COMP3050 Individual Dissertation Interim Report

Project name:Non-invasive Operator Monitoring

Submitted December 10, 2020, in partial fulfillment of the conditions for the award of the degree **Computer Science with Artificial Intelligence.**

Student Name: TANG Huimin
Student ID:20030248

Supervised by Matthew Pike

School of Computer Science University of Nottingham Ningbo China

Contents

1 Introduction

1.1 Background

As one of the greatest inventions of the 20^{th} century, the airplane has been playing an increasingly important role in people's travel and freight transportation since its inception. Because of its advantages of convenience and speed, the number of people choosing to travel and transport goods by air is increasing year by year. The data provided by the International Air Transport Association (IATA) [1] shows that air cargo demand grew 3.4% in 2019 compared with 2018, while the number of passengers choosing travels by airplane grew 3.72%. Increased reliance on the airplane makes the accuracy and safety of pilot's operations more important.

Except the increase in transportation demand, according to the data provided by General Aviation Aircraft Shipment Report [2], business jet deliveries have been growing for three consecutive years with an increase of 15% in 2019 compared with 2018, which means an increasing number of airplanes are being put into service. In other words, the future of air traffic will become more congested, pilots will face to complex flight environments and a growing number of parallel tasks, which means pilots will require more information regarding routes and traffic dynamics which creates the need for more complicated systems.

1.2 Motivation

There are a vast number of operations for current pilots that need to be done correctly if they want to safely complete their flights. The demand for concentration of pilots and the capacity to deal with numerous tasks will continue to increase because of the improvement of operational requirements and the increasingly sophisticated system, which means the number of attributes of the pilots' cognitive demand will increase. Cognitive demand represents the level of memory and attention required to process a specific task and it is related to the complexity of the task [3] which can be explained as the result of some information processing demands given by the task forcedly [4], these information processing demands imposed by the task are the attributes of the cognitive demands. If the task has higher complexity, which means it has larger number of required attributes, the cognitive demand will increase relatively [3]. Additionally, according to the data form Wilson [5], when the Mental Workload (MWL) is minimum, it corresponds to the lowest cognitive demand period and the MWL will increase as the cognitive demand increases. In conclusion, if the cognitive demand from a flight task containing too many attributes are presented to a pilot, it will lead to an increase in his MWL leading potentially to a state of mental overload [6]. Based on the theory that anticipated cognitive demand plays an important role in behavioural decision-making [7], when pilots are under a high MWL, their abilities will decline, especially the ability to deal with more than two chunks of auditory data will deteriorate rapidly [6], which means the judgement of the pilots will be influenced and might make some unwise decisions resulting to accidents. Increased risk of insecurity is neither good for the passengers nor the airline company. Moreover, excessive MWL will have a bad impact on pilot's physical and mental health [8], it is not wish to happen. Therefore, it is imperative to keep pilots' MWL at a normal level when they pilot a airplane.

During the flight, there will be many factors influence the MWL of pilots, such as the environment of the flight deck and the stages of flight. This project will focus on the factor related to the design of the aircraft cockpit display interfaces (CDI). When the pilot is piloting a airplane, the bridge that connects the pilot and the airplane is the CDI. During the flight, dynamic vision, which means the pilot needs take a quick look at the CDI and get all the needed information from it, is an important method to operate the airplane. If the CDI has not to be designed reasonable enough, pilots will need more time and attention to get the data, which will increase the MWL and lead to a higher possibility of the wrong operation.

Nowadays, several types of airplanes are widely used, such as Airbus 318 and Boeing

737. Different airplanes use different designed CDI, which means when the pilot operates different airplanes, his MWL would be different. Therefore, it is worthy to investigate the negative effects of these CDI on MWL and improve the insufficiency to develop a better designed CDI in order to help the pilots operate the airplane at a normal level of MWL in the future flight environment.

1.3 Aims and Objectives

The aim of this project is to find out what features on the CDI of the airplane will influence the MWL of the pilot during the flight, such as the different layout of the CDI in different types of airplanes and the different types of the CDI (traditional or modern). Besides, the impact of the familiarity with flight process and the CDI on pilot's MWL will be discussed.

The key objectives of this project are:

1) Compare several different types of CDI and analyse the key differences among them.
2) Do a number of flight tasks related to the different CDIs mentioned in 1) on a flight simulator and gather the data about participants' heart rate and brain activity.
3) Analyse the collected data and find out how the familiarity with the flight progress and the CDI influence MWL.
4) Analyse the collected data and combine the results with the results form 1). Find out the relationship between MWL and the different design of the CDI.
5) Use the conclusion to improve the design of the CDI.

2 Literature review/Related work

Although the concept of MWL has been put forward and studied for nearly 40 years, its definition is still controversial [9]. Intuitively, the construct of MWL can be described as the amount of necessary mental work consumed by a certain task over a given period of time [10], which is affected by multiple extra factors at the same time such as physical demand [9]. In general, MWL is a key criterion for evaluating operator and system performance [10], designers can design a more suitable system for their users by estimating their MWL. Furthermore, based on the theory that an individual's MWL is more likely to increase when he operates a complex system [3], the potentiality of mental overload is higher in these days because modern systems have become increasingly complex as a result of technological development [11]. According to Young and Stanton's opinion, both mental underload and overload will have a negative influence on human performance [11], it is important to keep the human MWL at an optimal level to minimize human errors and maximize system safety. Therefore, it is useful to investigate pilot's MWL for the purpose of keeping it at a normal level and improving the flight safety.

Based on the theory that MWL is susceptible to physiological measures [12], it is reliable to evaluate MWL by related assessment techniques. Hence, this section will first describe three psychophysiological measures and the reason why using them in this project. Besides, current studies related to pilot's MWL will be discussed.

2.1 MWL related assessment techniques

Cognition can be defined as the mental process of getting knowledge and understanding, these processes are called as cognitive processes which include "thinking, knowing, remembering, judging, and problem-solving" [13]. For the reason that pilots are required to use various cognitive processes during the flight task, numerous assessment techniques are needed for measuring MWL [5]. According to O'Donnell and Roberts [14], most procedures for scientific classification evaluation can be classified into three main categories.

