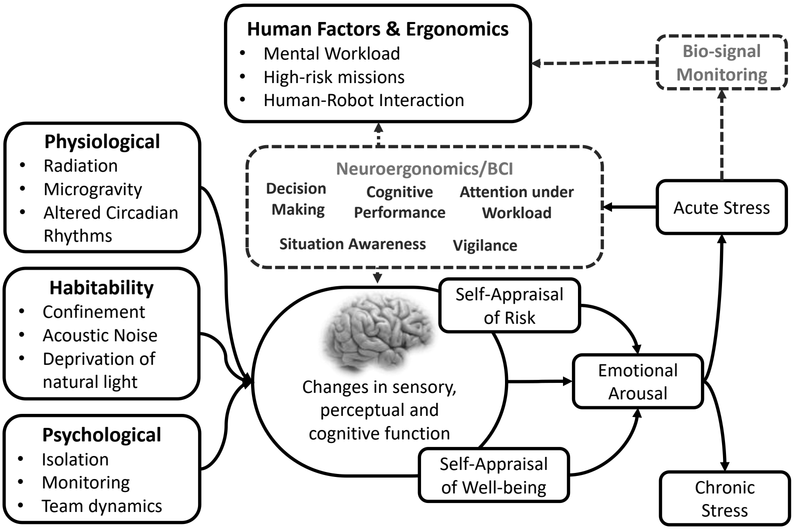


Figure 4:Human Factors and NeuroErgonomics in Space Flights. Current research needs to be seen in light of the homeostatic adaptations that take in space due to physiological, psychological and habitability factors in space.

direction, and theories based on working memory models have been employed to explain the information processing that underlies brain function.

These functions are supported by interconnected circuits that involve several brain regions and they are significantly affected under stress and workload. In particular, the prefrontal cortex (PFC) has been implicated in attention control, concentration, executive function and decision making [25-27]. Furthermore, ventromedial PFC and dorsomedial PFC are implicated in social processing and anxiety [28]. In sustained attention both top-down and bottom-up networks act in parallel to facilitate selective attention [29-31]. Selective attention is referred to as the ability of the brain to prioritise sensory information. The former originates from forebrain regions that include the PFC, the parietal cortex, somatosensory cortices and subcortical structures, such as thalamus and basal ganglia. It encodes brain states related to working memory, reinforcement learning, selection of task-relevant processes and inhibition of task-irrelevant processes [32]**.** On the other hand, the midbrain network is thought to exert a bottom-up regulation of attention related to the saliency of the stimulus. This process is thought to have an evolutionary purpose in order to alert humans when a ‘threatening’ stimulus, as for example when a pop-up stimulus enters their peripheral vision, whereas the forebrain enables humans to concentrate on a particular target.

In humans, only a small fraction of the visual field corresponding to the fovea is perceived in fine detail. Visual spatial attention refers to the ability to select relevant objects/stimulus and process information within the underlying area of the visual field [33]. Spatial attention results in increasing the gain of the mean firing rate and decreasing the noise correlations across neuronal populations related to the relevant location/objects [34]. This improves detection and discrimination of relevant stimulus and shortens reaction time. The processing of visual information is complex and involves the ventral and dorsal neuronal pathways that start from the primary visual cortex and extend to the temporal lobe and parietal lobe, respectively. The former pathway is involved in object recognition, whereas the dorsal pathway is mostly related to spatial awareness and spatial attention [33].

Damage of the spatial attention pathways could result in directional bias in orienting attention. This is a common clinical syndrome called spatial neglect. It is thought that spatial neglect is the interaction of several deficits that involve, directional bias in competition for selection, spatial working memory deficits and sustained attention deficits [35, 36]. Although, the exact brain regions involved in visual selective spatial attention is under debate, the Parietal Eye Fields (PEF), Frontal Eye Fields (FEF) and the Temporal Parietal Junction (TPJ) have been highlighted as most likely to play an important role [35-38].

Eye movements and spatial attention are interrelated, since it is evident that usually eye movements follow attention (overt orienting gaze). In fact, neuronal circuits that control attention are also related to eye movements/saccades [31]. For example, animal studies have shown that after unilateral removal of the FEF in prefrontal cortex, the animal could not direct gaze in the affected hemisphere. Overt orienting can be either reflexive or controlled. Reflexive movements are related to midbrain, bottom-up attention selection mechanisms rather than conscious processing of the visual field. In this context, the consensus is that there is a single mechanism that drives both selective attention and motor preparation. Nevertheless, humans are able to mentally direct their attention to spatial locations without moving their eyes (covert orienting gaze). This has been found to slower saccades and to alter the underlying perceptual processes.

1.2.1 Neuroergonomics and Spatial Attention

Several theories have been developed to explain and model how brain processes information with relation to attention and decision making [30, 39]. These models introduce the concept of working memory, which describes the ability of the brain to hold a limited amount of information for a short period of time while it is processed [29]. Working memory does not only refer to the ability to memorise but also the ability to suppress irrelevant information. Stimuli compete to gain control over working memory, whereas gaze and spatial attention processes play an important role in it, Figure 5. Visual, auditory and haptic information enter the ‘short-term sensory stores’ (STSS), which retain information of up to a fraction of a second [39]. STSS is thought to have a large capacity but short duration, whereas working memory can persist for several seconds but is restricted to few items [39]. This is also referred to as an ‘attentional blink’ and is the reason that stimulus experimental setups allow a 300 millisecond gap. This restricted attentional capacity emerges within sensory modalities. In other words, concurrent attention to visual stimuli limits attention to another visual stimuli but it does not limit concurrent attention to auditory stimulus [40]. Furthermore, it was shown that processing and recognition of a scene takes 100 milliseconds, though work has showed that humans are able to

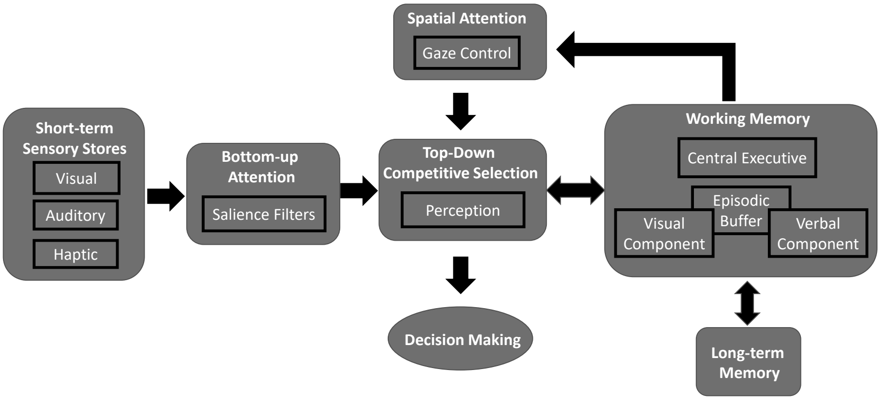


Figure 5: A consensus cognitive model that show how attention and working memory interact to process information.

recognize scenes better than chance even with rapid stimulus presentation of less than 50msec [41, 42].

1.3 Workload and Performance Measures in Human-Robot Collaborative Tasks

There is an overwhelming amount of evidence to suggest that workload and performance are strongly related. In fact, when humans face increased demand their performance may deteriorate, they will perform more errors, their tasks will be less accurate, they often lose awareness of their surroundings and they become increasingly frustrated and fatigued [39]. However, the relationship between workload and performance measures is not linear. Lower workload than normal can also result in similar performance deterioration, possibly due to boredom and drifts of attention. Therefore, it seems that there is a subject-specific point of workload where performance reaches its maximum, which normally reflects that the task is challenging, yet does not overwhelm the operator. Overall, human performance depends on a number of factors that include time pressure, operator's fatigue, training level, innate abilities to adapt to the task in hand and resilience to stress and anxiety. Furthermore, workload and performance may need to be compromised to satisfy operational goals.

Performance in human-robot collaborative tasks should take into account the ability of the human and also consider the ability of the robot to adapt and optimise its actions with relation to human responses. There is considerable effort in the research community to develop human-robot collaborative strategies to ease the cognitive and physical load and thus minimise workload. However, there is a risk that with increased automation, humans become observers and they are not actively engaged in the loop. This can also cause boredom, loss of awareness and lack of the ability to comprehend the complexity of the system. All these factors can result in the inability of the human operator to control the system if the automated mode is inaccurate or fails.

Mental workload can be sub-categorised based on human sensing and cognitive processes into visual, auditory, tactile and cognitive workload [43]. Visual perception is influenced by contrast, colour, dark-adapted vision, depth perception, movement detection and glare. Hearing perception is also influenced by loudness, pitch and location. Finally, with tactile sense, we perceive differences in temperature, pressure, and the frequency of vibrations of our skin. Human capacity to absorb and process information is called perception and is influenced by internal cognitive models and expectation. Human perception has the ability to fill as well as remove information based on contextual information and this mismatch between reality and perception of sensory information could lead to misinterpretations and illusions.

Characterising workload is an important yet quite complex process. Most common subcategories include mental/cognitive demand, physical demand, temporal demand, performance, frustration and effort. In fact, the NASA Task Load Index (TLX) has adopted this method to measure workload via subjective self-reports. Self-assessment reports are normally used shortly after the task in hand while user’s memory is still fresh. To overcome the fact that subjects perceive types of workload differently, the TLX index requires them to rank the order of each subcategory. Subjective assessments are disruptive, they are not continuous, and they suffer from scaling problems, since most operators do not translate increases in workload, linearly.

Task-specific performance measures provide an objective way to measure workload based on the assumption that the ability to perform a task well is affected by workload. Primary task measures include task analysis, speed, accuracy and levels of activity. Task analysis includes various methods that break user’s actions into sub-tasks, and they count how many are completed successfully. Measuring activity involves counting the number of steps per time required to finish the task. Actions may include control inputs, verbal responses, mental arithmetic, visual searches and decisions. Large numbers of measured activity imply a high workload. Task analysis and measuring activity requires the ability to break down the task and response, respectively, into specific modules, which might not be trivial in real-world dynamic environments. Speed and accuracy are the simpler performance measures to estimate but they cannot disassociate operator condition from system failures, such as slow response. Furthermore, workload in decision making tasks is difficult to characterize based on speed and accuracy alone. None of these measures take into account the skills of operators.

