Assessing Workload in Human Swarm Interaction

Samuel Clark

University of New South Wales at the Australian Defence Force Academy

This paper will assess the impact of workload in human swarm interaction. There exist many examples within society where workload influences performance both positively and negatively. Furthermore, the topic of human swarm interaction is becoming increasingly popular as an increasing number of possibilities for its application arise. Military drone pilots are an example of a possible ambassador for human swarm interaction. United States Air Force (USAF) drone pilots in theory could command multiple drones in operations. However, to achieve this, a greater understanding of the implications of increased workload in a human swarm environment is required. Research has been conducted into areas of workload in human control. For example, the effects of workload on public transport operators and workload in nursing. [27, 28] However, there have been no prevalent studies into workload in human swarm interaction. Consequently, this paper investigates the results obtained through an experiment conducted with the intent of replicating an increasing workload in human swarm interaction. The paper finds that an improvement in performance of the operators is seen as the workload is increased. The findings support the inverted U model for workload which states that poor performance will result from both underload and overload with optimum performance being achieved between the two.

Table of Contents

I. Introduction ......................................................................................................................1
II. Aim ....................................................................................................................................2
III. Literature Review .........................................................................................................2
   A. Workload ....................................................................................................................2
      1. Arousal .....................................................................................................................3
      2. Fatigue .....................................................................................................................3
      3. Workload Model .......................................................................................................3
   B. Measurement Methods ...............................................................................................3
      1. Subjective Measures ...............................................................................................4
      2. Performance Measures .........................................................................................4
   C. Considerations .............................................................................................................5
   D. Recommendations ......................................................................................................5
IV. Methodology ................................................................................................................6
   A. Ethics Clearance Application ......................................................................................6
   B. Experiment Population ...............................................................................................6
   C. Equipment ...................................................................................................................6
   D. Experiment Design .....................................................................................................6
V. Hypothesis .......................................................................................................................7
VI. Results ............................................................................................................................7
VII. Discussion ....................................................................................................................9
VIII. Conclusion ................................................................................................................10
IX. Future Work ................................................................................................................10
References ..........................................................................................................................11
Annex A ............................................................................................................................12
Annex B ............................................................................................................................13

I. Introduction

Many people work in highly complex and multiple task oriented roles. A common challenge involved with such positions is assessing the effects that increased workload has on performance. A Large number of studies have been conducted into this area of research. Many have provided limited insight into the concept of workload

---

1 PLTOFF, School of Engineering & Information Technology. ZEIT4500

Final Project Paper 2016, UNSW Canberra at ADFA
and its consequent effects on performance. This paper will provide an analysis of workload in human swarm interaction.

Workload is an important area of study. It holds direct relevance too many roles within society. One of the more prominent positions within society which has gathered attention over recent years is commercial pilots. A study conducted by Harvard University instructor David Robelik found that the odds of a plane crash occurring are 1 in 1.2 million and the odds of dying in a plane crash are 1 in 11 million. [1] More often than not, plane crashes are caused by human error with statistics showing that this accounts for 80% of plane crashes. [2] Of this 80%, 20% are a result of pilot fatigue with a survey conducted by British Airline Pilots Association finding that 43% of the surveyed pilots had involuntarily fallen asleep while in command of a plane. [2] In addition to this, 31% of those pilots woke up to find the co-pilot also asleep. [2] Evidently a problem exists in the management of pilot fatigue.

There has been extensive support for reevaluation of current laws and standards surrounding the working conditions for pilots. Many of these conversations include stricter laws governing the total hours a pilot is allowed to fly or the number of hours a pilot is required to rest prior to flight. It is important to understand the statistics which are inherent in the profession in order to understand the limitations of such laws. Research shows that the average person can remain alert for 16hrs in a day. [2] Although this is true, there exist many other conditions which can influence the performance of the pilots. For example, a pilot landing a plane at 05:00 h in the morning experiences impairments equivalent to that of having a blood alcohol reading of 0.08%. [2] Workload is considered a significant contributor to fatigue both physical and cognitive. A possible solution is introducing stricter laws enforcing fatigue management for pilots. An additional and perhaps more appropriate suggestion is to investigate the effects of workload and attempt to reduce the impact of workload on the pilot thus reducing the cumulative fatigue.