The first one is *Subjective measures*, which means all the operators need to make subjective comments about their MWL according to various rating scales or a set of questionnaires. Secondly, *Physiological measures* represents inferring a value of MWL based on any physiological response from the operator, such as brain activity or the eye movement. The last category is *Performance-based measures*. An index of MWL can be deduced from the objective performance notions on the primary tasks. Only *Subjective measures* and *Physiological measures* will be discussed in this part.

2.1.1 Electroencephalogram and Event-Related Potentials

As the recordings of electrophysiology of the cerebral cortex, Electroencephalogram (EEG) (shown in Fig 1) is commonly used to measure MWL [15] as a *Physiological measure*. Event-related potentials (ERPs) can be seen as a combination of EEG data and its corresponding events. According to the research Charles and Nixon [12] did, the captured EEG data can be decomposed in a needed frequency range, which makes it more flexible for different researches. Additionally, although the analysis of the data is complicated, the brain activity in specific ERPs presents a good reflection of MWL, particularly sudden shifts in MWL levels [12]. Moreover, the flight task in this project requires various brain regions to be more involved than others, for example, the specifications of visual flight rules (VFR) and instrument flight rules (IFR) are very different on the visual system and higher processing capabilities [5]. Using EEG can show the distribution of electrical activity over the scalp clearly [5] by creating detailed maps and it makes EEG more suitable for this project. In fact, EEG has been successfully used for several pilot's MWL related experiments in past 30 years [16]. Hence EEG has been chosen as an assessment technique in this project.

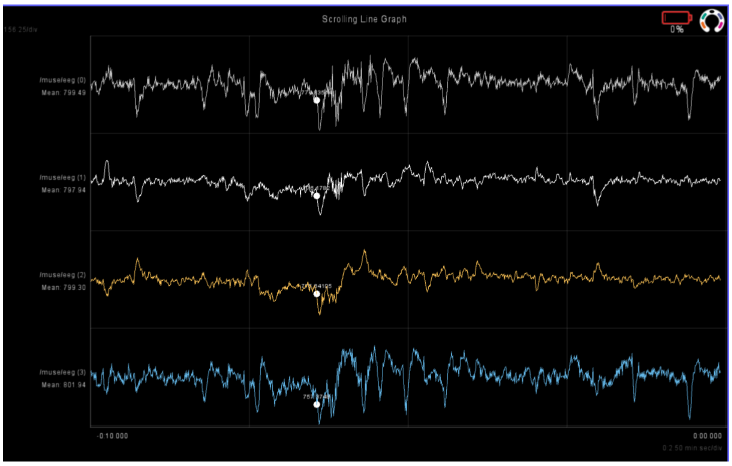


Figure 1: EEG data example

However, because the sensors that are used to measure EEG need to be placed on participant's scalp or forehead, its interference with ecological validity cannot be ignored. Ecological validity can be explained as the degree of whether the laboratory result can be generalized to the natural world [17]. In the natural world, people will not wear an EEG equipment during their daily life, hence when they put on these sensors in the laboratory environment, their behaviour might be influenced [18] and the ability of the gathered data to reflect the situations in the real world might decline, in other words, the ecological validity of the experiment will be reduced.

2.1.2 Heart Rate Variability

Heart rate (HR) and Heart rate variability (HRV) are two common *Physiological measures* in MWL measurement.

HR varies greatly under different conditions, for example, there is a significant difference in HR between a low demand task and a higher one [12]. However, for the reason that the mean HR observed under different conditions in the task had no notable different [19], HRV has been chosen as the measure in this project. For a healthy individual, the intervals between his successive heartbeats are different (shown in Fig 2), in other words, healthy person's heartbeats are "irregular". HRV is used to measure the irregularity of the heartbeats (shown in Fig 3). Compared with HR, Fallahi et al.'s [20] date shows that HRV can be affected as conditions changed from low to high traffic density in a traffic monitoring job. Additionally, the same change was observed by Delaney and Brodie during a Stroop test [21]. In addition, the pilot task did by Jaiswal et al. [22] proved that using HRV as an assessment technique is workable.

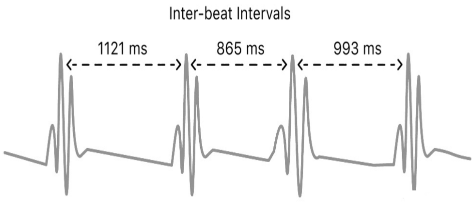


Figure 2: "irregular" heartbeats

Figure 3: HRV example

Nonetheless, there are diverse scale of methods used for HRV quantification, therefore the standard value of HRV becomes complicated. There are numerous of qualification of HRV and most of them are designed for a specific scenario, which means it is unreasonable to use these scale methods in other projects [23]. It is challenged for this project to find a suitable HRV standard value as the base-line. Despite of the difficulty of choosing standard value, the sensors used to monitoring wearer's HRV will also cause the issues related to ecological validity (as mentioned in 2.1.1). Extra sensors placed on the participant might influence his HRV and the reliability of the HRV data might be interfered.

2.1.3 NASA Task Load Index

As one of the validated largely-used *Subjective measures* of MWL, NASA Task Load Index (NASA-TLX) acts as base-line in MWL measurement [24]. It is a multidimensional scale that was originally created for the aviation industry. The final MWL score is calculated as a weighted average of 15 attributes in 6 dimensions depends on the questions in NASA-TLX (shown in Fig 4). Moreover, NASA-TLX has already been successfully used in thousands of studies related to assess workload in various environments such as aircraft cockpits and laboratory tests [25]. Except the high authority and high operability, NASA-

TLX can use the attributes in the questions to infer whether the participant's answer is reliable or not [26]. For example, if a participant's answer is high "physical demand" and low "effort", there exists a conflict and his answers might be considered meaningless. It is useful for researchers to identify the bad data from all the gathered data. Therefore, NASA-TLX has been selected for assessment.