The workload is also modulated under single or multiple task demands. Secondary task measures estimate the workload of a task by looking into how well the operator performs a second task simultaneously. These techniques quantify how many ‘spare resources’ the operator has and provide more information about the condition of the operator. However, they rely on the assumption that both tasks are competing for the same resources and that the performance of the first task remains constant. Furthermore, different operators may have different strategies to complete the first or second task. Careful design of the interaction of two tasks is important to provide meaningful conclusions.

A recent study has used a dual task design to understand the effect of engagement on workload and performance during driving [44]. The driver was instructed to maintain a specific distance from the vehicle in front, which was changing its speed at random. Primary performance measures included speed control and braking response time, while the NASA TLX index was also recorded after task completion. An auditory stimulus, which was selected to be interesting, boring or neutral was playing at random during the task. The results revealed that the time required to brake (response time) was longer while the driver was listening to an interesting stimulus. Also, the drivers perceived the interesting auditory stimulus to be less demanding, although the stimulus has been objectively chosen with similar difficulty index.

It is important to note that auditory processing channels are considered to be independent from visual processing channels. This is the reason that in ergonomics studies they have been suggested as effective communication channels when the primary task requires visual attention. Nevertheless, further processing of the information would require the allocation of more cognitive resources.

Dual-task designs are powerful as they can disentangle the influence of multiple sensing and cognitive pathways. For this reason, they have been used extensively in several studies in aviation, driving and surgery. It should be noted that the ability of astronauts to successfully perform dual tasks is affected both during the early adaptation to microgravity and towards the end of the flight. It is thought that the cause of this deficit is due to the increased fatigue and stress associated with both of these phases.

Fatigue and sleep deprivation have been associated with several serious accidents that resulted in collisions in space. In 1997, the Progress spacecraft collided with the MIS space station and caused extensive damage on the solar array modules. Although the astronauts claimed that there was a delay in the navigation system, NASA attributes the accident to workplace stress, fatigue and sleep deprivation. Space imposes unique challenges on astronauts that also result in an increased level of fatigue. For example, 60-80% of astronauts will be affected by space motion sickness, micro-gravity also affects their sleeping patterns, along with background noise, lack of adequate thermal control, lack of fresh air and so on.

Fatigue could be muscular or mental and is caused by prolonged physical or mental tasks, respectively. When it relates to emotional stress, fatigue can also be characterised as acute or chronic. There is no clear distinction between workload and fatigue. A problem of disentangling workload from fatigue is that there is no clear definition. Furthermore, most studies do not distinguish fatigue from sleepiness, since it is far more difficult to disentangle one from the other. Countermeasures of fatigue target regulation of circadian rhythm with scheduled sleep breaks along with well scheduled meals. 20-30 minutes sleep before night shifts helps to increase alertness along with administered caffeine and/or other pharmacological agents. Robotic exoskeletons have been suggested to tackle muscle fatigue by providing support to both the lower body and upper body [45].

Several measures have been proposed to quantify fatigue. The occupational fatigue inventory represents fatigue in five physical dimensions (lack of energy, physical exertion, physical discomfort) and two mental dimensions (lack of motivation and sleepiness). Other measures, such as the occupational fatigue exhaustion recovery scale, quantify the need for recovery. This measure describes fatigue in three scales that include acute fatigue, chronic fatigue and intershift recovery.

1.4 Brain Computer Interfaces in Workload and Attention

1.4.1 EEG based BCI

Electroencephalography (EEG) is a sensing modality which measures electrophysiological signals coming from the brain. EEG sensing systems can come in many forms, from implantable systems requiring surgery [46] to dry, wearable caps that favour efficient setup as opposed to optimal signal to noise ratio. Because most healthy individuals would not require or want invasive brain surgery for the monitoring of workload, this chapter includes consideration of only wearable non-invasive EEG systems.

Much of the foundational work related to human EEG recordings was established in a series of 15 reports by Hans Berger [47]. Berger investigated many of the fundamental questions related to EEG recordings, such as types of electrodes, recording locations, artifact removal, and Fourier analysis of the recorded signals. There has been significant development with regard to each of these topics in the last century, but there is still much to learn about what EEG signals can indicate and how they can be used.

One of the first considerations for EEG data recording is the removal of unwanted

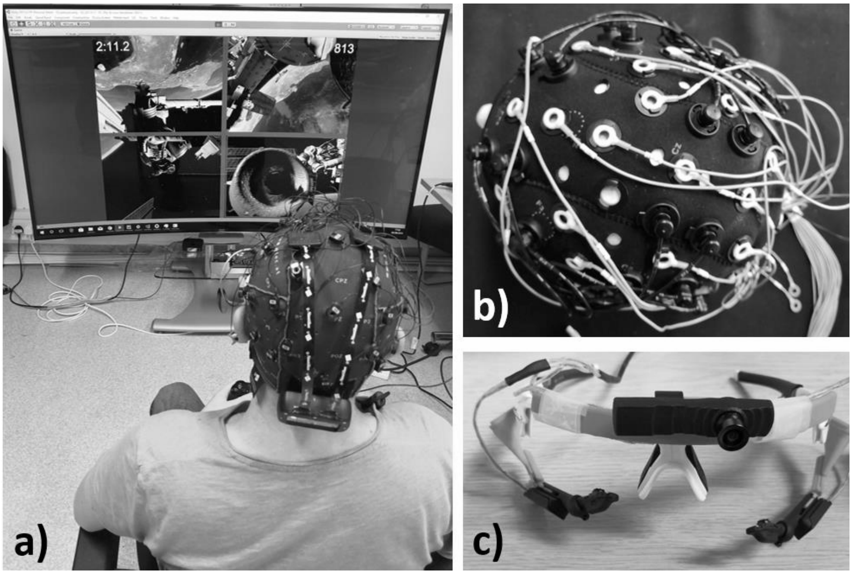


Figure 6:Brain Computer Interfaces are coupled with advanced 3D space simulator to monitor human attention, cognitive workload and spatial learning. a) An advanced 3D simulator has developed to allow users to control canadarm2 under realistic scenarios [1, 2]. Wearable EEG and eye-tracking technology is combined to monitor neurophysiological responses. b) Combined fNIRS and EEG technology promises to improve our understanding on learning processes and cognitive workload, c) An example of wearable eye-tracking device.

artifacts that are picked up by the electrode which may contaminate the signal. There are several known artifacts which originate from the eyes, muscles, heart, and environmental factors which researchers have attempted to remove based on knowledge of the artifact [48]. One way of doing this is by simultaneously recording data from a separate reference and altering the recorded EEG signal based on the recorded reference. When using the same type of electrode, the reference signal could be simply subtracted from the recorded signal. Other methods include the use of unsupervised learning algorithms such as principal component analysis (PCA) and independent component analysis (ICA), and the use of filters to remove either known or learned noise from the environment [48].

Wearable EEG systems typically use multiple electrodes which can provide spatial information about a person's ongoing cognitive activity. The 10-20 system is an international standard for electrode placement, which labels each electrode placement based on region such as Frontal (F), Central (C), Parietal (P), Occipital (O), and Temporal (T) regions. The system also labels electrodes based on the longitudinal fissure, where all electrodes along the fissure have a z placed after their regional marker. An example of a labeled wearable EEG system can be seen in Figure 6a. This labeling standard makes it easier for studies to easily compare their results even if they are using different EEG recording platforms.

Consideration of the spatial location of EEG electrodes is often not straightforward because the source of a particular brain response is not always known. For this reason, the spatial relationship between different electrodes can be numerically defined, which could provide a more intuitive understanding of the meaning of changes in various features. For example, researchers may want to define a spatial filter that only allows for consideration of the frontal regions of the brain. Therefore, in this instance electrodes nearest to the frontal region would have most influence over the signal, while electrodes that are far away may be entirely removed from the signal. These spatial relationships between electrodes can also be learned through algorithms such as Common Spatial Patterns (CSP), which can reduce the feature space for a classifier. However, the feature space should only be reduced in such a way that maximises its practical utility. For example, using miniaturised EEG caps with few channels may not need any further spatial filtering, but could provide information that is unclear due to noise and cannot be verified by comparison with neighbouring electrodes. From this perspective, utilising more electrodes is more practical, despite the fact that features may not be directly extracted from each electrode.

As compared to other brain sensing modalities, EEG is desirable because of its high temporal resolution, with recording frequency up to 1000 Hz. Such high rates allow for analysis of not only temporal features, but also features that are calculated in the frequency domain, such as Power Spectral Density (PSD). Use of the Fourier transform in EEG processing is common, as the signal's changing frequency profile is often indicative of changes in mentality.

EEG signals are commonly separated into several broad frequency bands such as the delta (1.5-4 Hz), theta (4-7.5 Hz), alpha (7.5-12.5 Hz), and beta (12.5-30 Hz) bands. While approximate values are given here, different studies may use slightly different values for the boundaries of these frequency bands. It has been indicated, for example, that a decrease in alpha power is associated with an increase in mental arousal, resource allocation or workload, while theta power tends to increase along with task requirements [49]. These power changes are also most noticeable in pre-defined regions, with alpha decreases being mostly noted in parietal regions, and theta increases being noted in frontal regions. However, research has indicated that mu and alpha rhythms increase in power for humans in a microgravity environment, so previous assumptions about changes in EEG signals may need to be adapted based on new research into the effects of microgravity [50, 51].

EEG has been used to measure mental workload in studies of air traffic controllers, airline pilots, drivers, and a wide range of humans performing cognitive tasks, such as memory or visuospatial tasks [52]. One of the most common tasks for workload measurement has been the n-back task, in which users have to remember whether the currently displayed stimulus was the same as the stimulus shown n trials ago. The difficulty of this task can be modulated simply by changing the value of n, so different methods of measuring mental workload can be easily validated and compared [49, 53]. Other ways of validating workload include relatively simple pre-defined simulated tasks. Because the tasks are simulated, parameters can be easily changed to influence task difficulty. The signals for each task difficulty level can then be compared to make data-based theories about how brain function changes with regard to task workload.

One typical way of validating a method of workload measurement is via classification, where preprocessing leads to feature selection, and eventually the selected features are fed through a classifier such as SVM [49] or discriminant function analysis (DFA) [52]. The ability of the model to determine workload is thus evaluated based on the accuracy of the classifier and other related metrics. If the model is evaluated to work well in classification, then the features that led to higher accuracies can be analysed to draw conclusions about brain function with increased workload.