Workload is an important factor when considering these debates. There exists a very limited understanding of its effects and its degree of impact on factors such as performance. As stated prior, workload is known to be a contributor to fatigue both physical and cognitive. The measure and understanding of these effects are important to the development of techniques used to manage workload. These methods could ultimately be used to reduce its impact. In the case of the pilot fatigue example, this would enable control measures to be implemented and through these measures the compounding fatigue could be reduced. A simple improvement in our understanding of workload could reduce the 20% statistic related to fatigue resultant commercial plane crashes.

II. Aim

The purpose of this paper is to explore the combination of workload and human swarm interaction. The paper will investigate the effects that workload has on the human swarm interaction and ultimately evaluate the impact on performance.

III. Literature Review

This section will examine the concept of workload and its associated effects. We will first explain workload and investigate how it is defined and understood in the academic community. A brief analysis of arousal and fatigue theory will be performed as well as a model for workload and performance. This will be followed by a discussion of measurement methods, including both subjective and performance measurement techniques. The literature review will conclude with considerations and recommendations.

A. Workload

Workload is a prominent subject amongst academics and practitioners alike. Although this is true, there is no unanimously agreed upon definition for the term. Originally the term was used to describe the physical phenomenon however has today evolved to be used predominantly for the cognitive phenomenon. The terms definitions consequently range from those of physical workload to those of cognitive workload. For the purposes of this paper the focus will be on the latter, cognitive workload.

The term workload describes a person’s ability to perform a task given its demands. It describes the balance between resource supply and resource demand. Workload is quantified using a number of measures many of which are subjective. [3] This is a consequence of the limited knowledge which exists surrounding the topic and ultimately the complexity of human behaviour and its influence on the term. Although many methods exist for its measure, there is no prominent link between them. [4] Consequently, multiple methods can be difficult to compare and analyse. Ironically, the most common analysis methods used for its measure still rely on comparison techniques. Because of this, workload is described as disparate and multidimensional. [5] In order to measure the

Final Project Paper 2016, UNSW Canberra at ADFA
effects of workload, it is paramount to understand the influencing factors. Workload is defined by a number of factors which when observed simultaneously can indicate workload and subsequently performance.

There are many factors which can be observed as influencing workload. These include factors such as; attention, arousal, motivation and complexity. [5, 6] The greatest difficulty associated with workload results from its subjective nature. Measuring these factors for one individual will likely under the same conditions yield different results for a different individual. For example, motivation will be highly dependent on the person and their individual characteristics, personality and beliefs. It is important to gain an understanding of the participant’s perceived level of engagement in the task as a supplementary source of information to compare with other collated data such as physiological data or performance data. Additionally it is important to include as large as possible group in the experiment to reduce the effect which personal characteristics and possible outliers may have on the results. It is common for a small experiment to use between 20 and 30 participants. [7] The expected error margin can be calculated using 1/sqrt(N) where N is the sample size. [7] This yields an 18% error margin for an experiment with 30 participants.

1. Arousal
   The literature often discusses the influence of arousal on workload. One of the most accepted and comprehensive definitions of the term was proposed by Razmjou in 1996. The term arousal is used to describe the human cognitive state. [8] There exist many cognitive states which can be experienced. The two extremes are the sleep state and the alert state. The term is therefore used to define the cognitive state which is observed through human behaviour. [9] Arousal is important because it regulates the response in given situations. Too high or too low will cause the performance to suffer.

2. Fatigue
   A sustained increased in workload can cause fatigue, but workload itself does not necessarily lead to fatigue. Fatigue is defined as the difficulty in performing/sustaining voluntary activity. [10] A common effect of cognitive load inducing fatigue is a decrease in parasympathetic activity and an increase in sympathetic activity. [10] The parasympathetic nervous system is responsible for rest and recovery and the sympathetic nervous system is responsible for preparing the body for activity; raising heart rate, release of sugar in the body ect. [11] This relationship suggests the subconscious tendency to strive to maintain performance. This subconscious tendency is greatly influenced by individual motivation. [12, 13] The relationship of decreasing parasympathetic activity and increasing sympathetic activity is described as a state of autonomic hypervigilance. [11] This combination has been shown to cause a gradual reduction in brain activity responsible for cognitive processing. [11] Essentially, this is how sustained workload can induce fatigue.