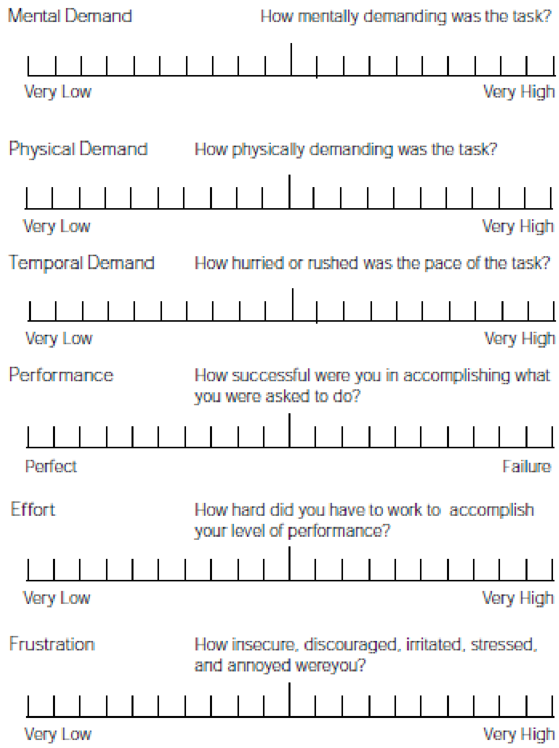


Figure 4: NASA-TLX [25]

2.2 Existing studies

2.2.1 Human-system interface & MWL studies

Plenty of relevant studies proved that a human–system interface (HSI) can have a significant impact on operator's MWL. According to the data provided by Piechulla et al. [27], an adaptive HSI is helpful to reduce driver's MWL during driving. Besides, Hancock et al. [28] proved that MWL is dynamic when using different HSI. Moreover, the same conclusion has been drawn by Jou et al. [29] when they tried to find out how the advanced nuclear power plants HSI automation affect the operator's MWL.

2.2.2 Pilot MWL studies

The researches related to pilot's MWL have been going for many years and some conclusions about the relationship between flight performance and different MWL have already been drawn. However, most of these studies were related to the performance when the participant flew the same type of airplane under different pressures. For example, Morris and Leung's [6] study focused on how different did pilot perform when he tried to finish a desktop, computer-based exercise that simulated a set of flight deck activities on a particular type of airplane under different MWL, Wei et al.'s [8] study concentrated on the performance under increasingly MWL on a specific airplane on a flight simulator. Except the experiments on the computers or simulators, studies on real flight process with different MWL have been done, such as explore the relationship between pilot performance and MWL in a rotary wing aircraft [30].

Current studies have shown that the participants' MWL are different when they are dealing with information on different CDI [31]. Nevertheless, there are not many studies related to how different CDI impact pilot's MWL during flight. Based on the theory that HSI has a significant impact on MWL [27], it is deserving to find out the relationship between different layouts of airplanes' CDI and MWL of pilots which has not been studied in depth.

3 Experiment Design & Procedure

A small pilot relevant study has been done for the reason to gather enough experience for the formal and bigger pilot study in this project. The objective in this small study is to investigate whether the traditional CDI and the modern CDI will have different impact on participants' MWL.

3.1 Equipment & Techniques

This section will introduce the equipment and techniques which will be used in this study, a diagram will be shown in the end of this section to describe the whole data gathering process.

3.1.1 Lab Streaming Layer

As Wilson [5] mentioned that pilot related study needs not only one technique to measure participant's MWL, a challenge related to synchronize the data gathered from different equipment has arose. In general, there are three common timing errors might occurred during the data gathering stage. The first one called constant offset can be corrected after this offset being measured. However, the other two errors named drift and jitter are difficult to be identified and corrected because they are some random variations. Therefore, if multiple sensors are used in this experiment, it is difficult to synchronize these data streams and the reliability of the analysis result will be influenced. Hence, Lab Streaming Layer (LSL) has been used during the experiment.

LSL is "a system for collecting and unifying the measurement time series in research experiments" [32] with a core transport library named liblsl. As shown in Fig 5, the signals come from data gathering sensors will be wrapped with their corresponding header data as streams and transferred by liblsl library embedded in LSL to the data processing software. The streams transferred through LSL will use the LSL built-in synchronized functionality which is designed to "achieve sub-millisecond accuracy on a local network of computer" [32], so that the impact of timing error can be minimized.

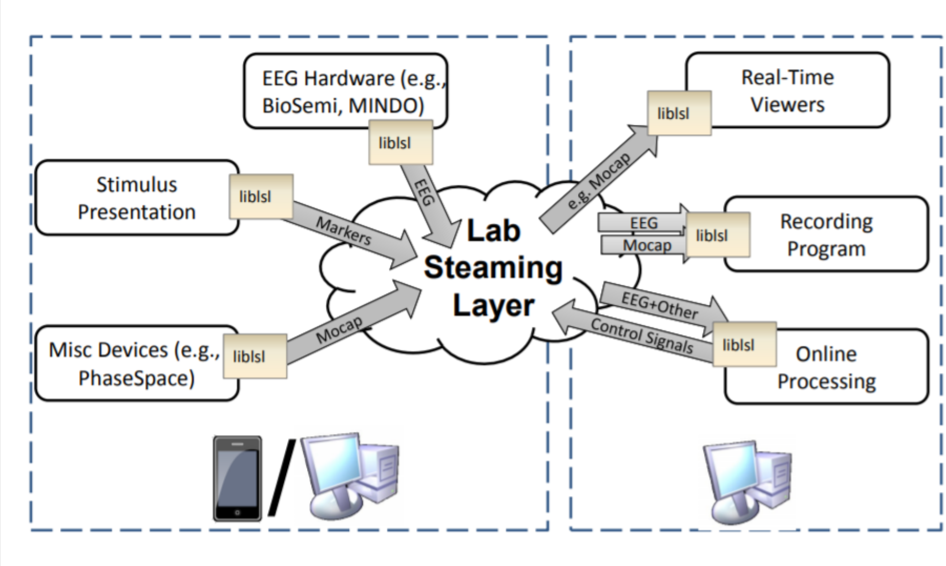


Figure 5: LSL working principle [33]

3.1.2 N-Back Task

The N-Back task is a widely used working memory measure task [34], participants need to determine whether each stimulus in a sequence matches the one that previously appeared in n items ago in the task. Fig 6, 7, 8 and 9 show an example of completing the N-Back task.

