A usability perspective into gesture for human-swarm teaming

Adam N. Rogers*

University of New South Wales at the Australian Defence Force Academy

From a very young age, before we can walk or talk, humans communicate non-verbally by using body language, also known as gestures. Studies have shown that young children can readily learn how to communicate with gestures before they can talk [5]. This thesis will look into how effective gestures are at controlling complex tasks such as swarm motion over the traditional keyboard control method. Swarm robotics is an emerging field which aims to use lots of smaller robots that can cover a larger area then one large robot. The challenge with this is that traditional methods such as a keyboard might not be suitable to control a swarm. As swarm motion can be very complex, using a more intuitive way to control the swarm will result in increased efficiency of the swarm [14].

Contents

I Introduction 2

II Literature Review 2
   A Gestures vs. Poses 2
   B Gestures 2
   C Opportunities and challenges of Gestures 4

III Methodology 4
   A DiddyBorg Robots 5
   B Choice of Input 5
   C Gesture Design 6

IV Experimental Design 7
   A Hypothesis 7
   B Primary Task 8
   C Secondary Task 8

V Results 8
   A Completion Time 9
   B Participant Perception 10
   C Researcher Notes 10

VI Conclusion 11

VII Future Work 11

*PLTOFF, School of Engineering and Information Technology, ZEIT4501 - 17 October 2016
I. Introduction

Increasing the performance of the Human-Machine interaction can have some life savings effects which can be seen in the use of Search and Rescue Robots (SARR). During the 2001 September 11 attacks on the World Trade Center, SARR were used to find casualties in a mass casualty incident resulting in the Centre for Robotic Assisted Search and Rescue (CRASAR) deploying their SARR. A report into the Human-Robot interactions during the events of September 11 found that the Human-Computer interface and robot systems needed to support people working without sleep. This was due to the operators going in-excess of 50 hours without sleep [7]. Finding a better way to control these robots that is less effected by fatigue should be top priority.

In a mass casualty event such at September 11, the SARR operators will be working as long and as hard as possible to save as many lives as they can. Two variables that effect this are speed and fatigue, the rescuers needs to be fast but also mitigate fatigue to stay effective for a long period of time. This can all be done by making a natural control system which will lower the effects of fatigue and increase effectiveness. This thesis will be looking at how this could be done by investigating gestures as a way to control SARR.

Gestures are used everywhere as a way of communication, it is seen in body language, used as a language for those people who can’t talk and is even seen in the animal kingdom where chimpanzees have been observed using gestures to communicate [20]. This is because gestures are a very natural way to communicate which is why it’s our primary form of communication before we can speak [3]. This form of natural communication is exactly what is needed when controlling a swarm of robots.

II. Literature Review

A. Gestures vs. Poses

This thesis defines a gestures as “a particular movement of the body, typically the fingers or hand, used to control or interact with a digital device” [10]. Using this definition, a gesture is any movement of the body, static or dynamic, that has some communicative information associated with it. Posing is defined as “Assuming a particular position” [9] which is a static discrete act and doesn’t convey language like a gesture does. A pose becomes a gesture when it is used to communicate some sort of information either to a digital device or another living thing.

B. Gestures

Over the past 30 years gesture control has come a very long way. It all started with Bolt [6] who used a wrist mounted sensor and microphone to detect gestures and voice commands that created and moved objects around a large screen. In recent years this technology has made it into the mainstream gaming community with the introduction of the Nintendo Wii and Sony’s PlayStation EyeToy. The Wii and PlayStation were able to make the technology available to the public in a way that was cheap and enjoyable. Before this, to capture gestures, the user had to wear gloves that were directly connected to a computer via wires or use very expensive inferred projectors and still cameras to capture the gesture [5]. This meant research was very limited due to cost and portability of the technology. Recent developments in image processing and camera hardware has made it possible for technology like the Microsoft xBox Kinect to become available to the public for the use in gaming as well as research. This has enabled the research community to expand their research into other areas that weren’t before possible such as health care or helping the elderly [17]. This research will help to improve the quality of life for those with disabilities by using the technology to do such activities as controlling home lighting. This is important as the population in Australia is getting older so systems will need to be designed to help these people [24]. A real world example of this technology being used now is a company called Gestsure who have designed a hands free solution for doctors to navigate MRI or CT scans without the use of a keyboard or mouse. This keeps the environment sterile and saves time, allowing the surgeon to become less fatigued and focus on the surgery [11].