3. Workload Model
   A model for describing the observed relationship between workload and performance is shown in figure 1. The model represents the relationship between cognitive workload and performance. This design shows many similarities to the famous Yerkes Dodson law of arousal. As can be seen the model for cognitive workload vs performance describes three zones in which a person can exist when conducting a task. These include underload, acceptable level and overloaded. [19] Underload causes a person to experience boredom, lack motivation and consequently perform poorly. Overloaded causes the person to be unable to appropriately perform due to the task demands being too high. Ultimately, the acceptable range falls in-between where the person is under enough load to make the task stimulating and thus increase motivation for the task and consequently performance. This model suggests that increasing workload will cause a possible initial improvement in performance and if increased further past a certain threshold, a possible decline in performance.

B. Measurement Methods
   Numerous papers have investigated workload and related topics. These papers range from investigations into effort and its relation to motivation through to physiological studies and how eye patterns can indicate fatigue. This range in research highlights the multidimensional nature of workload. Furthermore it outlines the difficulty involved in its exact definition. The importance lies in its measurement and how the measurements are used. With no singular definition of the term, there exists many measurement methods and types. It is therefore important to consider a number of factors when investigating the appropriate measurement methods.

Final Project Paper 2016, UNSW Canberra at ADFA
There are three main method types for measuring workload. These include subjective, physiological and performance method types. [4] Subjective methods are those which are based on scales, surveys and self-evaluation. [14] They are a measurement of a person’s perception of the workload which they are under. Physiological measures are based on known physical responses resulting from mental workload demands. These include changes in cardiac activity, brain activity and eye activity. Performance measures observe and identify success in given tasks corresponding to task demand. This is achieved through observing the change in performance as a person’s workload is increased. This enables an estimate for workload and its effects to be established.

When selecting a method for measuring workload, it is important to consider a number of criteria. The criteria includes: sensitivity, validity, reliability and intrusion. [4] Sensitivity is perhaps the most important as it is essential for the measurement method to be able to identify variations in workload and subsequent performance. Validity is also important as the measurement method must measure the required variables. Furthermore the measurement method must be reliable and consequently always yield the same results for identical conditions. Lastly intrusion of the measurement method must be considered. This is required as the measurement method must not influence the performance.

1. Subjective Measures

Subjective measurements are the simplest and arguably the most appropriate method for measuring workload. This is common belief as workload is highly subjective and affects people differently based on numerous factors many of which cannot be measured. Therefore many researchers conclude that if a person feels that they are under a lot of workload, then they likely are under a lot of workload. [4] When compared to the other types of measurement methods such as physiological measurement, the subjective approach lacks precision. However, subjective methods naturally account for all influencing factors and differences in personal characteristics and are therefore argued to be more practical. [4, 14] Subjective measures are also less intrusive and normally cheaper. Furthermore, studies have shown that physiological measures will often correlate closely with the subjective measures. [4] As a result of this, it is more common to adopt subjective measurement techniques.

There are two categories in subjective measuring. These include the unidimensional scales and the multidimensional scales. [4] The simplest method to use is the unidimensional method. This is simple because it does not require any complicated analysis techniques. As suggested by its title, it only has one dimension. The multidimensional scales are more complex and time consuming.

Multidimensional and unidimensional measurements are both commonly used in subjective measurement methods. Historically the multidimensional measurement techniques were considered superior. [15] Studies have recently indicated that the unidimensional methods can be equally effective. However, workload is considered multidimensional. [16] This makes the multidimensional methods more desirable for the measure of workload.

2. Performance Measures

Performance measures is one of the methods available for measuring workload. There are two ways of using performance measures which are commonly agreed upon. These include primary task performance measurement and secondary task measurement. [4, 17] These types of measurement techniques take advantage of the resource supply and demand limitations. It is suggested in the literature that tasks that employ the same or similar resource structures will over time result in deterioration in performance. [4] The primary and secondary tasks can be used in reference to each other as a means of measuring workload. Ultimately the idea being to use the decrease in performance of either method to estimate the workload.