1.4.2 fNIRS based BCI

Functional Near Infrared Spectroscopy (fNIRS) is a non-invasive, optical neuroimaging technique that allows the measurement of oxygenated hemoglobin (HbO) and deoxygenated hemoglobin (HbR). fNIRS light sources are normally arrays of LEDs or lasers that emit light in at least two wavelengths. The light penetrates the scalp and cortical regions and its relative absorption is measured by the fNIRS detectors. This allows the detection of relatively small changes in near-infrared light absorption, which relates to changes in HbO and HbR according to Beer-Lambert law [54].

fNIRS is resilient to eye-movement artefacts and this is one of the reasons it has been particularly common to measure prefrontal activation, which is related to executive function and its function has been found to be modulated with workload and training. Furthermore, miniaturised sensing technology has allowed fNIRS systems to become portable and wireless. This facilitates the continuous acquisition of brain signals in real-world settings. However, fNIRS electrodes only measure light within a few centimeters from the scalp, which limits the application of fNIRS to cortical regions only.

Several fNIRS studies aimed to replicate and confirm findings of functional magnetic resonance imaging (fMRI), which is the gold standard in functional neuroimaging and has shown average increases in oxygenation with increased workload/difficulty. Typically, in these studies, workload is modulated by N-back tasks [55]. In these scenarios, fNIRS data are recorded while the participant is serially presented with stimulus and he/she is instructed to respond when the stimulus matches the nth stimulus ago. N-back tasks have been found to activate the dorsolateral (dlPFC) and the ventrolateral (vlPFC) prefrontal cortical (PFC) brain regions. Furthermore, these studies show increased frontal-parietal connectivity [54].

In neuroergonomics, fNIRS exploits basic neuroscience principles to assess the design of new systems in terms of workload, parameter optimisation to achieve best cognitive capacity and training. The application scenarios span from air traffic control (ATC), aviation pilots and surgery [25]. In ATC, several studies based on fNIRS aim to assess the ergonomics of new human-machine interaction designs along with the number of aircraft that can safely operate in an airspace. In these cases, fNIRS activation would increase with workload up to a safety critical point, where activation is plateaued. It is evident that any further workload will not be safely handled from the operator and the probability of an accident increases dramatically. It should be highlighted that this point cannot be detected based on self-reported measures.

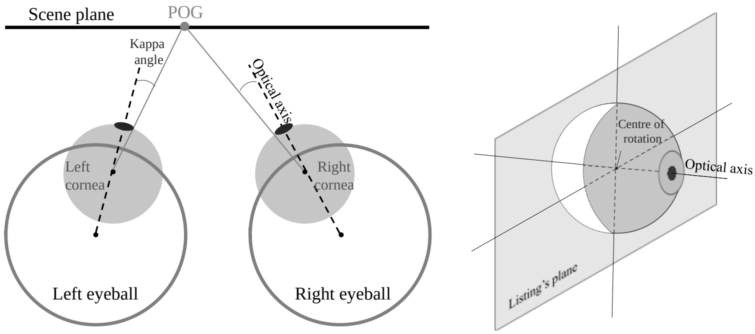
fNIRS activation in the prefrontal lobe with relation to workload is also modulated by the expertise level of the operator. Training results in cognitive adaptation processes that optimise attention control and problem-solving and thus it releases cognitive resources. The PFC activity in experienced operators is reduced compared to novice operators under the same tasks. This finding has confirmed both in ATC as well as in surgical tasks [25]. Therefore, neuroimaging studies offer a way to track the efficiency of training. Few days training in unmanned aerial vehicle piloting highlighted distinct phases of learning. Initially, increased activity in fNIRS reflected increase performance. In later stages of training, increased performance was associated with reduced activity in fNIRS. Similar results have been observed in surgical residence, where it was also shown that training phases is also modulated by the complexity of the underlying task.

1.4.3 Eye-Tracking based BCI

Eye-tracking data (e.g., eye movements and the pupillarity response) can elucidate visual interaction with complex user interfaces, such as where and what the operator looks at, how long the operator looks at it, and which eye movement happens when looks at it. This information also reflects the cognitive workload of the operator during teleoperation. Recently, eye parameters have gained extensive popularity in the estimation of mental workload for those needing to perform complex tasks under stress, such as pilots [56], drivers [57, 58], and surgeons [59, 60]. The most significant workload metrics based on eye parameters can be categorized into pupillary response (meaning pupil diameter and pupil diameter deviation), fixation (number of fixations, fixation duration, and fixation frequency), saccades (speed and amplitude of saccades), and blinks (blink frequency, number of blinks, blink duration) [61].

1.4.3.1 Point of Gaze and Eye Movements

The 3D eye model is typically simplified as the model demonstrated in Figure 7a, which consists of eyeball, cornea, and iris. From this model, the optical axis can be defined as the line that passes through the centres of the eyeball, cornea and iris, and the visual axis indicates the line from the Point Of Gaze (POG) to the corneal centre. There exists a constant deviation, namely *Kappa* angle, between the visual axis and



(a) (b)

Figure 7: a) Demonstration of the top view of two 3D eye models and the point of gaze (POG) on the scene plane. b) Illustration of the listing's plane and the 3D orientation of the eye and its axes of rotation.

the optical axis. The final POG is determined by averaging the estimated gazes of left and right eyes while both eyes are gazing at the same point. In addition, the direction of the optical axis and the rotation of the eyeball can be characterized by Listing's law, which describes the 3D orientation of the eye and its rotation axes by defining a Listing's plane[62]. Specifically, the vertical and horizontal axes of rotation formulate this Listing's plane as illustrated in Figure 7b, and the optical axis that is orthogonal to the plane indicates the torsional rotation. For eye tracking research in space, it should be pointed out that gravity has a critical impact on eye movement and head-eye coordination [63]. It has also been proven that the orientation of Listing's plane significantly changes under microgravity, where the elevation can be tilted backwards by approximately 10 degrees during a parabolic flight experiment [64].

With the successful tracking of the movement of eyeballs, three basic eye movements can be additionally defined as illustrated in Figure 8, including saccades, smooth pursuits, and fixation [65]. A saccade indicates the rapid movement of the gaze point from one position to another, which can also be regarded as shifts between fixations [66]. Fixation indicates the gaze fix or pause on a small region of interest [66]. It can be typically detected when the POG is within a particular area or if the gaze velocity is smaller than a threshold. Smooth pursuit represents the eye movement that follows a moving object [65]. For the teleoperated task described in Freer et al. [1], for example, the saccade movement occurs when the user switches the activated camera or avoids the debris by observing different cameras. The user will achieve fixation while adjusting the robot arm in a fine manner or thinking about the control strategy. During the teleoperation, the users will tend to have smooth pursuit while performing the translation of the end-effector.

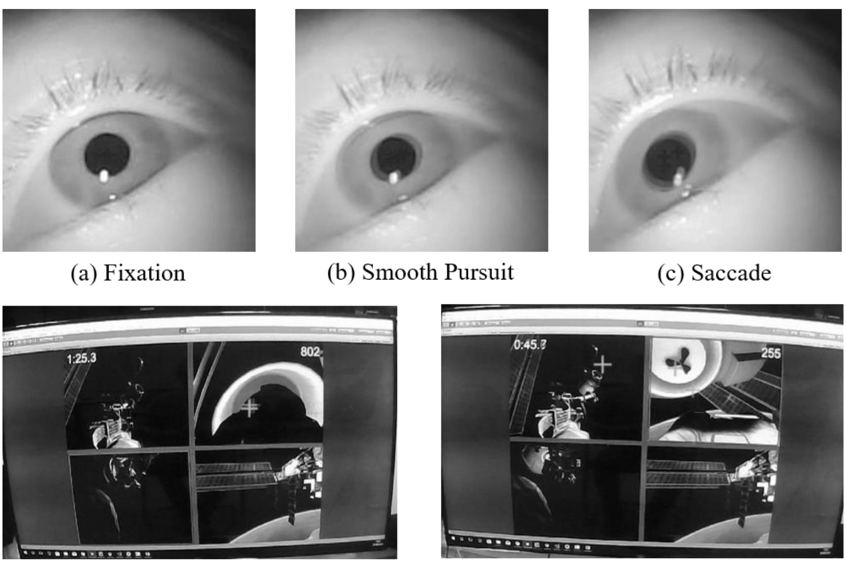
1.4.3.2 Eye Tracking Systems

The existing techniques for eye movement and gaze detection include magnetic search coil, Electro-Oculography (EOG), and Video-Oculography (VOG).

Magnetic search coil: For human eye movement tracking, the subject needs to wear the contact lens that contains coils of wire, which are also known as Helmholtz coils. In the experiments, the subject sits inside a specified area with a magnetic field, then eye movement can induce a variation of voltage in the contact coils. Compared to other systems, the search coil system can achieve high detection precision of eye movement in both spatial and temporal resolution, though it typically causes some discomfort for the human due to its semi-invasive nature and additionally involves a complicated setup procedure [67].

Electro-oculography (EOG): Another popular technique for tracking eye movement is to measure the corneo-retinal standing potential differences between the front and the back of the human eye, which is known as Electro-Oculography (EOG) [68]. EOG is advantageous in measuring the horizontal/vertical rotation of the eyeballs by attaching two surface electrodes to the edges of the orbits along the horizontal/vertical direction. However, the EOG signal is susceptible to noise and cannot measure the pupil diameters.

Video-oculography (VOG): Recent advances in computer vision technologies have led eye-tracking systems to adopt video-based techniques. Most video-based gaze tracking systems focus on the estimation of gaze direction, which can be categorized into remote and wearable gaze trackers. The basic idea is to detect the 2D or 3D parameters of the near-circular pupil from single/stereo camera. Then the rotation of the eye with respect to the camera can be determined after the appropriate calibration procedures. VOG eye-tracking accuracy heavily depends on the quality of the calibration process before recording and the pupil detection during the recording [69]. Pupil detection is more challenging with remote eye-tracking systems as compared to wearable ones due to the corneal reflections, occlusions, and eye blinks. We refer the readers to [67, 68] for more details on various eye-tracking techniques.



(a) Fixation (b) Smooth Pursuit (c) Saccade

(d) Fixation on the image plane (e) Saccade between two visual displays

Figure 8: Illustration of three basic eye movements and the gaze points of fixation and eye saccade on the image planes. Accordingly, features such as fixation duration and saccade speed can be extracted.

1.4.3.3 Eye-tracking based Mental Workload Detection

In the following, we give a brief review of the relationship between mental/cognitive workload and different commonly used eye tracking parameters.