Today’s gesture detection systems use a wide variety of sensors to gather data which is used to interpret the users gestures. These sensors include gloves embedded with microcontrollers, inferred projectors
and cameras, gyroscopes, HD video cameras or a mix of all of these technologies. The most common one due to its accuracy and cost is an infrared projector and camera system such as the one found in the Microsoft xBox Kinect. The infrared projector illuminates the object which is detected by the camera. This type of camera allows for the depth of the object to be measured in real time, a real advantage when creating gestures. Being able to map an object in 3D allows for more complex motions to be detected and analysed.

**Figure 1. Australian Defence Force Aircraft Move Ahead Signal**

Gestures have been used as a form of Human-Human communication for many years. Two good examples of this are aircraft marshalling and sign language. Aircraft marshalling is used in an environment where communicating in traditional means like verbally can’t be used. The noise of the aircraft means that it’s impossible for the pilot to hear what the marshaller is saying. Looking at the Australian Defence Force manual [23] for marshalling gestures it can be seen that the code book is made up of 44 different discrete and continuous gestures. One of the common themes throughout all the gestures is that they are all big movements and are intuitive. One gesture that displays this is the ‘Move Ahead’ gesture seen in Figure 1.

Sign language, just like aircraft marshalling, is a very effective way of communication when a person is unable to hear [4]. This shows that sign language which uses gestures to communicate is capable of keeping up with the human races main form of communication. This is because interpretation of gestures is a very natural thing for humans to do. A 2007 study showed that listeners rely on speakers’ gestures to disambiguate communicative intent where understanding may be impeded [30]. This points out how much humans rely on body language to understand the meaning of what is being said even when we aren’t thinking it. The reason why humans can understand what is being said even when the environment is too loud to hear is that gestures are processed as linguistic information rather than spacial information. Susan Goldin-Meadow from the University of Chicago gives the example of when deaf children whose hearing losses prevent them from acquiring spoken language even with intensive oral instruction, and whose hearing parents have not yet exposed them to sign language. Gesture is the primary means by which children in this situation communicate. The question is whether their gestures assume the language-like forms characteristic of a codified communication system like Americal Sign Language. The answer is that they do on all levels that have been examined thus far [12].

Once common way to detect this form of communication is with a camera system but as with all systems they come with advantages and disadvantages. A camera based system has a lot of benefits over other systems like microcontroller gloves or wrist mounted sensors. Some of the advantages identified in [29] are accessing information while maintaining total sterility, overcoming physical handicaps and exploring big data sets. These three advantages are particularly useful in the medical practice when doctors need to be sterile while operating on their patients. It will help them navigate through computers so when they are in the operation room the surgeon can look at the patients 3D MRIs without using a keyboard which would contaminate their hands. Outside the medical area, camera based gesture systems could be used to explore large and complex data volumes in a more intuitive way.

A camera based system also allows the user to do things that would otherwise be limited by other technologies such as wrist or glove mounted sensors. The users is not restricted by wires connected to a computer or a large sensor mounted on the body or wrist allowing them to do tasks that aren’t possible with these sensors. It also makes the system very portable as it can be embedded into the computer/laptop making it very rugged and capable of being used for search and rescue without worry of damage.

Other advantages of visual based hand gesture systems is with Human-Robot interaction. The idea behind a visual based gesture system is to have it resemble natural human dialogue as much as possible. Imagine if you could tell your personal cleaning robot to clean up a area by just pointing where to clean. Gestures also provide a source of expressiveness and immersion that can’t be found with other technologies[29].
One of the benefits of using a visual gesture based system over a traditional keyboard system is that it can extract biological data from the user over time rather than over discrete time periods. One area where this is currently being used is the early detection of Parkinson’s disease [27] by measuring the tremors of the users. Being able to detect the dominant frequency of Parkinson’s disease early on means the early diagnosis and therapy can be given to help control the tremors [15].