The primary task measurement is more widely accepted for estimating workload. This method is more reliable for detecting changes in workload. [17] This is because the primary task is generally more demanding making changes in performance more noticeable. Furthermore the primary task is non-intrusive. The primary task should attract the majority of the participant’s focus thus making it more stable and less vulnerable to irregularities. Although the primary task is the most reliable and accepted for workload measurement, it is also beneficial to consider the secondary task.

In addition to the primary task measure, the secondary task measure can offer further information to assist in estimating workload. The secondary task is important as it is a measure of the performance resulting from the remaining cognitive capacity. This relates closely to resource theory and is caused by the competition of cognitive resources between the primary and secondary tasks. This forces prioritisation and resource management and can result in the remainder of the cognitive capacity depleting. This makes it an appropriate measure for workload. This is more reliable when used in conjunction with the primary task measure.
An important component of performance that could be tracked would be keypress data. More specifically this would include the time between keypresses. This would identify in high tempo periods the user’s reaction time. Research shows that the average person has a reaction time of between 300-400 milliseconds for visual stimulus. [18] This could be used as a primary task performance measurement method for workload whereby a deterioration in reaction time indicates the effects of an increase in workload.

C. Considerations

There are a number of considerations which should be made when determining a measurement method. As previously mentioned, considerations include sensitivity, validity, reliability and intrusion amongst others. These are important as they govern the usefulness of the subsequent data. It is also important to consider the type of data that will be produced. More specifically, whether the data is continuous or discrete as this defines the measurement method. Table 1 shows an assortment of the considerations for selecting the measurement method along with brief descriptions as defined in the literature.

<table>
<thead>
<tr>
<th>#</th>
<th>Consideration</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Sensitivity</td>
<td>Changes in workload must be measurable.</td>
</tr>
<tr>
<td>2</td>
<td>Validity</td>
<td>The data measured must be dependent on workload.</td>
</tr>
<tr>
<td>3</td>
<td>Reliability</td>
<td>Must work consistently.</td>
</tr>
<tr>
<td>4</td>
<td>Intrusion</td>
<td>Must not influence the measurement.</td>
</tr>
</tbody>
</table>

Ultimately there exists three measurement method types. As discussed in this literature review, these include subjective, performance and physiological measures. All three of these approaches are alternative means in accomplishing measurement of the workload. The term workload has been defined in a number of ways but ultimately is the relationship of cognitive load to cognitive supply. The most prevalent problem identified through the literature is the limited knowledge supporting any one method of measure over another. Table 2 has been created to display the findings of the literature review. This outlines the pros and cons of many of the discussed measurement methods which belong to one of the three measurement method types.

<table>
<thead>
<tr>
<th>Measurement Type</th>
<th>Measurement Method</th>
<th>Increasing Workload</th>
<th>Positives</th>
<th>Negatives</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subjective</td>
<td>Unidimensional (Vertical Analogue Scale (VAS))</td>
<td>Marked at a higher level along the VAS.</td>
<td>Simple to use and can be used during the experiment with minimal intrusion.</td>
<td>Only one dimension, not considered reliable.</td>
</tr>
<tr>
<td>Subjective</td>
<td>Multidimensional (NASA Task Load Index Scale)</td>
<td>Score higher on the index scale.</td>
<td>It is considered accurate and supported by the academic community as valid.</td>
<td>It is resource intensive to use and analyse.</td>
</tr>
<tr>
<td>Performance</td>
<td>Primary</td>
<td>Change in performance.</td>
<td>Valid measurement method, accurately measures changes in workload.</td>
<td>Diminishing accuracy if under low workload.</td>
</tr>
<tr>
<td>Performance</td>
<td>Secondary</td>
<td>Change in performance.</td>
<td>Provides a support measurement to the primary.</td>
<td>May have adverse effects on the primary task</td>
</tr>
<tr>
<td>Physiological</td>
<td>Heart rate</td>
<td>Heart rate will increase.</td>
<td>Literature supported and accepted.</td>
<td>Reliability can be questionable, many factors can influence heart rate.</td>
</tr>
</tbody>
</table>

D. Recommendations

Workload is a difficult term to measure. Based on the literature and the subsequent analysis, it is evident that to achieve a reliable data set through experimentation there are a number of requirements which must be satisfied. The first requirement is to include at least one of the three accepted measurement types (subjective, performance and physiological) in the experimental design. It is also recommended that if two measurement methods are selected, that one measurement be of continuous nature and the second be discrete. Most common and accepted processes include a continuous measure and a discrete measure for the performance and subjective types respectively. Through adhering to these guidelines the risk of error in results due to measurement type and method is reduced. This, based on the literature is the most reliable and validated means of measuring workload.
IV. Methodology

The experiments methodology primarily relates to four sections which include important design decisions. In this section of the paper these four sections will be discussed and explained. These sections include the ethics clearance application, the experiment population, the equipment and the experimental design. Each of these components were essential to the conduct of this study.