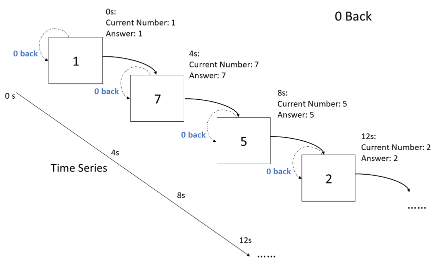


Figure 6: 0-Back

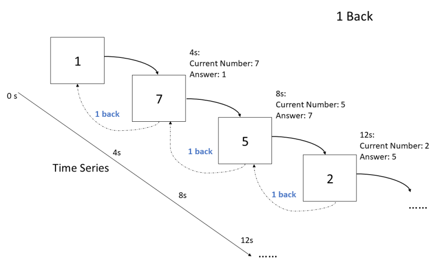


Figure 7: 1-Back

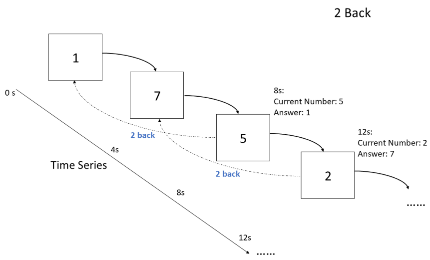


Figure 8: 2-Back

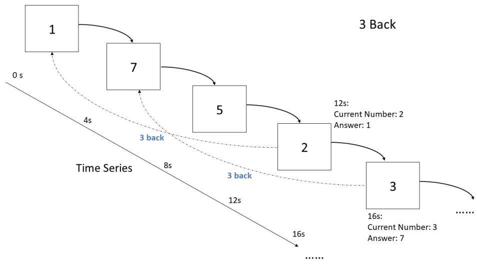


Figure 9: 3-Back

Because this experiment is aimed to find out how the different CDI of the airplane will influence the MWL of the pilots under different levels of workload, it is important to design a parallel task which can induce particular's different workload based on the its design. According to the findings from Brouwer et al. [35] and Berka et al. [36], workload induced by N-Back task can be categorized in single trail, which means N-Back task can act as the inducer of different workload. Hence N-Back task has been chosen to simulate the different difficulties of the parallel task in this study.

3.1.3 Muse2, Polar H7, PsychoPy & LabRecorder

Muse2 An EEG equipment called Muse2 (shown in Fig 10 & 11) has been chosen as the EEG measure in this experiment which is a headband that needs to be placed on participant's forehead for measuring (shown in Fig 12). 4 EEG sensors are placed on Muse2 (shown in Fig 13) which positions are based on the international EEG placement system (shown in Fig 14). AF7 and AF8 are forehead electrodes whereas TP9 and TP10 are ear electrodes, these embedded electrodes can detect the wearer's brain activity. Because it is a widely used tool to help with meditation, the comfort can be guaranteed so that the effect on participant's mental pressure is relatively small.

Figure 10: Muse2 #1

Figure 11: Muse2 #2

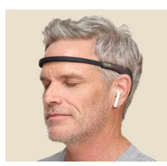
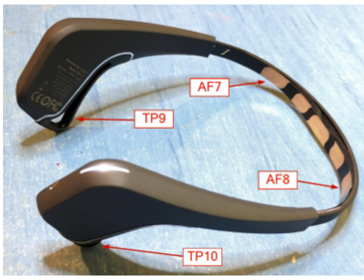


Figure 12: wearing Muse2



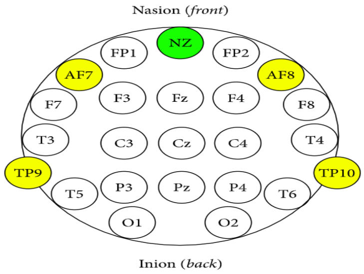


Figure 13: Muse-EEG-Sensors

Figure 14: EEG-Placement-System [37]

A software named *BlueMuse* is used in order to transfer the brain activity gathered by Muse2 to the data processing software through LSL. It is a Windows application that can detect Muse headset and stream data from it via LSL [38]. After the data from Muse2 has been transferred by *BlueMuse*, it becomes a stream (as mentioned in 3.1.1) and can be streamed by LSL to the data processing software.

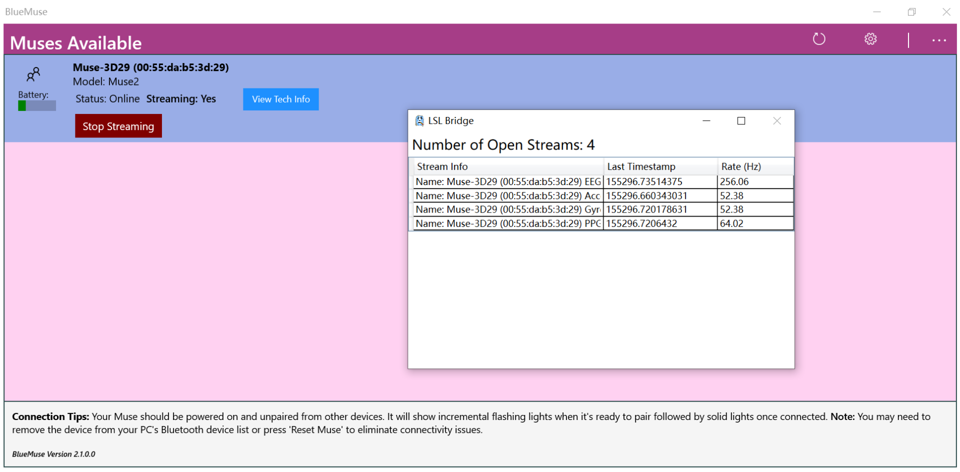


Figure 15: *BuleMuse*-Stream-Transferring-Interface

Polar H7 Considering to reduce the influence of ecological validity from HRV monitoring equipment, a sport chest strap named Polar H7 is chosen(shown in Fig 16) which can transfer wearer's HR and HRV through Bluetooth. Polar H7 is originally designed for collecting HR relevant data during its wearer's exercise process. For the purpose of decreasing the extra influence from the strap, this product is well designed to make sure it is comfortable to wear. Hence, the interference with participant's HRV caused by wearing extra sensors can be relatively minimized.

