

# **Metrics to Evaluate Human Teaching Engagement from a Robot's Point of View**

**Ori Novanda**

Submitted to the University of Hertfordshire  
in partial fulfilment of the requirements of the degree of

**Doctor of Philosophy**

February 2017

## Abstract

This thesis was motivated by a study of how robots can be taught by humans, with an emphasis on allowing persons without programming skills to teach robots. The focus of this thesis was to investigate what criteria could or should be used by a robot to evaluate whether a human teacher is (or potentially could be) a good teacher in robot learning by demonstration. In effect, choosing the teacher that can maximize the benefit to the robot using learning by imitation/demonstration.

The study approached this topic by taking a technology snapshot in time to see if a representative example of research laboratory robot technology is capable of assessing teaching quality. With this snapshot, this study evaluated how humans observe teaching quality to attempt to establish measurement metrics that can be transferred as rules or algorithms that are beneficial from a robot's point of view.

To evaluate teaching quality, the study looked at the teacher-student relationship from a human-human interaction perspective. Two factors were considered important in defining a good teacher: *engagement* and *immediacy*. The study gathered more literature reviews relating to further detailed elements of engagement and immediacy. The study also tried to link physical effort as a possible metric that could be used to measure the level of engagement of the teachers.

An investigatory experiment was conducted to evaluate which modality the participants prefer to employ in teaching a robot if the robot can be taught using voice, gesture demonstration, or physical manipulation. The findings from this experiment suggested that the participants appeared to have no preference in terms of human effort for completing the task. However, there was a significant difference in human enjoyment preferences of input modality and a marginal difference in the robot's perceived ability to imitate.

A main experiment was conducted to study the detailed elements that might be used by a robot in identifying a "good" teacher. The main experiment was conducted in two sub-experiments. The first part recorded the teacher's activities and the second part analysed how humans evaluate the perception of engagement when assessing another human teaching a robot. The results from the main experiment suggested that in human teaching of a robot (human-robot interaction), humans (the evaluators) also look for some immediacy cues that happen in human-human interaction for evaluating the engagement.

## Acknowledgements

It has been a long journey ever since I stepped into the University of Hertfordshire. Sometimes people say that all journeys have all its “ups and downs”. If I were to describe my journey, it has ups and downs, rounds and bends, intersections and junctions, whirling and many loop-the-loops, and obviously not to forget digging underground to the core of the Earth and flying back up with a helicopter. I haven’t even mentioned climbing every mountain and swimming across the depths of the ocean. You get the gist—it hasn’t been very easy. This journey has been challenging and difficult, and I’ve even considered stopping a few times. However, I got there in the end, and as headache-inducing my journey was, the rewards are worth the effort.

I would like to thank Kerstin for all the guidance, encouragement and critique she has provided me. Joe, for helping me to put all abstract ideas inside my head into words and to put the words into a human-readable form. Maha, for the greatly-appreciated critical analysis of my paper, and Mick, for supporting me in my studies and giving me the opportunity to be a visiting lecturer. I’d like to thank them all for their patience as my supervisors—I know I hadn’t been all bright and delightful throughout. My colleagues, Ben and Luke, whom I am greatly thankful for their personal help, especially outside of academic matters. Dag, for enlightening me when I was stuck with statistics. I’d also like to thank all my friends in the lab: Abu, Adelin, Alessandra, Frank, Kheng Lee, and Marcus; for being exceptional and supportive lab mates, and all persons who participated in the experiments on this study. Not to forget the University of Sumatra Utara and General Directorate of Higher Education of Ministry of Education and Culture of Indonesia, who granted me my scholarship and has allowed me to travel this far to see a broader vision of the world.

Last, but not least, I’d like to thank my family. My wife, Tika, for her eternal patience, when I did not write even though she nagged me to, encouraged me when I felt like discontinuing, and her home-cooked meals that fuelled me through nights that I had to work. My two awesome daughters, Rachel and Denisa, who greatly helped me when I’m clueless as whether to use past or present tenses, as well as being beta-testers for my experiments. And of course, my two noisy can’t-sit-still sons, Kevin and Cedric, for trying to sit down as long as possible and being quiet as they can as I worked. Well, it didn’t last very long, but they tried nevertheless.

To those names mention above, I thank you all greatly for your help and your support. You have all helped me throughout this journey. As I said, it had its ups and downs and loop-the-loops, but you have all made it possible for me to finish this journey. That, I am very thankful for. Alhamdulillah.