C. Opportunities and challenges of Gestures

Using gestures as a control device has a few advantages over some of the traditional keyboard and mouse controllers as outlined in [25]. Their paper employed a stylus on a tablet to look at handwriting and written gestures to perform common tasks that could be done with a standard keyboard and mouse. Their aim was to identify the human factors involved with the interface and look at the advantages/disadvantages of a stylus over a keyboard and mouse. The advantages identified in this paper were a single gesture can be equivalent to many keystrokes and mouse actions as well as the interface is silent. While some disadvantages are that handwriting is 2-5 times slower than keyboard for entering text and gestural interfaces will be more expensive than keyboard/mouse interfaces.

They found that using a stylus with gestures out preformed the keyboard when performing tasks like writing sentences as well as using gesture to input commands. Some gesture commands that were used were sum, move and copy. The biggest problem they faced was correctly identifying the correct gesture which is difficult, especially with handwriting, as everybody has their own style so the gestures needed to adapt. It was also difficult to determine the closure of the gesture as it might be a single continuous line or multiple lines that make up the gesture but at the same time a single small difference could mean a big difference in the gesture. The key thing that helped the system identify the gesture was that it gathered context about what was happening on the screen. This helped improve the accuracy of the gestures being performed.

Another big problem identified with gestures in [21] was the need for gestures that could be accurately identified without the user having to repeat the gesture. This is because if the person can’t trust that the gesture will be detected and recognised correctly they will be hesitant to use that gesture. It was found in a test that if the gesture can’t be recognised close to 100% of the time then user satisfaction will be low. This means the user will be less likely to use the gesture if there are alternative ways to do the same operation. It was also noted that if the training was too laborious the user would also be more likely to abandon the gesture. With a simple set of eight gestures that was used to control a DVD player an average accuracy of 97% was achieved over a sample size of 240.

To increase accuracy of the gesture to achieve close to 100% accuracy user defined gestures should be used [18][27]. A user defined gesture could be a completely new gesture or one that can be altered to better suit the user. By giving the user the ability to create or alter a gesture an accuracy of 96.67% was seen in the study by [27]. This accuracy included complicated gestures that were a string of up to three discrete gestures. It was also noticed that using a combination of discrete gestures actually increased accuracy from 95% to 96.67%. This is an huge positive for the use of gestures as being limited to single discrete gestures really limits the amount of activities that can be accomplished with gestures. Having the ability to string together discrete gestures and improve accuracy means the amount of activities gestures can do greatly increases.

III. Methodology

To validate the hypothesis a small experiment was conducted that looked into the differences between discrete gestures and keyboard. The basic design of the experiment required 20-30 participants to conduct a primary and secondary task. The primary task compared discrete gestures against keyboard which is what this thesis is about while the secondary task investigated fatigue detection using keystroke dynamics. Comparing Gestures to keyboard was chosen due to the gesture set and keyboard both being discrete actions.
A. DiddyBorg Robots

The DiddyBorg robot is the main vehicle that will be used for this thesis. The DiddyBorg robot is a six wheeled Raspberry Pi powered vehicle that is capable of being controlled over WiFi or Bluetooth. The robot is fitted with a front facing camera and four HC-SR04 ultrasonic sensors that can range from 2cm to 4m. The ultrasonic sensors are used for obstacle avoidance and will prevent the robot moving in a direction where it could potentially damage itself. Figure 2 shows the DiddyBorg with ultrasonic sensors attached.

B. Choice of Input

There are many different types of inputs that can be used for this project including: Video cameras, 3D mapping and microcontroller gloves. In the early stage of gesture research, gloves with microcontrollers were used due to limitations in available technology. Now days with advancements in 3D scanning and signal processing their is no need for the user to be wearing special gear for gesture detection. Due to thesis goals and limitations the following criteria were used to select the input device; Availability, Cost, Setup, Data Streams and Programmability.

Availability: The input device needs to be readily available for a few reasons. This thesis has a time limit of eight months which in includes all the coding and testing meaning it needs to be off the shelf and available in Australia. Also it shouldn’t be specialised equipment where there are only a few available in the world. This is so the project can be expanded if warranted.

Cost: As with the availability this project doesn’t have unlimited funding so costs need to be kept as low as possible. Not only do the costs need to be low, one of the thesis goals was to make it as practical as possible and not just some research that is too expensive to use in real life. For gestures to be a viable option over keyboard it needs to be able to be mass produced like the keyboard.