Figure 16: Polar H7

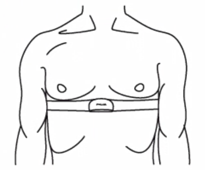


Figure 17: wearing Polar H7

Because the HR and HRV data come from Polar H7 need to be transferred via LSL, *BLEPolarDirect* [39] has been chosen to connect Polar H7 with the computer. It is a Windows application that can convert the HR and HRV data from Polar H7 into LSL stream.

PsychoPy PsychoPy [40], a widely used psychological task design software, is selected to display the N-Back task. It can simulate different kinds of stimulus (shown in Fig 18) and receive several kinds of responses (shown in Fig 19) as event codes, so it is not difficult to use PsychoPy to simulate the stimulus in the N-Back task and collect participants' responses. Additionally, it is easy to transfer the streams from sensors through its embedded LSL functionality. Hence it is suitable for this project.

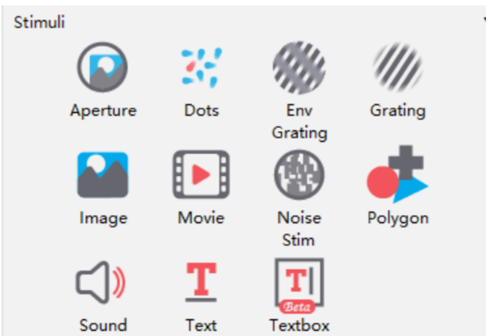


Figure 18: Stimulus

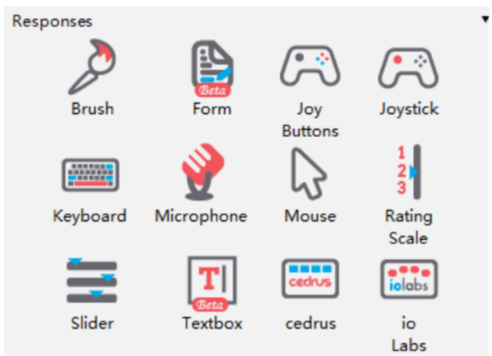


Figure 19: Responses

Figure 20: An experiment interface

Figure 21: PsychoPy-Code-Interface

LabRecorder LabRecorder is LSL's default recording program, it is used to receive the streams transferred by LSL and record them together into a XDF file with time synchronization between them [41]. XDF is a "general-purpose container format for multi-channel time series data with extensive associated meta information" and is widely used for storing biosignal data [42]. After the stream starting to be transferred through LSL, it will show up in the LabRecorder's stream box and user can choose the streams that he wants to record. In addition, user can choose when to start or stop the recording (shown in Fig 22). Besides, the storage path can be customized by user.

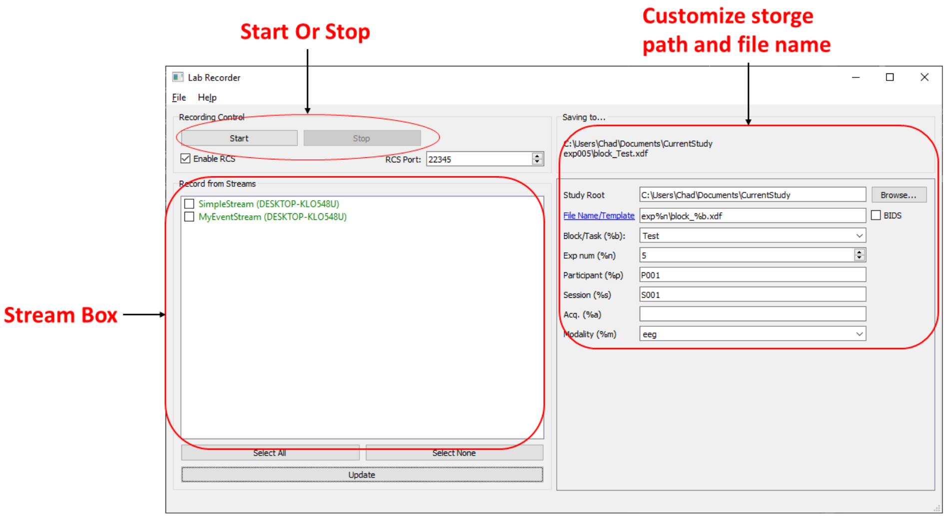


Figure 22: LabRecorder-Interface

The whole process of transferring collected data from sensors to the XDF file has been shown in Fig 23.