A. Ethics Clearance Application

The experiment required a number of considerations to enable its realization. The first and most important was the ethics clearance application through the Australian Defence Human Research Ethics Committee. This was necessary to obtain permission to conduct an experiment using human participants. The ethic clearance process involved many hours of work compiling documentation, obtaining signatures and general project management tasks. The final deliverable to the ethics board was a document exceeding 30 pages outlining the proposed experiment. It included a literature review, experimental design, risk assessment, participant rights documentation and more. The point being, the ethics board application and its approval was a significant component of this project which shaped and influenced its design and conduct.

B. Experiment Population

The experiment was conducted over a two week period. A total of 20 participants completed the experiment. The statistics for the experiment show that 85% were male and 15% female. The participants varied between 18 and 25 years of age. The population was made up of 85% right handed participants and additionally 85% engineering students. Each of the participants are students at University of New South Wales Canberra. They were each contacted via email. An expression of interest was released over the Australian Defence Force Academy webpage (Forcenet) and those that responded were contacted with an information package. Additionally a number of participants were approached in person in an attempt to make up more numbers for the experiment. These participants were also sent an email information package.

C. Equipment

The experiment primary task employed the DiddyBorg robots. These are a Raspberry Pi powered vehicle which can be seen in figure 2. These robots are capable of Wi-Fi and Bluetooth control. Each has 4 ultrasonic sensors on each corner with a 2-4cm obstacle detection range. the reason these were chosen is because they are reasonable cheap compared to other options and they are highly customizable. In addition to this, they are very well supported. This made their implementation simple and reliable.

D. Experiment Design

The primary task is the main focus of the experiment. This task involved the control of a three robot swarm. The user was required to navigate the swarm numbered 1 through 3 to visit their corresponding numbers located on the platform (figure 2). The robots were attached together with a fixed length to force simultaneous movement and in turn swarm behaviour. These characteristics can be seen in figure 2. This primary task is completed with both the input controls keyboard and gesture. For each control method the user conducts 4 run throughs each with the secondary and subjective tasks separating them. This amounts to a total of 4 keyboard completions, 4 gesture completions, 8 subjective measures and 8 secondary tasks completions for each participant. A visual diagram for the experiment sequence can be seen in annex A. Each run through is both recorded through a bird’s-eye video feed and a time stamped keypress tracker. A complete run through of the experiment takes roughly 90 minutes per participant. In addition, 20 participants conducted the experiment to ensure reliability in results. Through this design, the appropriate swarm behaviour was created and additionally, appropriate secondary and measurement tasks were performed in order to create the desired environment.
The secondary task is designed to target and induce cognitive load. The task simulates the effects of prolonged exposure to workload. This task was conducted at the completion of each primary task round. Conducting this task at intervals induces a progressive cognitive load. This allows the progressive effects of workload to be observed. The task remains constant ensuring that the cognitive assemblies targeted remain unchanged. A working memory span test was used for the secondary task. This involved a cycle of 5 sentences being read to the participant with the participant typing each out. The participant’s goal was to remember the last word from each of the sentences and recall them at the completion of the 5 sentence group. This cycle of 5 sentences followed by recall was repeated for 3 minutes before the task was completed. The measurement of the impact of this task is conducted through subjective measurement task.

The subjective measurement task enables the measurement and quantification of the participants perceived workload state. For this task the NASA TLX was adopted. As mentioned in the literature review, this method has many attractive characteristics. It provides a workload rating for each round of the primary task and has a breakdown of the contributing factors for the associated workload rating. In addition, it enables quantitative comparisons between subjective measurements for different rounds. The NASA TLX is simple to use containing a number of binary questions and scale ratings. An example NASA TLX can be seen in annex B.