Setup: The system needs to be easy to setup with the device being able to be used in multiple different environments. This means the system can be used outside in the real world. The device should be portable i.e. it shouldn’t need a ‘green screen’ to detect the person.

Data Streams: Most input devices these days have multiple different output streams, for example the Microsoft Kinect V2 can output a 16-point skeleton of the person in front of the camera. The ability to have different data streams and on board signal processing improves the functionality of the camera.

Programmability: The device needs to be easily programmed which means it needs to come with a SDK and be able to be programmed on many different languages. Having a SDK for a particular programming language adds a lot of the functionality, for example using the kinect V2 SDK means you can extract the skeleton data with outdoing the classification yourself.

Looking at Table 1 the Microsoft Kinect V2 is the obvious choice for this thesis. The Kinect is available in Australia, easy to setup, has seven different data streams and can be programmed on multiple different languages including C++ and C#. The Kinect is the most expensive out of the five input devices listed here but has by far the most functionality. The Microsoft Kinect uses a method called indirect time-of-flight (TOF) to calculate distances to objects [19]. Indirect TOF works by using the phase difference between the reference and measured signals. When light travels a distance, $d$ and reflects off an object, the reflected light is delayed by phase $\omega$. The phase difference and the frequency of the light is then used with Eq. 1 to calculate the distance [16].

$$d = \frac{\Delta \omega}{4\pi f} \cdot C \tag{1}$$
Table 1. Input Selection Data Sheet

<table>
<thead>
<tr>
<th>Device</th>
<th>Available</th>
<th>Cost</th>
<th>Setup</th>
<th>Data Streams</th>
<th>Programmability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kinect V2</td>
<td>Yes</td>
<td>$204.90</td>
<td>Easy</td>
<td>Colour, Depth, Full Body Skeleton (16-Point), Body Index, Infra-red and Audio Array</td>
<td>C++, C#, JavaScript NB: SDK Provided</td>
</tr>
<tr>
<td>Intel SR300</td>
<td>No (In Pre-Order)</td>
<td>$131.15</td>
<td>Easy</td>
<td>Colour, Depth (Short Range), Infra-red, Skeletal hand, Audio</td>
<td>C++, C#, JavaScript, Java (JDK 1.7.0.11 or later) NB: SDK Provided</td>
</tr>
<tr>
<td>Leap Motion</td>
<td>Yes</td>
<td>$144.25</td>
<td>Easy</td>
<td>Skeletal hand</td>
<td>C++, C#, Unity, Objective-C, Java, Python, JavaScript, Unreal Engine NB: SDK provided</td>
</tr>
<tr>
<td>USB Web Camera</td>
<td>Yes</td>
<td>$39-$179</td>
<td>Hard</td>
<td>Colour</td>
<td>C++, C#, MATLAB, Java, JavaScript, Python NB: No SDK</td>
</tr>
</tbody>
</table>

C. Gesture Design

The creation of the gesture criteria started with a review of the literature to gather information about what makes a good gesture [2] [8] [22] [28] [29]. This information was then collated with four criteria being developed. These criteria are:

**Gross Motor Movements:** The gestures need to use gross motor movements. This is a limitation due to the Microsoft Kinect V2 [22] but also to reduce the effect of fatigue on the user. Unlike fine motor skills, gross motor skills are less effected by fatigue. So using gross motor movements will be more accurate once fatigue sets in [2] [8].

**Learnability—Easy to learn:** Gestures need to be easy to learn otherwise the users will get frustrated with the gesture and wont use it. This means the gesture needs to be simple and avoid complex movements that aren’t natural. The challenge with learnability is that the learning rate depends on the task, user experience and user cognitive skills [29]. A possible solution for this is to use user defined gestures or adopt gestures that are natural and intuitive.

**Intuitive—Easy to use:** The gesture set needs to have a clear cognitive association with the action. For example, moving your hand left can represent “turn left” or an open palm could mean “Stop”. If the actions are unnatural or represent complex movements the user will be reluctant to use them [29]. The biggest challenge with this is that intuitiveness is strongly related to cultural background. This means what is intuitive to me wont necessarily be to you [28].