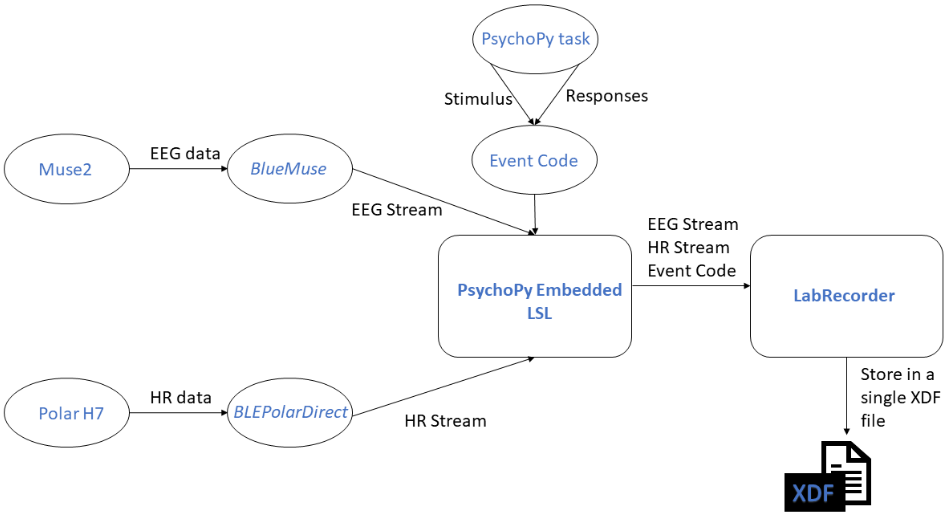


Figure 23: transferring process

3.2 Experiment Design

3.2.1 Task

For the purpose of making the experiment results more focus on the influence coming from CDI, the tasks had to try to minimum the impact from other factors. Firstly, the influence from different operating systems on different planes should be considered. Therefore, the planes used for the tasks have been chosen as Cessna 172 and Cessna 172s (shown in Fig 24 & 25), the only difference between these two types of planes is that Cessna 172 uses traditional CDI whereas Cessna 172s uses the modern one so that the influence from flight operating system can be minimized. In addition, Cessna is a type of plane which is usually used in flight school for training new pilots because of its relatively simple operating system, which means it is not difficult to operate for the participants who were involved in the study. Secondly, the tasks should try to reduce the extra operations during the flight which might distract participants. Thirdly, because this experiment focuses on the CDI, the tasks should make sure that participants need to use displays on the panel frequently. Fourthly, each task cannot take too long time to complete because it might have a negative impact on participants' physiological indicators and influence the accuracy of the gathered data [43].

Hence the experiment was designed as letting the participants finish a "Final Landing" task and a N-Back task at the same time. "Final Landing" is the final part during a flight, which means the plane is in the sky when the task starts, and it needs to be landed in the designated runway at the end of the task. The landing stage mainly requires pilots combine the outside environment and the displays on the panel so that the distractions from extra operations can be reduced. N-Back task was designed to display on a laptop's screen next to the flight simulator, a serial of numbers would be showed on the screen one by one, participants need to enter the number corresponding to the N that came before the current one by pressing "0" to "9" key on the keyboard.

Four tasks were created based on different CDI and the difficulty levels of N-Back task. For each CDI, participants need to finish 1-Back and 2-Back task.

Figure 24: Cessna172

Figure 25: Cessna172S

3.2.2 Participants

Four participants (3 male, 1 female) with an average age of 21 years were recruited to take part in the experiment. Participants were all from Computer Science department of the University of Nottingham Ningbo China. All participants had normal or corrected vision and reported no history of head trauma or brain damage, they were all in good health without any heart problem. The study was approved by the school's ethics committee. Before the study started, they all received the *Participant Inform Sheet* which introduced the content of the study and signed the *Participant Consent Form*.

3.3 Procedure

Because of the professional requirements of the flight simulator, participants were asked to watch a tutorial video about how to manipulate the simulator and some basic knowledge about aircraft before coming to the flight simulator studio. After they had some understanding of the aircraft and landing operation, a practice session was provided so that they can come to the studio and get familiar with the flight simulator. Besides, an introduction related to N-Back task was been provided before the formal experiment began.

Before each task began, Muse2 was placed on participant's forehead to record the brain activity of the participant, Polar H7 was wore by participant for HR and HRV recording. At the start point of each task, the plane was 1000 foot in the air and was within three nautical miles of the target landing point. Participants were asked to keep the approach speed at 65 to 75 knots per hour so that the whole task took about 5 to 6 minutes to finish. The N-Back task started in the third minute and ended in the fifth. The respond time and the answer given by participants were recorded through PsychoPy and later analysed by the researcher.

Figure 26: During Task #1

Figure 27: During Task #2

After each task, participants were required to complete the NASA-TLX form to evaluate their MWL during the task by themselves. Then they can relax themselves for 2 minutes so that they can start the next task in a good state.

3.4 Reflection

Although until now the data analysing process has not started, several issues related to experiment design have been detected during the experiment procedure.

Firstly, the flight simulator studio existed several uncontrollable factors. For example, it is difficult to keep the room and its surroundings silent enough during the experiment, the unexpected noises might influence the participant's concentration to the task and have an impact on both the performance during the task and the quality of gathered biosignal data. Besides, the temperature was difficult to control for the reason that this studio is a large room, which might cause the room temperature to be different every experiment time and become a extra experimental variable. The solution to this problem is considered as moving a flight simulator to a specific room in order to make sure all the needed conditions can be conformed.

Secondly, the N-Back task designed in PsychoPy was not reasonable enough. The task was designed as quit and stop recording the transferred data from sensors when the participant presses the last corresponding key, however, the flight task did not finish as the same time as the N-Back task during the study. Hence, the flight task corresponding data after the N-Back task ended has not been recorded into the XDF file. Besides, three of the four participants were found to be helping with their hands when they tried to complete the N-Back task. They put one figure over the number they had seen on the screen so that they can directly press the key next trail without memorizing it, it is contrary to the aim of this task because their workload did not be induced. Additionally, the action of pressing keys on the laptop had a great impact on flight task. Therefore, the responding way of the N-Back task is considered to change as letting the participants speaking out the required number instead of pressing the corresponding key. Except the way of responding, the way to quit the task has been changed into stopping by the researcher.

Thirdly, the content of the flight task needs to be reconsidered. By observing the experiment process, two of the participants operated the simulator by just looking at the surrounding environment without using the data on the CDI, which makes the gathered data useless because it did not relate to the design of the CDI. In addition, the landing operation is difficult for untrained participants, all of them crushed during the task. So the flight task could be changed as turning right four times in a row. The task could be designed above a field or a forest, hence there will not have any reference or light outside so that the participants can only focus on the CDI to finish the task.

4 Progress

4.1 Progress Management

Because this project needs human participants, the development began with submitting the related ethic forms. Then after getting familiar with the EEG and HRV technology which will be used during the project, the tests' design began as soon as the Ethics Committee give the approval. The most time-consuming part of this project is analysing the collected data and combined it with the design of the panels, so part of the winter vacation will be used to data analysis. The detailed timeline has been presented as a Gannt chart shown in Table 1.