V. Hypothesis

The hypothesis for the experiment is based on the findings of the literature review. The inverted U model for workload versus performance (Figure 1) suggests the expected relationship between the two terms. Therefore it is hypothesised that an increase in workload will cause an initial increase in performance followed by a subsequent decline in performance. This hypothesis is based directly off the workload performance model discussed in the literature review. Since a number of different measures are used in the experiment for performance it is important to define the term in the context of this experiment. Performance is indicated by; completion time, time between keypresses and keypress distribution.

VI. Results

In this section of the paper, the results gained through the experiment will be shown. These include results for the NASA TLX, the primary task completion time and the primary task keypress data. These are each necessary in assessing workload in human swarm interaction and furthermore to the evaluation of the experiment hypothesis which states that an increase in workload will cause an initial increase in performance followed by a subsequent decline in performance.

The first of the experimental data shown is the NASA TLX. This is displayed in figure 3. It is evident through the graph that as the experiment progressed, both groups 1 and 2 experienced a steady increase in workload. It can also be seen that beginning with gesture control significantly increases the workload experienced by the participants. The statistical significance for the data can be seen in table 3 for appropriate rounds.

![Figure 3: NASA TLX Workload Ratings](image)

<table>
<thead>
<tr>
<th>Round 1-4</th>
<th>Group 1</th>
<th>Group 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>42.95</td>
<td>55.24</td>
</tr>
<tr>
<td>t Stat</td>
<td>-7.98</td>
<td></td>
</tr>
<tr>
<td>P(T&lt;=t) two-tail</td>
<td>0.0040</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Round 5-8</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>48.48</td>
</tr>
<tr>
<td>t Stat</td>
<td>-13.93</td>
</tr>
<tr>
<td>P(T&lt;=t) two-tail</td>
<td>0.0008</td>
</tr>
</tbody>
</table>

The results for completion times indicate that the participants are improving as the experiment progresses. These results are displayed in figure 4. The graph portrays the comparison of gesture controlled completion times to keyboard controlled completion times. This shows that keyboard is on average a faster control method than gesture. Additionally, as stated, the completion time is decreasing. Furthermore the statistical significance t-test is presented in table 4, illustrating that both data sets are statistically significant.
The keypress data obtained from the primary task demonstrates a number of important relationships. Figure 5 shows the average distribution of keypresses with respect to time. The keypress distribution has been separated into 400ms bands. It is evident that for each round, the data in the figure is positively skewed. This is indicated by the long tails to the right of the graph for each round. Additionally this is shown in table 5 which displays a number of calculations required for the analysis of the graphs. These include the skewness, kurtosis and total keypresses within the first 400ms band.

The skewness indicates the distribution of the data about the mean. In round 1 the highest skewness is seen with round 4 having the lowest. Since the greater positive skewness represents the greater the grouping of the data in the left extremities of the plot, round 1 has the least average time between keypresses. This by extension means that round 4 has the greatest average time between keypresses. It is also apparent that the total total keypresses within the first 400ms is greatest in round 1 and fewest in round 4. Lastly the t-test can be seen in table 6 showing the statistical significance of the data.

**Table 5: Keypress Data**

<table>
<thead>
<tr>
<th>Round Number</th>
<th>Skewness</th>
<th>Kurtosis</th>
<th>Total Keypresses within 400ms (Average)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.43</td>
<td>3.57</td>
<td>55.50</td>
</tr>
<tr>
<td>2</td>
<td>1.35</td>
<td>3.25</td>
<td>48.57</td>
</tr>
<tr>
<td>3</td>
<td>1.39</td>
<td>3.54</td>
<td>49.21</td>
</tr>
<tr>
<td>4</td>
<td>1.29</td>
<td>3.28</td>
<td>45.06</td>
</tr>
</tbody>
</table>

**Table 6: Statistical significance (figure 5)**

<table>
<thead>
<tr>
<th>T-test Keypress Distribution</th>
<th>Round 1 vs Round 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>T stat</td>
<td>-1.86</td>
</tr>
<tr>
<td>P(T&lt;=t) two-tail</td>
<td>0.078</td>
</tr>
</tbody>
</table>
The percentage of presses for each key are shown in figure 6. The distribution is generally consistent between rounds however some of the small variations indicate changes in behaviour and subsequent improvements in performance. The variation between keypress percentages for each round are small. It is evident however that the highest percentage of keypress ‘4’ are seen in round 4. This represents the selection of all robots. Additionally the lowest percentage of keypress ‘n’ are seen in round 4. This represents the stop command to the robots. The results for the t-test can be seen in table 7 indicating the statistical significance for the data.