**Low mental load—Easy to remember:** Having to recalling gestures will add to the users mental load so having this as low as possible is desirable. Doing this will allow more cognitive capacity to be used on the task which will in turn improve the quality of the task [29]. The challenge with this is to make the gestures as simple and natural as possible but also maintain the complex actions that are required for complex tasks such as swarm motion.

The gestures for this thesis were designed only using the gross motor movements and learnability criteria excluding intuitive and low mental load. This was due to intuitive and low mental load being difficult to measure so fell outside the scope of this thesis. The right hand was chosen to control the direction of the robots due to the function of the left side of the brain. This half of the brain is used for the functions of logic, language and reasoning so everything known for controlling [26]. As well as this 90% of the population is right handed [13] so the use of the right hand for control is perfect. As the right hand was being used control the left hand is used for the selection of the robot.
Figure 3 shows the gesture set used for this thesis and experiment. The gestures work by using the left and right shoulder as the origin for the left and right hand respectively. When the wrist of each hand moves into a zone as seen in Figure 3 it activates the respective discrete action. The gestures were chosen so that they were easy to learn and are gross motor movements. This was done by having the zones near the maximum reach of each person’s arm, so to activate Robot one the left hand needs to be stretched straight up to select robot one. The gestures were also designed so that when in the natural position—hand over the axis origin—all the robots will stop, this is done so if physical fatigue sets in, all the robots will stop preventing the user from doing something they didn’t want to do.

Figure 4 shows the block diagram of the gesture detection system. The system takes in the gesture from the user, a gesture could be moving the right wrist right parallel to the ground out as far as possible. While this is happening the Kinect V2 is continuously tracking the gesture and sending the data to the classifier. The classifier is a simple pose detection system but to the user the full gesture is being detected and hence all the properties of gestures are applicable. Once the wrist enters the zone the gesture is detected and the specific command is sent to the robot, in this example spin right is sent. As such the accuracy of this system is 100% due to the system always detecting the wrist location. The dead zones for the left hand were designed to reduce the sensitivity of detection so that the user couldn’t accidentally select the wrong robot. This wasn’t done to the right hand as in some situations it is preferable to quickly give a Backwards or Forwards command straight after spinning the robot left or right.

IV. Experimental Design

The participants were split into two groups, the first starting with keyboard then gesture while the second group would do the opposite. This was done so not to skew the data as it was expected that after a few runs the participant would start to learn the fastest way around as well as to mitigate fatigue due to the secondary task. This section will explain what the primary and secondary tasks were and how they were run.

A. Hypothesis

“The use of gestures is a more effective way to control a swarm”
B. Primary Task

The primary task is designed to make the participants operate three robots as a swarm to complete a task. This was achieved by connecting the three DiddyBorg by a length of rope suspended above such that it doesn’t fall in front of the ultrasonic sensors. The objective of the task is to move the three DiddyBorg robots around the arena seen in Figure 5 and visit each of the checkpoints in any order with the corresponding robot. Once the robot had visited its check point it was allowed to leave. The position of the check point and length of rope were designed so all three check points couldn’t be reached at the same time. Once again this was done to force the participant to operate the robots as a swarm rather then as singular entities. Each participant was given five minutes to complete the task as fast as possible with the completion time and all gestures/keypresses recorded for analysis.

Before the start of the first primary task the participants were given a training task that lasted five minutes. The training task involved the participants controlling the swarm around the arena with the keyboard and gestures for two and a half minutes each. This was designed to get the participants familiar with the keyboard layout and the feel of the gestures. The participants were limited to two and a half minutes so they could get to know the controls but not the task as this would skew the results. Additionally, at the end of the gesture training participants were asked to perform a set of gestures as a test of objectives with feedback given to help improve how they used the gestures.

At the completion of the experiment the participants were given a questionnaire to get subjective data about how the participant perceived the experiment. They were also asked questions about how the experiment, particularly the gestures, could be improved.

C. Secondary Task

The secondary task was designed to introduce fatigue into the participant by cognitive loading their memory. This was achieved by the participant typing out five sentences into a computer while remembering the last word of each sentence. At the end of the 5 sentences the participant has to recall the last word of each sentence and type them back into the computer in any order. The participant was given three minutes to complete this as many times as possible. As this task would be done after each primary task the effect on the primary task would be constant for each participant.