In the last century, pupil diameter has already been used as an index of cognitive workload. Extensive research has revealed that the pupil diameter increases with the task difficulty, which is highly related to the mental workload [70, 71]. One of the remaining challenges is that various confounding factors unrelated to workload, including changes of luminance condition and emotional arousal, may also affect the pupillary response [72]. The Index of Cognitive Activity (ICA) [73] can provide an estimation of cognitive workload level by disentangling pupil dilations caused by cognitive activity (small rapid dilations). The ICA is determined by the changes of pupil dilation while performing a specified complex task. The core idea of ICA is to perform wavelet analysis to identify the abrupt pupil dilation in the eye-tracking data [73]. When the rapid pupil dilation is larger than a specified threshold, it reflects the effect of cognitive activity. Higher ICA levels per second represent a higher degree of cognitive workload [59].

With the recent advances in gaze tracking technologies, studies have been conducted to find the relationship between eye fixations and cognitive processes since the 1970s. [74] has demonstrated that the locus and duration of eye fixations are all closely related to the activity of the central processor. Following this, extensive studies have proven that the duration of fixation has a negative relationship with the mental workload during various complex tasks [56, 75, 76]. In other words, higher fixation duration will be observed during lower mental workload condition and lower fixation duration will occur during higher mental workload tasks. Results in [76] also emphasized that the fixation duration is the most suitable metric among various parameters to estimate mental workload. Furthermore, in a study on visual attention of pilots, researchers have found that expert pilots will have more fixations on different instruments/displays with shorter duration time [56].

Eye saccades have become a popular metric for studying motor control, cognition and memory [77], however, a significant relationship between the amount of saccades and mental workload has not been observed [75, 76]. [76] conducted the N-back memory experiment with four difficulty levels to induce mental workload, in which 17 eye parameters were evaluated to investigate the relationship with mental workload. Results have demonstrated that the fixation duration and eye blink parameters show a significant relationship with workload, while the saccade related parameters failed to show a significant relationship. In [78], driver distraction was evaluated by detecting eye saccade movements. They have found that the older group shows worse performance with mental workload under the distracted driving conditions. [56] also suggested that expert pilots have more saccades on different instruments compared to novices.

Previous studies also found that blink frequency and blink duration showed a significant positive relationship to mental workload [61, 76]. [57] demonstrated that blink duration, compared to blink frequency, is a more sensitive and reliable indicator for workload detection. Similar to blinks, the percentage of eye closure over time can be used as a measure of fatigue, which has been extensively applied for driving fatigue detection [58].

1.4.3.4 Eye-tracking based Skill Assessment

Eye-tracking data can also be adopted as an objective tool for skill assessment [60], with potential applications in training for improving performance. Recent research has demonstrated that the eye movements have significant differences between novices and experts in aviation [56] and medicine [59]. [56] has shown that expert pilots tend to have more fixations on different instruments and shorter dwell time on each instrument. Meanwhile, expert pilots manage to extract more relevant information from their peripheral information. For surgeons, differences in eye metrics reflecting focused attention can also be found between junior and senior surgeons. Senior surgeons have higher fixation rates because they know what they are looking for and where to locate it, and simultaneously, they have lower ICA values over junior surgeons as they do not experience the same degree of cognitive workload in surgical procedures [59, 60].

Figure 9: University of North Dakota Space Analog Simulations (Photo provided by Dr Travis Nelson and Prof. Pablo De Leon.)

1.4.4 NeuroImaging in Space

Neurophysiological responses are altered in space due to several factors listed in Figure 4 that include physiological, psychological and habitability issues. Some of the physiological changes include cognitive/neurological alterations, increased fatigue, changes in circadian rhythm, changes in stress hormone levels and immune function [79]. Most of the variations in brain neurophysiology occur as a result of microgravity. Microgravity results in a shift of body fluid towards the head and this has been implicated in neurophysiological adaptations that last several weeks after the space flight [80]. In fact, both structural and functional brain changes have been demonstrated in studies [50, 51, 80, 81].

The brain adjusts to changes such as a shortened sleep cycle and microgravity, which has been reflected in changes in heart rate variability (HRV), changes in brain rhythms and regulatory brain connectivity networks, such as the default mode network (DMN) [80]. The DMN emerges during rest and its function manifests from interactions across several brain regions. It is thought to regulate the autonomic nervous system and its interaction with other major brain networks, such as the Salience Network, reflects shifts in focused attention [80, 82]. In space the DMN plays an important adaptation role that is also reflected in changes in the human oscillatory brain activity [50, 51]. Furthermore, the close relationship of the DMN to the autonomic system results in changes observed to the HRV.

It is important to realise that most of the scientific evidence for these changes does not come from studies conducted in space. Instead, some of the studies compare neuroimaging/physiological data obtained from astronauts before and after space flights. Other studies use simulation environments to imitate some of the common conditions encountered in space. The most notable conditions are confinement and isolation along with microgravity. Some of these simulation experiments take place in cages to resemble the extreme psychological and physical space conditions, such as isolation and confinement. For example, the Space Studies Department at North Dakota has a specialized facility to resemble 'Mars' missions, Figure 9, and to test new spacesuit technologies along with how they affect mental workload [83]. There are also human activities, such as Antarctic expeditions, that in some cases provide a close analog to space missions. On the other hand, microgravity effects can be simulated with bedrest approaches, in which the subject is asked to lie on a bed that is inclined downwards by roughly six degrees [79].

These experiments exert mental and physical pressure on individuals and they are difficult to complete. One major limitation is the small number of subjects and thus the inability to extract statistically significant results. This problem is more profound in early studies and it is exacerbated by the fact that experimental conditions between studies and exact protocols differ significantly from mission to mission. As the quality of portable neuroimaging equipment improves and becomes more practical and less cumbersome, BCI will be adapted to more realistic scenarios in space and aviation. In addition, space agents like NASA have already developed technology to reduce motion artifacts and improve accuracy of similar systems [84].

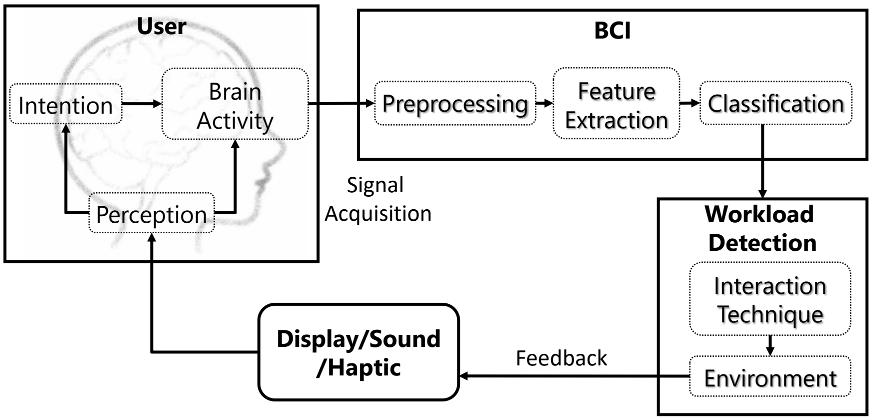


Figure 10: Classical BCI framework it involves signal acquisition followed by machine learning techniques to preprocess the signal, extract features and perform classification. Intuitive interactive approaches can increase the robustness of the BCI [3, 4].

1.5 Artificial Intelligence in BCI-based Workload Detection

Typically brain computer interfaces rely on signal preprocessing to remove artifacts, which depend on the nature of the signal, followed by feature extraction and classification Figure 10. There is large literature that has accumulated knowledge derived from well-controlled lab experiments on tasks that modulate workload, such as the n-back task mentioned earlier. However, there is far less work to address challenges in real environments [85]. This is partly because BCIs are sensitive to motion artefacts and interference. Even in the most well-controlled environments subject variability in BCIs hinders the robustness of the device and high classification rates are difficult to achieve for more than two classes [3]. Furthermore, reliable signal detection entails good coverage of key brain regions, which results in cumbersome equipment that are not pleasant to wear.

Currently, there are three notable categories in workload detection and modeling that drive research forward: i) work that extracts cognitive models derived from neuroimaging data that allow us to explain differences in the signal between workload conditions [86]; ii) Multi-modal fusion, which exploits machine learning algorithms to extract features from several modalities that include EEG, fNIRS, eye-tracking and physiological measures to improve classification rates [85]; and iii) BCIs that are coupled with realistic paradigms of tasks that account for the complexity of the task in hand [1].

Cognitive models of workload shed a light on the neurophysiological origins of the signals and can help eliminate spurious results due to motion and physiological related artefacts, such as breathing and heart rate. Growing evidence show that breathing patterns change under mental workload [86, 87]. This implies that occasionally BCIs classification rates may reflect motion artefacts and therefore do not generalize across subjects. Furthermore, this would hinder detection of peaks of performance with relation to optimum workload.

Information fusion of neurophysiological, physiological and behavioral data is important in real world experiments because they have the potential to enhance reliability and sensitivity of workload detection, while they reduce uncertainty and address inter-subject variability. Neurophysiological recordings refer to modalities that directly measures brain signals. For example, combining neurophysiological modalities such as EEG and fNIRS can enhance temporal and spatial coverage [88, 89].

On the other hand, physiological modalities normally include eye-tracking, heart-rate measurements and electrodermal activity (EDA) are indices of acute stress, Figure 4 and they are sensitive to mental states [90]. Normally, measurements can be obtained with far less sensors than classical BCIs and they are more discreet and comfortable to wear. However, they are also affected by physical activity and their ability to detect fine changes in mental workload is inherently limited when they are used alone.

Behavioral measures include computer mouse movements and clicks and the reason they are used in workload detection is the fact that they reflect engagement and attention. Human pose tracking is among the most useful behavioral measure in human machine interaction as well as mental state detection [85, 90, 91]. Body posture reveals attentional engagement and it also relates with the difficulty and complexity of the task in hand [91]. For example, with increased workload the person may lean forward and the distance from the monitor is smaller.

It is well known in signal-processing that information fusion from independent sensors improve signal-to-noise ratio and reduces bias of confounding factors. Multi-modal fusion is normally implemented in four levels: sensor level, feature level, decision level and hybrid models. The level indicates at which point information fusion takes place along the processing pipeline [85]. At sensor level fusion takes place after preprocessing and it is suited when the raw data reflect the same physical aspect. At feature level, fusion takes place after feature extraction and thus extracted features are estimated independently from each modality [88]. At decision level, feature extraction and classification has performed independently across modalities and fusion takes into consideration classification outcomes to reach a final decision.