**Figure 6: Keypress percentages**

![Keypress Percentage Distribution](image)

**Table 7: Statistical significance (figure 6)**

<table>
<thead>
<tr>
<th>T-test Keypress Percentage</th>
<th>Round 1 vs Round 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>T stat</td>
<td>-1.68</td>
</tr>
<tr>
<td>P(T&lt;=t) two-tail</td>
<td>0.129</td>
</tr>
</tbody>
</table>

**VII. Discussion**

The paper aims to interpret the result and evaluate the hypothesis. Recall that the hypothesis stated that an increase in workload will cause an initial increase in performance followed by a subsequent decline in performance. From the results section it is evident that a change in performance has occurred as a consequence of increasing workload. This is seen in each of the sets of data presented. These include the completion time data, the keypress distribution data and the keypress percentage data.

The workload comparison plot is important as it enables the hypothesis to be evaluated. This shows that the workload on the participants is increasing between rounds (figure 3). This is essential to the analysis and evaluation of the presented hypothesis.

The graph for the primary task completion time indicates an improvement in performance. This is seen for both keyboard and gesture control (figure 4). There are likely a number of reasons for this improvement. The initial improvement which is seen between rounds 1 and 3 has the greatest positive rate of change. This is indicated by the steepness of the gradients. The initial improvement is likely due to the learning curve effect; becoming familiar with the task, controls and interactions. However, it is important to recall that the workload is continuing to increase whilst the experiment progresses through 8 total rounds of the primary task. This suggests more than a learning curve, it suggests improvement in performance is influenced positively by the increasing workload.

The results for the keypress distribution also indicate improved performance throughout the completion of the primary task. As suggested by the skewness of figure 5 shown in table 5, round 1 has the largest positive skewness and round 4 the lowest. This implies that the least time between commands occurs in round 1 and conversely the most time between commands occurs in round 4. Recall from figure 4 that the participants are completing the task faster on average in round 4 and slowest on average in round 1. These results when analysed concurrently indicate that the participants in round 1 are requiring near instant corrections to many commands which they provide. This is further supported by the total number of keypresses within the first 400ms band. These are representative of keypresses within the average person’s reaction time. Hence round 1 having the highest average in this category and being the slowest round, indicates a high rate of near instant follow up commands indicating corrections to commands. This is converse to round 4 where the command rate is slowest yet the completion time is faster on average. This indicates that the participants are thinking more about the task between commands and making more deliberate movements.
Lastly the keypress percentages also indicate improvement. To achieve better performance, in general the participants should be using the keys ‘n’ fewer times and ‘4’ more times. This is because ‘n’ indicates the stop command where stopping the least often will likely result in faster completion times and ‘4’ because this is the key for selecting all robots which allows for complex but high reward simultaneous movement of the robots. As shown in figure 6, both of these traits occur in round 4. Thereby further suggesting an increase in performance from round 1 through to round 4. Again this appears to be occurring as a result of the increasing workload.

The results indicate very strongly that the participant’s performance is improving as the experiment progresses. Recall that the hypothesis stated that an increase in workload will cause performance to initially improve followed by a subsequent decline in performance. The results show the initial improvement but no succeeding deterioration in performance. This increase in performance was suggested by the model discussed in the literature review seen in figure 1. However the model also states that increasing workload can cause a decrease in performance if it surpasses into the overload threshold. Evidently, this decrease in performance is not seen through the results of the experiment.

The cognitive workload vs performance model describes the improvement in performance seen in the results. At the beginning of the experiment when the participants are under very little workload they perform poorly. This is seen through their completion times in round 1 compared to in round 8 (figure 4). Additionally, the keypress data presented indicates more mistakes and less consideration for strategy in round 1 than in round 4. This is shown in the analysis of figure 5 and the analysis of the time between keypresses and total number of keypresses in the first 400ms band. Furthermore, the keypress percentages in figure 6 suggests that participants are performing more complex commands through simultaneous movement of the robots in round 4 than in any other round.