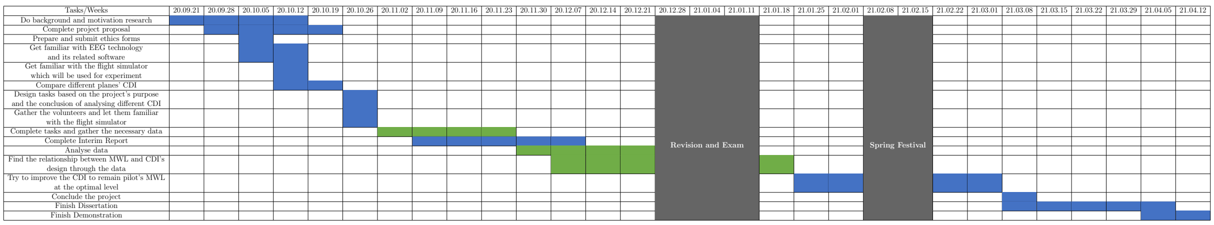


Table 1: Original timetable

Because the device for monitoring HR was not be considered to use at the beginning of the study, the timeline related to equipment familiarity has been adjusted. Besides, a new small pilot relevant study has been designed for the purpose of getting familiar with the whole process of gathering data before the formal pilot study. Because of this new experiment, the formal pilot study has been scheduled after the interim report finished. Another change in the timeline is about the revision week, the exam timetable shows that the final exam will start on December 28^{th} and finish on January 4^{th} so that the length of the revision week has been modified. The new time table has been shown in Table 2.

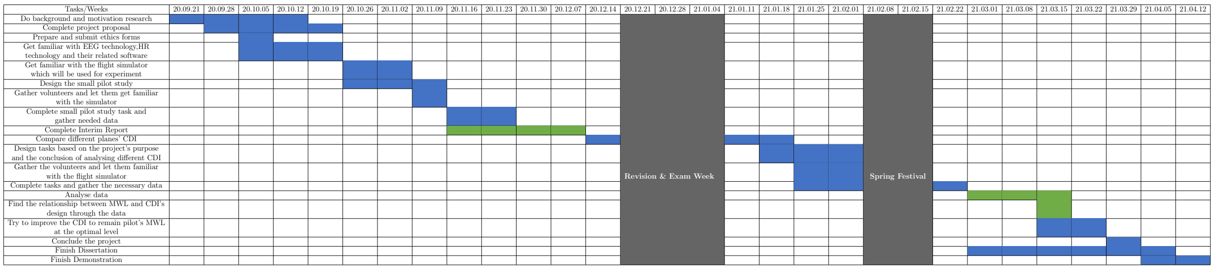


Table 2: Updated timetable

4.2 Progress Reflection

Until now the achievement of the project is good. Firstly, plenty of experiences about how to use related equipment and software have been gained and it is useful for the coming study. Secondly, except gathering some valuable data, based on the small pilot study several problems in the design of the task have been detected, which is helpful to the design of the formal pilot study. Besides, EEG data processing relevant knowledge has been learnt during November and it can reduce the pressure during the second semester.

Although the project is basically carried out as planned, some problems are existed. One of the problems it that due to the lack of relevant knowledge, a lot of time were wasted on finding the usable software for transferring data from sensors to the computer. Additionally, because there was some bugs in HRV transferred software, the time consumed on fixing it was unplanned, which had some impact on the time allocation. Moreover, because of the lack knowledge of aviation, the time spent on task design exceeded the plan.

Overall, the progress of the project does not have big issues at this point. Besides, because the new timetable has considered the difficulties encountered so far, it is more reasonable for the further project. Hence, the next step of this project is to follow the new timetable and focus on designing the formal pilot study.

Bibliography

[1] International Air Transport Association. ANNUAL REVIEW 2019, 2020. Retrieved October 1, 2020 from `https://annualreview.iata.org/?_ga=2.183168800.780959666.1602231137-1689187455.1601106589`.

[2] General Aviation Manufacturers Association. General Aviation Aircraft Shipment Report: 2019 Year End, 2020. Retrieved September 21, 2020 from `https://gama.aero/wp-content/uploads/2019ShipmentReport03162020.pdf`.

[3] G Tsaparlis. Cognitive demand. *Encyclopedia of science education*, pages 1–4, 2014.

[4] Peter Robinson. Task complexity, task difficulty, and task production: Exploring interactions in a componential framework. *Applied linguistics*, 22(1):27–57, 2001.

[5] Glenn F Wilson. An analysis of mental workload in pilots during flight using multiple psychophysiological measures. *The International Journal of Aviation Psychology*, 12(1):3–18, 2002.

[6] Charles H. Morris and Ying K. Leung. Pilot mental workload: how well do pilots really perform? *Ergonomics*, 49(15):1581–1596, 2006. PMID: 17090505.

[7] Wouter Kool, Joseph T McGuire, Zev B Rosen, and Matthew M Botvinick. Decision making and the avoidance of cognitive demand. *Journal of Experimental Psychology: General*, 139(4):665–682, 2010.

[8] Zongmin Wei, Damin Zhuang, Xiaoru Wanyan, Chen Liu, and Huan Zhuang. A model for discrimination and prediction of mental workload of aircraft cockpit display interface. *Chinese Journal of Aeronautics*, 27(5):1070 – 1077, 2014.

[9] Luca Longo. Formalising human mental workload as a defeasible computational concept. 2014.

[10] Brad Cain. A review of the mental workload literature. Technical report, Defence Research And Development Toronto (Canada), 2007.

[11] Luca Longo. Formalising human mental workload as a defeasible computational concept. 2014.

[12] Rebecca L. Charles and Jim Nixon. Measuring mental workload using physiological measures: A systematic review. *Applied Ergonomics*, 74:221 – 232, 2019.

[13] Gary R VandenBos. *APA dictionary of psychology.* American Psychological Association, 2007.

[14] Robert D O'DONNELL. Workload assessment methodology. *Cognitive processes and performance*, 1986.

[15] Ying Lean and Fu Shan. Brief review on physiological and biochemical evaluations of human mental workload. *Human Factors and Ergonomics in Manufacturing & Service Industries*, 22(3):177–187, 2012.

[16] John A Caldwell and JA Lewis. The feasibility of collecting in-flight eeg data from helicopter pilots. *Aviation, space, and environmental medicine*, 66(9):883–889, 1995.