1.6 Cognitive Workload Estimation during Simulated Teleoperations – A Case Study

To investigate neurophysiological indices of cognitive workload, eye-tracking and EEG data were simultaneously acquired from 10 healthy volunteers during teleoperations on the photo-realistic Unity simulator of the Canadarm at the ISS shown in Figure 2 [1, 92]. The aims of the study were to understand how exogeneous factors such as time-pressure and time-latency affect the performance of operators and how these changes link to neurophysiological indices related to spatial attention and cognitive workload. Under the time-pressure condition, users need to complete the task within a certain time (4 minutes), whereas in time-latency condition, 0-1.5sec delays were added to the operator motion control to reflect the round-trip time that it takes for the control signal to reach the robot and the visual feedback to travel back to the operator. Time-pressure is common in safety critical operations and the ability of a user to operate under limited time can have a profound influence in the success of the operation. The relationship between time-pressure and performance has been highlighted in section 1.3 and is supported by several studies in teleoperated robotics, such as surgical robots and other systems [86]. On the other hand, time-latency is very critical in space/teleoperation applications due to the considerable time it takes the signal to cover large distances in space. Communication delay can also reflect hardware, design or software limitations and it is a well know problem in master/slave robotic systems.

The simulation task has been picked to reflect a real-life scenario of the seven degree of freedom teleoperated Canadarm at the ISS and is described in detail in previous work [1, 92]. It requires the user to locate a new module close to the ISS, navigate the robotic arm to the grapple fixture of the module so the end-effector can attach to it and then move and dock the new module next to the Columbus module. As in real-life it is important that the operation is performed safely, so that the Canadarm and the module do not collide with the ISS or other objects. To ensure that the task is sufficiently difficult and it requires enhanced spatial abilities, debris

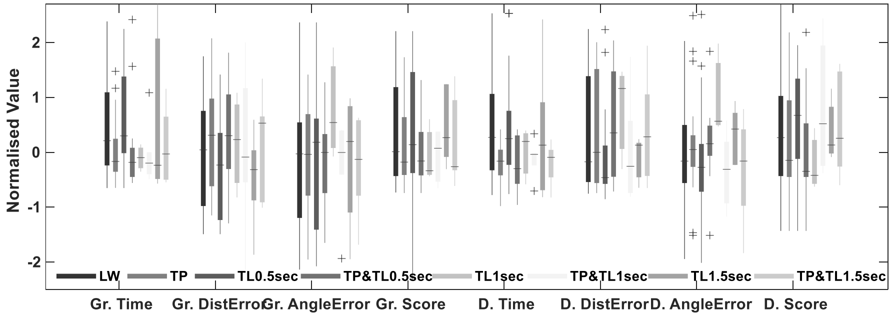


Figure 11: Normalised values of performance measures during teleoperations via VR simulation of the Canada arm at the ISS that include: the grasp time (Gr. Time), grasping distance error (Gr. Distance Error), grasping angular error (Gr. Angle Error), grasping score (Gr. Score), docking time (D. Time), docking distance error (D.DistError), docking angular error (D. AngleError) and docking overall score (D. Score). These performance measures are examined in conditions that induce varying cognitive workload: The 'Low Workload' condition refer to performing the task without additional time-pressure and time-latency (LW), under time-pressure only (TP), under time-latency of 0.5sec only (TL), under both time-pressure and time-latency of 0.5 sec (TP&TL0.5sec), under time-latency of 1 sec only (TL1sec), under both time-pressure and time-latency of 1sec (TP&TL1sec), under time-latency of 1.5sec (TL1.5sec) and under time-pressure and time-latency of 1.5sec (TP&TL1.5sec).

objects were added to the scene so that the user would have to find a way to navigate around them.

Users were asked to complete three blocks that included: i) a familiarisation block (data from this part was not included in the analysis), which gave them the opportunity to familiarise themselves with the simulator without any added time-pressure or time-latency. To start the main experiment, users were required to complete a simple version of the task without obstacles under four minutes. ii) A block of nine trials with randomised order that included three trials of time-pressure, three trials with a time-delay of 0.5 sec and three times with neither factors. Users in this block had the additional difficulty of navigating the robot around obstacles. iii) A latency block that included two trials for 0.5, 1.0 and 1.5 seconds of latency with and without time-pressure, a total of six trials. This block also required the users to navigate around obstacles.

Several performance measures were taken into consideration that included i) grasp time (time between the beginning of the experiment and the grasping of the module), ii) dock time (time from the grasping of the module to the docking to another part of the ISS), iii) grasp distance error (distance error from the optimum grasping point), iv) grasp angular error (angle error from the grasping angle between the module and the robotic end effector) , v) dock distance error (distance error from the optimum docking point) vi) dock angle error (angle error from the optimum docking angle between the module and the docking station), vii) Grasp score (score after grasping the module) viii) Dock score (score from grasping to docking the module), ix) Number of collisions with obstacles per trial. ANOVA results for pair-wise comparisons showed that the grasping and docking time between low workload and time-pressure conditions is statistically significant with a p-value smaller than 0.05 and 0.005, respectively. Furthermore, the number of collisions with obstacles per trial between time-latency condition and low workload condition was also

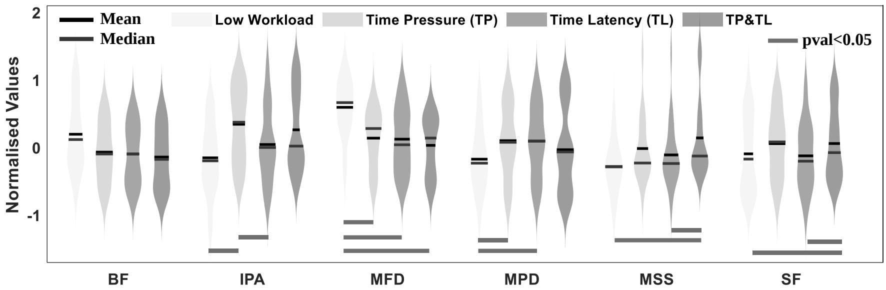


Figure 12: Normalised values of eye-tracking features with relation to different cognitive workload conditions. Blink Frequency (BF), Index of Pupillary Activity (IPA), Mean Fixation Duration (MFD), Mean Pupil Diameter (MPD), Mean Saccade Speed (MSS) and Saccade Frequency (SF).

statistically significantly with p-value smaller than 0.05. No significant differences were found to the precision and overall scores of the tasks.

On the other hand, eye-tracking features reveal a number of statistically significant differences. The most prominent features include saccade frequency, mean saccade speed, mean pupil diameter, mean fixation duration and the Index of Pupillary Activity (IPA) [92]. Pupillary response, such as pupil diameter has been related to cognitive workload factors. However, it is also sensitive to other factors including the brightness of the scene. The Index of Pupillary Activity (IPA), which reflect small variations in pupil diameter has been suggested as a more robust measure of cognitive workload. Mean frequency duration, mean pupillary diameter and IPA reveal statistically significant differences between the low-workload and time-pressure conditions. Significant differences between time-pressure and time-latency conditions are identified based on IPA only, whereas significant differences between low-workload and time-latency conditions were identified based on mean frequency duration and mean pupillary diameter. Finally, mean saccade speed and saccade frequency were more sensitive to time-latency conditions.

It is observed that in this scenario eye-tracking based features are more sensitive to different experimental conditions than performance measures. Furthermore, performance measures need to be redefined for each different task. For example, in this study we broke down the task into subparts of the grasping and the docking phase. This approach is ad-hoc, requires specific expert knowledge of the task at hand and thus performance measures are not generalizable to new situations. Another important limitation is that only few simplistic performance measures, such as time duration, can be measured outside the simulation environment. Therefore, there are not suitable for closed-loop feedback mechanisms that could help the operator to improve his/her performance and alert the team if unexpected situations emerge. This indicates how important is to acquire neurophysiological indices during teleoperations and in this way facilitate the development of dynamic human-in-the-loop systems.

Further analysis based on two-class classification results of with/without Time-Pressure and with/without Time-Latency, respectively, is demonstrated in Figure 13. The figure shows the results of SVM based on radial kernel functions across different time windows of 2secs, 5secs, 10secs, 20secs and the whole trial data [92]. Classification accuracy is shown for a number of eye-tracking features that include

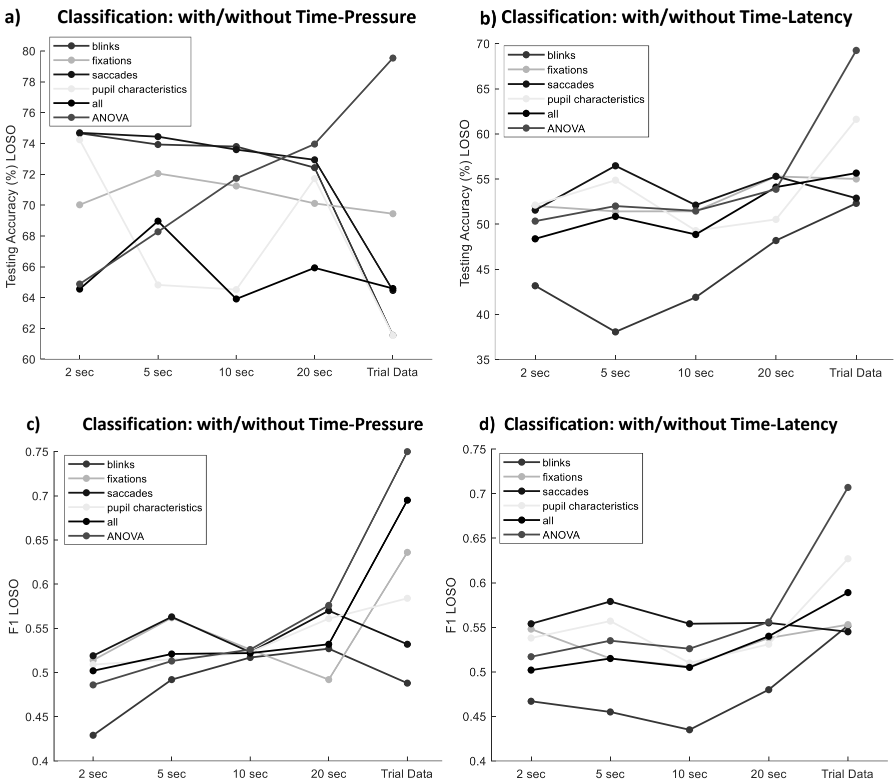


Figure 13: Two-class classification results on with or without time-latency and with and without time-latency under the Leave-One-Subject-Out (LOSO) cross-validation protocol. Testing accuracy and F1 scores are shown for window sizes of 2sec, 5sec, 10sec, 20sec and the whole trial data for each classifier respectively. The classifier is based on SVM with radial basis kernel functions. A number of eye-tracking features are examined to determine whether they can reliably identify cognitive workload.