V. Results

During the conduct of the experiment completion time, skeletal position, participant perception and researcher notes were recorded for data analysis. The experiment consisted of a sample size of 20 participants who were split into two groups of ten. The participants were drawn from undergraduates currently studying at the University of New South Wales, Canberra. The sample population was made up of 85% engineering students with a handedness split of 85% right handed and all participants being aged between 18-25 years old. The participants were recruited through an expression of interest to the university as well as through word of mouth to students that attended the same class as the principle investigators.

This section will investigate three pieces of data extracted from the experiment; Completion Time, Participant Perception and Researcher Notes. Every primary task was timed to get an average completion time for each group, separated into run number. This was done to determine if there is a speed difference between the two modes of control. At the completion of the experiment the participants were given a questionnaire that contained questions about their perception of keyboard and gesture at the start, middle and end of the task. This data was compiled to gain a perspective into the minds of the participants. The final piece of data analysis was notes from the researchers conducting the experiment. This was done to get an insight into the experiment and help explain the reason behind the results.

School of Engineering and Information Technology, UNSW Canberra at ADFA
A. Completion Time

Figure 6 shows the average completion time for each run with the corresponding T Test data shown in Table 2. Examining group one it can be seen that there is little significant statistical difference between the average keyboard and gesture times. This is seen by examining both the T Test results and looking at Figure 6 where the probability of rejecting the null hypothesis didn’t reach the 95% confidence level to be statistically significant. Group one started with keyboard then went onto gesture with the reverse being down with group two.

Examining group two’s T Test data in Table 2 and Figure 6 a significant statistical difference between the average completion time of keyboard and gesture was seen in run number one and two with 97.15% and 97.39% probability that the null hypothesis is rejected. Once the participant got to run three and four their was low significant statistical probability to reject the null hypothesis. One thing of note with the group two’s data is 50% of participants did not complete the first run through using the gestures. Though the probability of a difference is statistically significant for run number one the amount of incomplete data at the beginning of the test is enough to skew the data towards keyboard. Looking at the recordings of the participants one of the main reasons for participants failing to complete the task was due to the tactics being used and not the input device. The choice of input might have an effect on tactic but this was not tested so no data exists to prove this theory.

Figure 6. Completion time for each group over the 4 runs

Table 2. T Test P-values that the data rejects the null hypothesis: the means are the same, with corresponding degrees of freedom

<table>
<thead>
<tr>
<th></th>
<th>Run # 1 DF</th>
<th>Run # 2 DF</th>
<th>Run # 3 DF</th>
<th>Run # 4 DF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>13.64%</td>
<td>41.69%</td>
<td>94.3%</td>
<td>61.94%</td>
</tr>
<tr>
<td>Group 2</td>
<td>97.15%</td>
<td>97.39%</td>
<td>17.2%</td>
<td>76.24%</td>
</tr>
</tbody>
</table>

Considering the conflicting data between group one and group two the only conclusion is that the results are inconclusive to determine if gestures or keyboard is faster at completing the task. The reasoning for this is that group one’s data was inconclusive to determine if gestures or keyboard was faster, while group two’s first two runs were statistically significant that keyboard was faster then gesture. This changed for the final two runs for group two where the results were statistically inconclusive. One thing that can be determined from this data is that the learning rate for the gestures is very fast but learning the best tactic along with learning how to use the gestures at the same time is hard. This suggests a problem with the
experiment having a heavy reliance on the learning task that over shadowed the differences between keyboard and gestures. A longer experiment should have been conducted to get definitive results.

B. Participant Perception

Figure 7 shows the data recorded from a questionnaire that was given to the participants at the end of the experiment. It can be seen at the start of the experiment the participants found the keyboard easier then gestures. This is due to the participants being familiar with the operation of the keyboard while being the first time they had used gestures to control anything. By the end of the experiment, both keyboard and gesture became easy with gesture having the biggest change. In fact more people found that gesture were easier then keyboard at the end of the experiment even though the timing data may suggests otherwise. This is a positive sign for gestures as once the participants learnt the gestures they found it easier to use then keyboard. Even if gestures are slightly slower, being easier to user means the participant will be less fatigued due to lower cognitive loading.