Ultimately, the model supports these findings. It is evident that the participant’s performance is improving between rounds. This represents the transition of the participants between the underload range of figure 1 to the acceptable range. What the data hasn’t illustrated is the transition from the acceptable range to the overloaded range. The reason for this is the limitation of the length of the experiment. The experiment in its current form took on average 90mins to complete per person. As a result, the participant’s workload never reached the overload stage and thus there was no deterioration in performance observed. To reach this stage of workload, the experiment would need to be extended. Unfortunately, this was not within the scope of the experiment and could not be achieved. However this fall in performance would likely have occurred had the experiment been extended. As a result the hypothesis is partially supported by the results. The initial improvement in performance has been supported, the succeeding decline in performance resulting from workload overload is inconclusive.

VIII. Conclusion

The aim of the paper was to assess workload in human swarm interaction. In addition the paper has discussed workload and its measure through a literature review. The paper explained the methodology behind the experiment and justified many of the design decisions. The hypothesis which stated that an increase in workload will cause an initial increase in performance followed by a subsequent decline in performance was evaluated. It was found through the experiment that the results supported the hypothesis partially. The initial improvement in performance was observed, however the subsequent decrease in performance was not. This was a result of constraints on the experiment and the inability to induce the workload required to decrease the performance. Ultimately the results showed that the average completion times improved, the average time between keypresses increased (indicating more thinking and deliberate commands) and that the keypress distribution percentages changed to a more appropriate keypress distribution for success. All of this occurred as the workload of the individuals was increased.

This paper has identified that workload can be a very positive force when considering performance. In the experiment that was conducted, increasing workload had a significant positive affect on performance. It is noted however, that increasing workload beyond a threshold (which was not reached in this experiment) would likely be detrimental to performance. Equally, workload being too low can have negative effects on performance as shown through this paper. It is therefore evident that the inverted U model holds true when assessing workload in human swarm interaction.

IX. Future Work

As stated in the hypothesis, increasing workload will at its extreme cause a deterioration in performance. This was not evaluated through the experiment as the design did not induce a high enough workload to cause this effect. Therefore, future work on this project could investigate the higher ends of induced workload and establish whether this component of the hypothesis is supported.
References


Annex A

Experiment Flow Chart:

BT = Baseline Task
TT = Training Task
PT = Primary Task
AT = Auxiliary/Secondary Task
Orange = Keyboard Control
Green = Gesture Control
Annex B

NASA TLX spreadsheet:

The spreadsheet below was used to calculate the workload for each participant. This is a mock example of how it was filled out. Ultimately the NASA TLX enabled the workload to be quantitatively measured.

<table>
<thead>
<tr>
<th>Participant ID</th>
<th>Task ID</th>
<th>2</th>
<th>Round</th>
<th>Source of Workload Tally Sheet</th>
<th>Weighted Rating Worksheet</th>
</tr>
</thead>
<tbody>
<tr>
<td>Composition Code</td>
<td>1 or 2</td>
<td>Rating Sheet</td>
<td>Score (0-100)</td>
<td>Scale Title</td>
<td>Weight</td>
</tr>
<tr>
<td>Effort</td>
<td>1</td>
<td>Mental Demand</td>
<td>80</td>
<td>Mental Demand</td>
<td>4</td>
</tr>
<tr>
<td>Temporal Demand</td>
<td>2</td>
<td>Physical Demand</td>
<td>20</td>
<td>Physical Demand</td>
<td>0</td>
</tr>
<tr>
<td>Temporal Demand</td>
<td>1</td>
<td>Temporal Demand</td>
<td>60</td>
<td>Temporal Demand</td>
<td>0</td>
</tr>
<tr>
<td>Performance or Frustration</td>
<td>2</td>
<td>Performance</td>
<td>65</td>
<td>Performance</td>
<td>1</td>
</tr>
<tr>
<td>Temporal Demand</td>
<td>2</td>
<td>Frustration</td>
<td>70</td>
<td>Frustration</td>
<td>5</td>
</tr>
<tr>
<td>Total</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Final Project Paper 2016, UNSW Canberra at ADFA