blinks, fixations, saccades, pupil characteristics, all features and ANOVA features with a significant difference extracted from each trial. Blinks and eye saccades features performed well in identifying time-pressure even with very small window size of 2secs. These eye-tracking characteristics could enable real-time evaluation of cognitive workload related to time-pressure. Pupil characteristics show less stable performance, whereas concatenating all features does not increase performance. Perhaps, this reflects that there are small number of samples compared to features. On the other hand, ANOVA features perform best in terms of classification in both with/without time-pressure and with/without time-latency. They work by evaluating the variance of the predictive variables on the response. However, their performance is low with small window sizes and reaches its maximum with the whole-trial data. Classification of with/without Time-pressure performs significantly better than classification of with/without Time-Latency condition. Further validation is required to understand whether this difference in performance reflects uneven sample sizes or

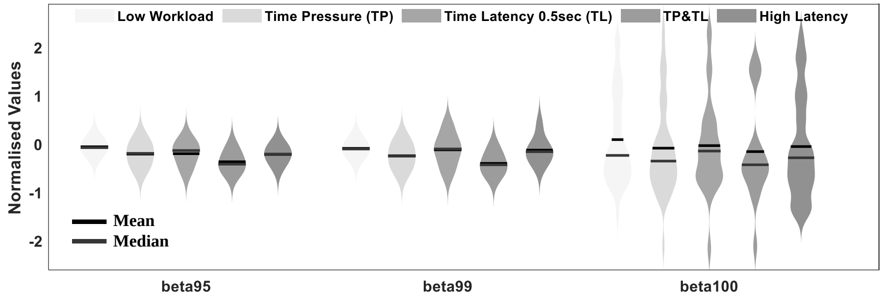


Figure 14:Beta power as a neurophysiological index of cognitive workload. EEG data are z-normalised. A large variability is observed and for these reasons outliers have been excluded by estimating confidence intervals at 95% (beta95) and 99% (beta99), respectively.

whether time-latency is inherently more difficult to identify based on eye-tracking features.

EEG data were also acquired simultaneously and analyzed to extract features and neurophysiological indices that are related to differences in cognitive workload [1]. Previous work hypothesizes that specific spectral powers, such as theta, alpha and beta as well as derived indices such as ratios of spectral power in these bands can robustly identify workload [93]. It is common for EEG data to be band-passed into five bands, which roughly categorized as delta (1-4Hz), theta (4-8Hz), alpha (8-13Hz), beta (13-30Hz) and gamma (30-70Hz) [94]. These definitions relate to fundamental properties of human brain function that reflect distinct roles and underline communication between different brain regions. The theta band has been correlated with mental fatigue, higher demands of working memory and increased cognitive workload. It has been also linked to tasks that require sustained attention and it is anti-correlated with alertness and mental vigilance. On the other hand, occipital alpha oscillations are very well recognized during relaxed states and eyes closed. Alpha power is also anticorrelated with vigilance and attentional resources, whereas a suppression of alpha waves in occipital and parietal regions relates to more challenging and difficult tasks [93]. Beta power is also linked to cognitive workload as it correlates well with increased workload intensity and high levels of concentration and visual attention.

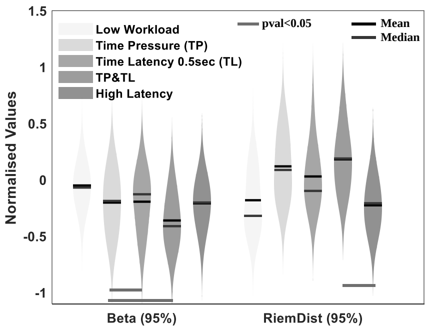


Figure 15:Beta power and Riemannian distance as measures of cognitive workload. Statistically significant differences are identified based on Kruskal-Wallis pair-wise comparisons.

EEG data are prone to artifacts, which result from muscle activity, eye blinks and movements. Independent component analysis (ICA) was used to decompose the EEG signal in statistically independent sources and reject the sources that are unlikely to have neuronal origin. Since beta power has been demonstrated as a strong predictor of cognitive workload, the EEG signal was band-passed in beta band and the beta power was compared across different conditions. EEG data were z-normalised and segmented in 2sec windows. Figure 14 uses violin plots to provide a sense of the data distribution and is shows a high variability and non-Gaussian distribution when all data are included (beta100). To alleviate this problem outliers are removed by retaining the 99% (beta99) and 95% (beta95) of data, respectively, within the confidence interval of each class.

Finally, Figure 15 shows how changes across conditions in beta power compare with changes in Riemannian distance between covariance matrices. Covariance matrices reveal brain connectivity and their geometric properties can provide a better estimation of distance compared to standard Euclidean distance [82, 95]. Riemannian distance has been used also in brain computer interfaces as a robust way to classify different brain states [2]. Here we hypothesize that cognitive workload increases will be also associated with changes in brain connectivity. This is an argument supported also from previous work in cognitive workload that revealed significant connectivity changes both across condition as well as across expertise in surgeons [86]. Statistically significant differences were identified based on the Kruskal-Wallis one-way analysis of variance, which is a non-parametric method that account for the fact that some of the EEG data distributions are not normal even after the z-normalisation and outliers removal.

1.7 Recommendations and Future Work

Realistic virtual reality environments along with augmented reality paradigms provide unique opportunities towards novel artificial intelligence approaches that improve performance in teleoperations by leveraging human factors information obtained via multi-modal neurophysiological indices. Human in the loop systems are challenging to design and require multi-disciplinary approaches that couple our understanding in human brain neurophysiology with powerful computational approaches. Future applications should exploit the rich information derived from numerous neuroscientific and psychology studies to develop real-time adaptive systems. These systems should be able to account for the complexity in real-life scenarios via data-driven approaches and provide subject-specific support. Therefore, future systems should focus on:

- The communication between key brain regions such as the prefrontal cortex and motor cortex is relatively unexplored. The function of the prefrontal cortex is key in decision making and as a high functional centre it is affected by anxiety and cognitive workload. Evidence shows that there are statistically significant differences in brain connectivity between conditions of time-pressure and self-paced tasks as well as between novice and expert users. The development of advanced machine learning techniques that exploit this information in real-time and provide continuous measures of workload is of paramount importance. The Riemannian distance of the covariance matrix of the EEG signal has shown some promise in characterising cognitive workload differences. Further work is required to develop intuitive and interpretable continuous measures that can detect subtle differences in brain connectivity and characterise cognitive workload conditions.

- Cognitive workload estimation exploit advances in brain computer interfaces both in terms of miniaturised, portable sensing devices as well as intelligent algorithms. Current systems do not generalise well across subjects due to high inter-subject and inter-session variability and require training to be adjusted to each new user. Typically, BCIs require subject-specific training and calibration prior to use. Nevertheless, there are significant advances in machine learning that pave the way towards subject-independent BCIs. These technologies will play a significant role in transferring knowledge from highly specialised neuroimaging technologies to wearable headsets thus enabling reliable systems to be developed for safety critical applications.
- It is also important to understand the differences in cognitive workload induced due to task specific difficulty and exogeneous factors such as time-pressure and time-latency. To this end several neurophysiological indices were explored as well as activity-depended measures that shed some light on the interaction level between the user and the system. Simulations environments can also provide opportunities to couple user's actions with robotic motion in a task-independent way. These causal relationships give intuition behind performance and errors, which should be predicted and avoided in space applications.
- Another important aspect of teleoperated systems with a human in the loop is semi-autonomous modes of operations. Semi-autonomy can enhance performance and reduce errors while it empowers users to be in control of the system. How these modes of operations can be defined in a principled way that it is independent of the task in hand and therefore their function generalise across robotic systems is an active area of robotics. Furthermore, how a system can go safely from one level of autonomy to another is of paramount importance in safety critical applications in space. In this scenario, transparency of the system achieved by intuitive, interactive designs and continuous measures of cognitive workload will play a critical role both during the cycles of system development as well as during the operation.
- Computational complexity can be prohibitive in translating current artificial intelligence success stories to space applications. Therefore, it is important to develop emulation systems that mimic the hardware capabilities in space and allow researchers to test their solutions.

References

[1] Freer, D., Guo, Y., Deligianni, F., and Yang, G.-Z., 'On-Orbit Operations Simulator for Workload Measurement During Telerobotic Training', *ACM Transactions on Human Robot Interaction (https://arxiv.org/abs/2002.10594)*, submitted.

[2] Freer, D., Deligianni, F., and Yang, G.Z., 'Adaptive Riemannian Bci for Enhanced Motor Imagery Training Protocols', *IEEE BSN*, 2019.

[3] Williamson, J., Quek, M., Popescu, I., Ramsay, A., and Murray-Smith, R., 'Efficient Human-Machine Control with Asymmetric Marginal Reliability Input