[17] Mark A Schmuckler. What is ecological validity? a dimensional analysis. *Infancy*, 2(4):419–436, 2001.

[18] Martin G Bleichner, Micha Lundbeck, Matthias Selisky, Falk Minow, Manuela Jäger, Reiner Emkes, Stefan Debener, and Maarten De Vos. Exploring miniaturized eeg electrodes for brain-computer interfaces. an eeg you do not see? *Physiological reports*, 3(4), 2015.

[19] Y.-H. Lee and B.-S. Liu. Inflight workload assessment: Comparison of subjective and physiological measurements. *Aviation Space and Environmental Medicine*, 74(10):1078–1084, 2003. cited By 78.

[20] Majid Fallahi, Majid Motamedzade, Rashid Heidarimoghadam, Ali Reza Soltanian, and Shinji Miyake. Effects of mental workload on physiological and subjective responses during traffic density monitoring: A field study. *Applied Ergonomics*, 52:95 – 103, 2016.

[21] J.P.A. Delaney and D.A. Brodie. Effects of short-term psychological stress on the time and frequency domains of heart-rate variability. *Perceptual and Motor Skills*, 91(2):515–524, 2000. cited By 170.

[22] Dibyanshu Jaiswal, Arijit Chowdhury, Tanushree Banerjee, and Debatri Chatterjee. Effect of mental workload on breathing pattern and heart rate for a working memory task: A pilot study. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 2202–2206. IEEE, 2019.

[23] Marek Malik and A John Camm. Heart rate variability. *Clinical cardiology*, 13(8):570–576, 1990.

[24] Sandra G Hart and Lowell E Staveland. Development of nasa-tlx (task load index): Results of empirical and theoretical research. In *Advances in psychology*, volume 52, pages 139–183. Elsevier, 1988.

[25] NASA. NASA TLX, Task Load Index, 2019. Retrieved October 23, 2020 from `https://humansystems.arc.nasa.gov/groups/tlx/`.

[26] Lucas Rizzo and Luca Longo. Representing and inferring mental workload via defeasible reasoning: A comparison with the nasa task load index and the workload profile. In *AI^3@ AI* IA*, pages 126–140, 2017.

[27] Walter Piechulla, Christoph Mayser, Helmar Gehrke, and Winfried König. Reducing drivers' mental workload by means of an adaptive man–machine interface. *Transportation Research Part F: Traffic Psychology and Behaviour*, 6(4):233 – 248, 2003.

[28] Peter A Hancock and Mark H Chignell. Mental workload dynamics in adaptive interface design. *IEEE transactions on Systems, Man, and Cybernetics*, 18(4):647–658, 1988.

[29] Yung-Tsan Jou, Tzu-Chung Yenn, Chiuhsiang Joe Lin, Chih-Wei Yang, and Chih-Cheng Chiang. Evaluation of operators' mental workload of human–system interface automation in the advanced nuclear power plants. *Nuclear Engineering and Design*, 239(11):2537 – 2542, 2009.

[30] ERIK J. SIREVAAG, ARTHUR F. KRAMER, CHRISTOPHER D. WICKENS MARK REISWEBER, DAVID L. STRAYER, and JAMES F. GRENELL. Assessment of pilot performance and mental workload in rotary wing aircraft. *Ergonomics*, 36(9):1121–1140, 1993. PMID: 8404840.

[31] Hengyang Wei, Damin Zhuang, Xiaoru Wanyan, and Qun Wang. An experimental analysis of situation awareness for cockpit display interface evaluation based on flight simulation. *Chinese Journal of Aeronautics*, 26(4):884 – 889, 2013.

[32] Christian Kothe, David Medine, Chadwick Boulay, Matthew Grivich, Tristan Stenner. Lab Streaming Layer, 2019. Version 1.13. 2018. URL: `https://github.com/sccn/labstreaminglayer`.

[33] The Qualcomm Institute. Demo 1 The Lab Streaming Layer, 2013. Retrieved October 20, 2020 from `https://www.youtube.com/watch?v=Y1at7yrcFW0`.

[34] Susanne M Jaeggi, Martin Buschkuehl, Walter J Perrig, and Beat Meier. The concurrent validity of the n-back task as a working memory measure. *Memory*, 18(4):394–412, 2010.

[35] Anne-Marie Brouwer, Maarten A Hogervorst, Jan BF Van Erp, Tobias Heffelaar, Patrick H Zimmerman, and Robert Oostenveld. Estimating workload using eeg spectral power and erps in the n-back task. *Journal of neural engineering*, 9(4):045008, 2012.

[36] Chris Berka, Daniel J Levendowski, Michelle N Lumicao, Alan Yau, Gene Davis, Vladimir T Zivkovic, Richard E Olmstead, Patrice D Tremoulet, and Patrick L Craven. Eeg correlates of task engagement and mental workload in vigilance, learning,

and memory tasks. *Aviation, space, and environmental medicine*, 78(5):B231–B244, 2007.

[37] Herbert H Jasper. The ten-twenty electrode system of the international federation. *Electroencephalogr. Clin. Neurophysiol.*, 10:370–375, 1958.

[38] Christian Kothe. Xdf, 2016. Version 1.12. Mar. 1, 2016 . URL:`https://github.com/sccn/xdf/releases/tag/v1.12`.

[39] Robert Twomey . nosc-toolkit, 2019. March. 12, 2019 . URL:`https://github.com/roberttwomey/nosc-toolkit`.

[40] Jonathan Peirce. Psychopy^3, 2020. Version 2020.2.8. Feb 2020 . URL:`https://www.psychopy.org/`.

[41] Tristan Stenner, Chadwick Boulay. App-labrecorder, 2020. Version 1.13.1. Aug. 12, 2020 . URL:`https://github.com/labstreaminglayer/App-LabRecorder`.

[42] Jason Kowaleski. Bluemuse, 2020. Version 2.1.0. Jun. 4, 2020 . URL:`https://github.com/kowalej/BlueMuse`.

[43] P. A. Hancock and M. H. Chignell. Mental workload dynamics in adaptive interface design. *IEEE Transactions on Systems, Man, and Cybernetics*, 18(4):647–658, 1988.