C. Researcher Notes

It was noted during conversation at the start of the experiment the participants felt that gestures would be inferior to keyboard due to their experience with keyboard throughout their high school and university studies. They believed that their familiarity with the keyboard would be a huge bonus that the gestures couldn’t compete with. On completion of the experiment a lot of their minds had changed and were surprised how well the gestures worked. This was backed up in the participant perception data where gestures outscored keyboard on ease of use.

Two unforeseen issues with the system were seen during the experiment, these were with the gesture and ultrasonic sensors sensitivity. It was observed, in particular with shorter people, gestures became very sensitive which required them to be more careful with the gestures. This effected these participants at the start of the experiment making it harder to learn the system. It was noticed these participants tended to make more mistakes while learning the gestures then the general population of participants. As such the participants lost trust in the system as they couldn’t be confident the robot would do what they wanted it to do. A potential fix is to have adjustable gesture zones to better suit short participants. This could be achieved automatically by using a non-linear function that sets the zones or a GUI that the user can change them selves.

The other issue was the ultrasonic sensors that would be very sensitive some of the time but not at others. This was caused by the angle of incident at which the robot approaches the obstacle causing the robot to think it’s further away from an object then it is. This caused the user to lose trust in the system as they couldn’t be confident that the obstacle would be detected which could lead the robot getting into sensor lock. Once a robot was sensor locked the only way to get out was for one of the supervisors to manually move the robot. As well as this it was hard to judge what sensors were triggered resulting in lost time moving robots that didn’t need to be moved. A simple solution for this would be to have a visual cue that would
tell the user when the sensor had been triggered in the form of a led on the sensor as well as increasing the number of ultrasonic sensors from four to eight.

VI. Conclusion

Concluding the data analysis, it was seen that there is no statistical difference between keyboard and gesture in group one while this is a statistical difference that keyboard was faster for run number one and two in group two when the invalid data was excluded for the first two runs. This changes for the last two where there is no statistical difference between keyboard and gesture to determine which input method was faster. As such by the end of the experiment for both group one and two their was no statistical difference between keyboard and gesture for mean completion time.

Examining the participant questionnaire, by the end of the experiment more participants perceived the gesture to be easier then the keyboard. This is what is expected if the gestures were designed correctly using the criteria of Gross Motor Movements, learnability, intuitive and low mental load. If this were not the case it would be expected that the participants would find the gestures hard to use. This goes towards confirming the hypothesis as more participants found the gestures easy then the keyboard.

Due to the timing data being inconclusive, the participants finding the gestures easier and that gestures have the benefit of gross motor movements the conclusion from the experiment is that the data validates the hypothesis of “The use of gestures is a more effective way to control a swarm”.

This thesis feeds into of a wider study called trusted autonomy which is looking into the interactions between humans and machines and how trust can be built between the two [1]. It is hoped that this thesis will help the research into this field as well as into the field of controlling search and rescue robots.

VII. Future Work

During the conduct of the experiment it was found that there was conflicting data between group one and two so no statistical difference could be see between that average completion time. It was noted in the researcher notes that this could be due to the participant still learning the gesture controls but their could also be no difference. Extending the experiment so that more primary tasks are conducted will extend the knowledge about gestures and the differences between it and keyboard. An extension to the experiment will also be able to investigate to what extent physical fatigue effects gestures and where the limit is. A study of engagement may also bring up some surprising results.

One interesting observation that came out of the data is the learning rate of the participant. An investigative experiment to see how fast the participants learn to use the gestures would be interesting. It would be a real benefit to see how fast somebody with out any training can learn gestures. If only experienced operators could benefit from gestures then it might not beneficial to use gestures as it would take too much training.

Finally, research into continuous gesture control where the distance the hand moves relates to how fast the robot moves. This will give the user more control over the robot and how it navigates the arena. A system such as this could be tested against a joystick which is also a commonly used control system to operate robots.

Acknowledgements

Firstly I would acknowledge my supervisor, Prof. Hussein A. Abbass, who with out his guidance and expertise would have not made this project possible. You have been a great help and have supported all my ideas as well as kept me on track. I would also like to thank Hung Nguyen for designing his robot control system that I used to conduct my experiment. Also Pilot Officer Samuel Clark and Pilot Officer James Hayes for helping with the design and conduct of our experiment.
References