Devices', *Plos One*, 2020.
[4] Evain, A., Roussel, N., Casiez, G., Argelaguet‐Sanz, F., and Lecuyer, A., Brain‐Computer Interfaces for Human‐Computer Interaction', *Brain‐Computer Interfaces 1: Foundations and Methods*, (John Wiley & Sons, Inc., 2016)
[5] Pieters, M.A. and Zaal, P.M.T., 'Training for Long-Duration Space Missions: A Literature Review into Skill Retention and Generalizability', *Ifac Papersonline*, 2019, 52, (19), pp. 247-252.
[6] Clark, T.K., Stimpson, A.J., Young, L.R., Oman, C.M., Natapoff, A., and Duda, K.R., 'Human Spatial Orientation Perception During Simulated Lunar Landing Motions', *Journal of Spacecraft and Rockets*, 2014, 51, (1), pp. 267-280.
[7] https://www.theverge.com/2019/3/15/18267365/boeing-737-max-8-crash-autopilot-automation, accessed Date Accessed 2019 Accessed
[8] Weller, A., Transparency: Motivations and Challenges', in Samek, W., Montavon, G., Vedaldi, A., Hansen, L.K., and Müller, K.-R. (eds.), *Explainable Ai: Interpreting, Explaining and Visualizing Deep Learning*, (2019)
[9] Murray-Smith, R., Control Theory, Dynamics, and Continuous Interaction', in Oulasvirta, A., Kristensson, P.O., Bi, X., and Howes, A. (eds.), *Computational Interaction*, (Oxford University Press, 2018)
[10] Kramer, L.J., Etherington, T.J., Bailey, R.E., and Kennedy, K.D., 'Quantifying Pilot Contribution to Flight Safety During Hydraulic Systems Failure', *Advances in Human Aspects of Transportation*, 2018, 597, pp. 15-26.
[11] Coleshill, E., Oshinowo, L., Rembala, R., Bina, B., Rey, D., and Sindelar, S., 'Dextre: Improving Maintenance Operations on the International Space Station', *Acta Astronautica*, 2009, 64, (9-10), pp. 869-874.
[12] Visinsky, M., 'Robotics on the International Space Station (Iss) and Lessons in Progress', *NASA Techical Reports Server*, 2017.
[13] Fong, T., Rochlis Zumbado, J., Currie, N., Mishkin, A., and Akin, D.L., 'Space Telerobotics:Unique Challenges to Human–Robot Collaboration in Space', *Reviews of Human Factors and Ergonomics*, 2013, 9, (1), pp. 6-56.
[14] Callen, P., 'Robotic Transfer and Interfaces for External Iss Payloads', *NASA Techical Reports Server*, 2014.
[15] https://www.nasa.gov/mission_pages/station/structure/elements/special-purpose-dextrous-manipulator, accessed Date Accessed 2018 Accessed
[16] https://www.asc-csa.gc.ca/eng/iss/robotics/robotics-training-centre.asp, accessed Date Accessed 2018 Accessed
[17] Belghith, K., Nkambou, R., Kabanza, F., and Hartman, L., 'An Intelligent Simulator for Telerobotics Training', *Ieee Transactions on Learning Technologies*, 2012, 5, (1), pp. 11-19.
[18] Fournier-Viger, P., Nkambou, R., and Mayers, A., 'Evaluating Spatial Representations and Skills in a Simulator-Based Tutoring System', *Ieee Transactions on Learning Technologies*, 2008, 1, (1), pp. 63-74.
[19] https://www.asc-csa.gc.ca/eng/multimedia/games/canadarm2/default.asp, accessed Date Accessed 2013 Accessed
[20] Stirling, L., Siu, H.C., Jones, E., and Duda, K., 'Human Factors Considerations for Enabling Functional Use of Exosystems in Operational Environments', *Ieee Systems Journal*, 2019, 13, (1), pp. 1072-1083.
[21] Ravi, D., Wong, C., Deligianni, F., Berthelot, M., Andreu-Perez, J., Lo, B., and Yang, G.Z., 'Deep Learning for Health Informatics', *IEEE JBHI*, 2017, 21, (1),

pp. 4-21.
[22] Andreu-Perez, J., Deligianni, F., Ravi, D., and Yang, G.Z., 'Artificial Intelligence and Robotics', *arxiv.org/abs/1803.10813*, 2018.
[23] Cinel, C., Valeriani, D., and Poli, R., 'Neurotechnologies for Human Cognitive Augmentation: Current State of the Art and Future Prospects', *Frontiers in Human Neuroscience*, 2019, 13.
[24] Navarro, J., Reynaud, E., and Osiurak, F., 'Neuroergonomics of Car Driving: A Critical Meta-Analysis of Neuroimaging Data on the Human Brain Behind the Wheel', *Neuroscience & Biobehavioral Reviews*, 2018, 95, pp. 464 - 479.
[25] Modi, H.N., Yang, H.S.G.-Z., Darzi, A., and Leff, D.R., 'A Decade of Imaging Surgeons' Brain Function (Part Ii): A Systematic Review of Applications for Technical and Nontechnical Skills Assessment', *Surgery*, 2017, 162, (5), pp. 1130 - 1139.
[26] Clery-Melin, M.L., Jollant, F., and Gorwood, P., 'Reward Systems and Cognitions in Major Depressive Disorder', *Cns Spectrums*, 2019, 24, (1), pp. 64-77.
[27] Hiser, J. and Koenigs, M., 'The Multifaceted Role of the Ventromedial Prefrontal Cortex in Emotion, Decision Making, Social Cognition, and Psychopathology', *Biological Psychiatry*, 2018, 83, (8), pp. 638-647.
[28] Xu, J., Van Dam, N.T., Feng, C.L., Luo, Y.J., Ai, H., Gu, R.L., and Xu, P.F., 'Anxious Brain Networks: A Coordinate-Based Activation Likelihood Estimation Meta-Analysis of Resting-State Functional Connectivity Studies in Anxiety', *Neuroscience and Biobehavioral Reviews*, 2019, 96, pp. 21-30.
[29] Knudsen, E.I., 'Fundamental Components of Attention', *Annual Review of Neuroscience*, 2007, 30, pp. 57-78.
[30] Knudsen, E.I., 'Neural Circuits That Mediate Selective Attention: A Comparative Perspective', *Trends in Neurosciences*, 2018, 41, (11), pp. 789-805.
[31] Moore, T. and Zirnsak, M., 'Neural Mechanisms of Selective Visual Attention', *Annual Review of Psychology, Vol 68*, 2017, 68, pp. 47-72.
[32] Clayton, M.S., Yeung, N., and Kadosh, R.C., 'The Roles of Cortical Oscillations in Sustained Attention', *Trends in Cognitive Sciences*, 2015, 19, (4), pp. 188 - 195.
[33] Gilbert, C.D. and Li, W., 'Top-Down Influences on Visual Processing', *Nature Reviews Neuroscience*, 2013, 14, (5), pp. 350-363.
[34] Ruff, D.A., Ni, A.M., and Cohen, M.R., 'Cognition as a Window into Neuronal Population Space', *Annual Review of Neuroscience*, 2018, 41, p. 77.
[35] Walle, K.M., Nordvik, J.E., Espeseth, T., Becker, F., and Laeng, B., 'Multiple Object Tracking and Pupillometry Reveal Deficits in Both Selective and Intensive Attention in Unilateral Spatial Neglect', *Journal of Clinical and Experimental Neuropsychology*, 2019, 41, (3), pp. 270-289.
[36] Husain, M., 'Visual Attention: What Inattention Reveals About the Brain', *Current Biology*, 2019, 29, (7), pp. R262-R264.
[37] Bartolomeo, P., Thiebaut de Schotten, M., and Chica, A.B., 'Brain Networks of Visuospatial Attention and Their Disruption in Visual Neglect', *Frontiers in Human Neuroscience*, 2012, 6.
[38] Bogadhi, A.R., Bollimunta, A., Leopold, D.A., and Krauzlis, R.J., 'Spatial Attention Deficits Are Causally Linked to an Area in Macaque Temporal Cortex', *Current Biology*, 2019, 29, (5), p. 726.
[39] Kanki, B., Clervoy, J.-F., and Sandal, G., *Space Safety and Human*

Performance, (Butterworth-Heinemann, 2017)
[40] Duncan, J., Martens, S., and Ward, R., 'Restricted Attentional Capacity within but Not between Sensory Modalities', *Nature*, 1997, 387, (6635), pp. 808-810.
[41] Boucart, M., Fabre-Thorpe, M., Thorpe, S., Arndt, C., and Hache, J.C., 'Covert Object Recognition at Large Visual Eccentricity', *Journal of Vision*, 2001, 1, (471).
[42] Potter, M.C., Wyble, B., Hagmann, C.E., and McCourt, E.S., 'Detecting Meaning in Rsvp at 13 Ms Per Picture', *Attention Perception & Psychophysics*, 2014, 76, (2), pp. 270-279.
[43] Jamison, H. and Julie, A.A., 'Multi-Dimensional Human Workload Assessment for Supervisory Human–Machine Teams', *Journal of Cognitive Engineering and Decision Making*, 2019, 13, (3), pp. 146-170.
[44] Horrey, W.J., Lesch, M.F., Garabet, A., Simmons, L., and Maikala, R., 'Distraction and Task Engagement: How Interesting and Boring Information Impact Driving Performance and Subjective and Physiological Responses', *Applied Ergonomics*, 2017, 58, pp. 342 - 348.
[45] Varghese, R.J., Freer, D., Deligianni, F., Liu, J., and Yang, G.Z., Wearable Robotics for Upper-Limb Rehabilitation and Assistance: A Review on the State-of-the-Art, Challenges and Future Research', *Wearable Technology in Medicine and Healthcare*, (Elsevier, 2018)
[46] Velliste, M., Perel, S., Spalding, M.C., Whitford, A.S., and Schwartz, A.B., 'Cortical Control of a Prosthetic Arm for Self-Feeding', *Nature*, 2008, 453, (7198), pp. 1098-1101.
[47] Berger, H., 'The Electro-Cephalogramm of Humans and Its Analysis.', *Naturwissenschaften*, 1937, 25, pp. 193-196.
[48] Jiang, X., Bian, G.B., and Tian, Z.A., 'Removal of Artifacts from Eeg Signals: A Review', *Sensors*, 2019, 19, (5).
[49] Brouwer, A.M., Hogervorst, M.A., van Erp, J.B.F., Heffelaar, T., Zimmerman, P.H., and Oostenveld, R., 'Estimating Workload Using Eeg Spectral Power and Erps in the N-Back Task', *Journal of Neural Engineering*, 2012, 9, (4).
[50] Cheron, G., Cebolla, A.M., Petieau, M., Bengoetxea, A., Palmero-Soler, E., Leroy, A., and Dan, B., 'Adaptive Changes of Rhythmic Eeg Oscillations in Space: Implications for Brain-Machine Interface Applications', *Brain Machine Interfaces for Space Applications: Enhancing Astronaut Capabilities*, 2009, 86, pp. 171-187.
[51] Leroy, A., De Saedeleer, C., Bengoetxea, A., Cebolla, A., Leurs, F., Dan, B., Berthoz, A., McIntyre, J., and Cheron, G., 'Mu and Alpha Eeg Rhythms During the Arrest Reaction in Microgravity', *Microgravity Science and Technology*, 2007, 19, (5-6), pp. 102-107.
[52] Berka, C., Levendowski, D.J., Lumicao, M.N., Yau, A., Davis, G., Zivkovic, V.T., Olmstead, R.E., Tremoulet, P.D., and Craven, P.L., 'Eeg Correlates of Task Engagement and Mental Workload in Vigilance, Learning, and Memory Tasks', *Aviation Space and Environmental Medicine*, 2007, 78, (5), pp. B231-B244.
[53] Murata, A., 'An Attempt to Evaluate Mental Workload Using Wavelet Transform of Eeg', *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 2005.
[54] Curtin, A. and Ayaz, H., 'The Age of Neuroergonomics: Towards Ubiquitous and Continuous Measurement of Brain Function with Fnirs', *Japanese*

Psychological Research, 2018, 60(4).
[55] Liu, Y., Ayaz, H., and Patricia, A.S., 'Mental Workload Classification with Concurrent Electroencephalography and Functional near-Infrared Spectroscopy', *Brain-Computer Interfaces*, 4, (3), pp. 175-185.
[56] Ziv, G., 'Gaze Behavior and Visual Attention: A Review of Eye Tracking Studies in Aviation', *The International Journal of Aviation Psychology*, 2016, 26, (3-4), pp. 75-104.
[57] Benedetto, S., Pedrotti, M., Minin, L., Baccino, T., Re, A., and Montanari, R., 'Driver Workload and Eye Blink Duration', *Transportation research part F: traffic psychology and behaviour*, 2011, 14, (3), pp. 199-208.
[58] Gao, X.-Y., Zhang, Y.-F., Zheng, W.-L., and Lu, B.-L., 'Evaluating Driving Fatigue Detection Algorithms Using Eye Tracking Glasses', in, *2015 7th International IEEE/EMBS Conference on Neural Engineering (NER)*, (IEEE, 2015)
[59] Richstone, L., Schwartz, M.J., Seideman, C., Cadeddu, J., Marshall, S., and Kavoussi, L.R., 'Eye Metrics as an Objective Assessment of Surgical Skill', *Annals of surgery*, 2010, 252, (1), pp. 177-182.
[60] Tien, T., Pucher, P.H., Sodergren, M.H., Sriskandarajah, K., Yang, G.-Z., and Darzi, A., 'Eye Tracking for Skills Assessment and Training: A Systematic Review', *journal of surgical research*, 2014, 191, (1), pp. 169-178.
[61] Van Orden, K.F., Limbert, W., Makeig, S., and Jung, T.-P., 'Eye Activity Correlates of Workload During a Visuospatial Memory Task', *Human factors*, 2001, 43, (1), pp. 111-121.
[62] Crawford, J., Martinez-Trujillo, J., and Klier, E., 'Neural Control of Three-Dimensional Eye and Head Movements', *Current opinion in neurobiology*, 2003, 13, (6), pp. 655-662.
[63] Clément, G. and Ngo-Anh, J.T., 'Space Physiology Ii: Adaptation of the Central Nervous System to Space Flight—Past, Current, and Future Studies', *European journal of applied physiology*, 2013, 113, (7), pp. 1655-1672.
[64] Clarke, A.H. and Haslwanter, T., 'The Orientation of Listing’s Plane in Microgravity', *Vision research*, 2007, 47, (25), pp. 3132-3140.
[65] Kowler, E., 'Eye Movements: The Past 25 Years', *Vision research*, 2011, 51, (13), pp. 1457-1483.
[66] Salvucci, D.D. and Goldberg, J.H., 'Identifying Fixations and Saccades in Eye-Tracking Protocols', in, *Proceedings of the 2000 symposium on Eye tracking research & applications*, (2000)
[67] Shelhamer, M. and Roberts, D.C., 'Magnetic Scleral Search Coil', *Handbook of Clinical Neurophysiology*, 2010, 9, pp. 80-87.
[68] Haslwanter, T. and Clarke, A.H., 'Eye Movement Measurement: Electro-Oculography and Video-Oculography', *Handbook of Clinical Neurophysiology*, 2010, 9, pp. 61-79.
[69] Nyström, M., Andersson, R., Holmqvist, K., and Van De Weijer, J., 'The Influence of Calibration Method and Eye Physiology on Eyetracking Data Quality', *Behavior research methods*, 2013, 45, (1), pp. 272-288.
[70] Kahneman, D. and Beatty, J., 'Pupil Diameter and Load on Memory', *Science*, 1966, 154, (3756), pp. 1583-1585.
[71] Hess, E.H. and Polt, J.M., 'Pupil Size in Relation to Mental Activity During Simple Problem-Solving', *Science*, 1964, 143, (3611), pp. 1190-1192.
[72] Wang, W., Li, Z., Wang, Y., and Chen, F., 'Indexing Cognitive Workload

Based on Pupillary Response under Luminance and Emotional Changes', in, *Proceedings of the 2013 international conference on Intelligent user interfaces*, (2013)
[73] Marshall, S.P., 'The Index of Cognitive Activity: Measuring Cognitive Workload', in, *Proceedings of the IEEE 7th conference on Human Factors and Power Plants*, (IEEE, 2002)
[74] Just, M.A. and Carpenter, P.A., 'Eye Fixations and Cognitive Processes', 1975.
[75] Schulz, C.M., Schneider, E., Fritz, L., Vockeroth, J., Hapfelmeier, A., Wasmaier, M., Kochs, E., and Schneider, G., 'Eye Tracking for Assessment of Workload: A Pilot Study in an Anaesthesia Simulator Environment', *British journal of anaesthesia*, 2011, 106, (1), pp. 44-50.
[76] Volden, F., Edirisinghe, V.D.A., and Fostervold, K.-I., 'Human Gaze-Parameters as an Indicator of Mental Workload', in, *Congress of the International Ergonomics Association*, (Springer, 2018)
[77] Leigh, R.J. and Kennard, C., 'Using Saccades as a Research Tool in the Clinical Neurosciences', *Brain*, 2004, 127, (3), pp. 460-477.
[78] Son, L.A., Hamada, H., Inagami, M., Suzuki, T., and Aoki, H., 'Effect of Mental Workload and Aging on Driver Distraction Based on the Involuntary Eye Movement', *Advances in Human Aspects of Transportation*, 2017, 484, pp. 349-359.
[79] Mogilever, N.B., Zuccarelli, L., Burles, F., Iaria, G., Strapazzon, G., Bessone, L., and Coffey, E.B.J., 'Expedition Cognition: A Review and Prospective of Subterranean Neuroscience with Spaceflight Applications', *Front Hum Neurosci*, 2018, 12, p. 407.
[80] Otsuka, K., Cornelissen, G., Kubo, Y., Shibata, K., Hayashi, M., Mizuno, K., Ohshima, H., Furukawa, S., and Mukai, C., 'Circadian Challenge of Astronauts' Unconscious Mind Adapting to Microgravity in Space, Estimated by Heart Rate Variability', *Sci Rep*, 2018, 8, (1), p. 10381.
[81] Schmitz, J., 'Structural Changes of the Brain During a Stay in Space Increased Intracranial Pressure for the Astronauts', *Flugmedizin Tropenmedizin Reisemedizin*, 2018, 25, (2), pp. 53-54.
[82] Deligianni, F., Varoquaux, G., Thirion, B., Sharp, D.J., Ledig, C., Leech, R., and Rueckert, D., 'A Framework for Inter-Subject Prediction of Functional Connectivity from Structural Networks', *Ieee Transactions on Medical Imaging*, 2013, 32, (12), pp. 2200-2214.
[83] Rabbi, A.F., Zony, A., Leon, P.d., and Fazel-Rezai, R., 'Mental Workload and Task Engagement Evaluation Based on Changes in Electroencephalogram', *Biomedical Engineering Letters*, 2012, 2, pp. 139–146.
[84] Harrivel, A. and Hearn, T., Detection of Mental State and Reduction of Artifacts Using Functional near Infrared Spectroscopy (Fnirs)', (NASA, 2014)
[85] Debie, E., Rojas, R.F., Justin Fidock, Michael Barlow, Kathryn Kasmarik, Sreenatha Anavatti , Matthew Garratt, and Abbass, H.A., 'Multimodal Fusion for Objective Assessment of Cognitive Workload: A Review', *IEEE Transactions on Cybernetics*, 2020.
[86] Deligianni, F., Singh, H., Modi, H.N., Jahani, S., Yucel, M., Darzi, A., Leff, D.R., and Yang, G.Z., 'Expertise and Task Pressure in Fnirs-Based Brain Connectomes', *https://arxiv.org/abs/2001.00114*, 2020.
[87] Jaiswal, D., Chowdhury, A., Banerjee, T., and Chatterjee, D., 'Effect of

Mental Workload on Breathing Pattern and Heart Rate for a Working Memory Task: A Pilot Study', *Conference Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2019, 2019, pp. 2202-2206.
[88] Aghajani, H., Garbey, M., and Omurtag, A., 'Measuring Mental Workload with Eeg+Fnirs', *Frontiers in Human Neuroscience*, 2017, 11, p. 359.
[89] Shin, J., von Luhmann, A., Kim, D.W., Mehnert, J., Hwang, H.J., and Muller, K.R., 'Data Descriptor: Simultaneous Acquisition of Eeg and Nirs During Cognitive Tasks for an Open Access Dataset', *Scientific Data*, 2018, 5.
[90] Deligianni, F., Guo, Y., and Yang, G.Z., 'From Emotions to Mood Disorders: A Survey on Gait Analysis Methodology', *Ieee Journal of Biomedical and Health Informatics*, 2019, 23, (6), pp. 2302-2316.
[91] Qiu, J. and Helbig, R., 'Body Posture as an Indicator of Workload in Mental Work', *Human Factors*, 2012, 54, (4), pp. 626-635.
[92] Guo, Y., Freer, D., Deligianni, F., and Yang, G.-Z., 'Eye-Tracking for Performance Evaluation and Workload Estimation in Space Telerobotic Training', *IEEE Transactions on Human-Machine Systems*, submitted.
[93] Rojas, R.F., Debie, E., Fidock, J., Barlow, M., Kasmarik, K., Anavatti, S., Garratt, M., and Abbass, H., 'Electroencephalographic Workload Indicators During Teleoperation of an Unmanned Aerial Vehicle Shepherding a Swarm of Unmanned Ground Vehicles in Contested Environments', *Frontiers in Neuroscience*, 2020, 14.
[94] Deligianni, F., Centeno, M., Carmichael, D.W., and Clayden, J.D., 'Relating Resting-State Fmri and Eeg Whole-Brain Connectomes across Frequency Bands', *Frontiers in Neuroscience*, 2014, 8.
[95] Deligianni, F., Varoquaux, G., Thirion, B., Robinson, E., Sharp, D.J., Edwards, A.D., and Rueckert, D., 'A Probabilistic Framework to Infer Brain Functional Connectivity from Anatomical Connections', *IPMI*, 2011, 6801, pp. 296-307.