# Table of Contents

Abstract.....	i
Acknowledgements.....	ii
Table of Contents.....	iv
List of Tables .....	x
List of Figures .....	xi
Chapter 1. Introduction .....	1
1.1. Motivation.....	1
1.2. Research Questions .....	1
1.3. Limitation of Work .....	3
1.4. Summary of Contributions.....	3
1.5. Overview of the Thesis.....	4
Chapter 2. Addressing the Who to Imitate in Imitation Learning .....	6
2.1. Tailoring Robots to Task.....	6
2.1.1. Machine Learning.....	6
2.2. Imitation Learning.....	8
2.2.1. Importance of Imitation Learning.....	9
2.2.2. Five Major Questions in Imitation .....	9
2.2.3. Existing Work on the “Who” and “When” Questions.....	11
2.3. The Relationship between Teacher and Student -- What Defines a Good Teacher? ...	13
2.3.1. Student Perspective .....	13
2.3.2. Teacher Perspective.....	13
2.3.3. Reciprocal Relation .....	14

2.3.4. “Bad” Teacher .....	14
2.3.5. Summary .....	14
2.4. Engagement and Immediacy.....	15
2.4.1. Definition of Engagement .....	15
2.4.2. Elements of Engagement .....	16
2.4.3. Definition of Immediacy.....	17
2.4.4. Elements of Immediacy.....	17
2.5. Conclusion.....	20
Chapter 3. Linking Engagement to Effort in Physical Activities .....	21
3.1. Engagement, Revisited.....	21
3.1.1. Effort in Teaching Perspective .....	21
3.1.2. Brief Summary .....	23
3.2. Communication Modality in Human-Robot Interaction .....	23
3.2.1. Voice.....	24
3.2.2. Gesture.....	25
3.2.3. Tactile.....	26
3.3. Measuring the Effort.....	27
3.3.1. Definitions of Effort.....	28
3.3.2. Speech.....	29
3.3.3. Gesture.....	30
3.3.4. Methods to Measure Arm Activity.....	31
3.3.5. Categories to Take Account When Measuring Physical Activity.....	33
3.4. Summary and Next Action .....	34
Chapter 4. Evaluating Modality Preferences .....	36
4.1. Background .....	36

4.2. Related Work .....	37
4.3. The Study .....	40
4.4. The Task .....	41
4.5. The Robot.....	42
4.5.1. Software Development .....	43
4.5.2. Physical Compliance.....	45
4.5.3. Voice Recognition .....	47
4.5.4. Visual Tracking .....	48
4.5.5. Graphical User Interface .....	48
4.6. Experiment Method.....	49
4.6.1. Ethics Approval .....	49
4.6.2. Target Participants .....	49
4.6.3. Tester Participants .....	49
4.6.4. Equipment and Setting .....	50
4.6.5. Interaction Scenario.....	51
4.6.6. Procedure.....	52
4.6.7. Dependent Measurements .....	56
4.7. Results.....	57
4.7.1. Participants .....	57
4.7.2. Data Analysis.....	57
4.8. Discussion and Conclusion.....	59
4.8.1. Summary of Findings.....	59
4.8.2. Relation to Literature.....	60
4.8.3. Limitation .....	60
Chapter 5. Main Experiment Part 1: Robot Teaching .....	61

5.1. Background .....	61
5.2. The Task .....	62
5.3. Robotics Software .....	63
5.3.1. Controller GUI .....	64
5.3.2. Participants' GUI .....	65
5.3.3. Start/Stop Gesture Detection .....	65
5.3.4. Imitation Behaviour .....	66
5.3.5. Limitation .....	67
5.3.6. Tester Participants .....	67
5.4. Experiment Setup.....	68
5.4.1. Ethics Approval .....	68
5.4.2. Inviting Participants .....	68
5.4.3. Equipment and layout.....	69
5.5. Experiment Procedure .....	70
5.5.1. Pre-trial Part.....	70
5.5.2. Introduction Session .....	70
5.5.3. Teaching Session .....	72
5.5.4. Post-trial.....	74
5.6. Data Collection.....	74
5.6.1. Dependent Measurement.....	74
5.6.2. Method .....	76
5.7. Results.....	76
5.7.1. Participants .....	76
5.7.2. Data Analysis.....	77
5.7.3. Data for the Teacher Evaluator Experiment .....	80

5.8. Conclusion.....	81
Chapter 6. Main Experiment Part 2: Teacher Evaluator .....	83
6.1. Background .....	83
6.2. The Task .....	84
6.3. Software.....	84
6.3.1. Video Rating Interface .....	85
6.3.2. Tester Participants .....	87
6.4. Experiment Setup.....	87
6.4.1. Ethics Approval .....	87
6.4.2. Equipment and Layout.....	88
6.5. Experiment Procedure .....	89
6.5.1. Pre-trial Part.....	89
6.5.2. Introduction Session .....	89
6.5.3. Evaluation Session .....	92
6.6. Data Collection.....	94
6.6.1. Post-trial Part .....	94
6.6.2. Dependent Measurement.....	94
6.6.3. Method .....	95
6.7. Results.....	96
6.7.1. Participant.....	96
6.7.2. Questionnaire Data Analysis.....	97
6.8. Comparing the Data from the Teacher and the Evaluator .....	101
6.8.1. Data Preparation.....	102
6.8.2. Video Annotation Software .....	103
6.8.3. Behaviour Analysis .....	103

6.9. Conclusion.....	106
Chapter 7. Conclusion and Future Directions .....	107
7.1. Conclusion.....	107
7.2. Findings and Review of the Research Questions .....	108
7.3. Future Directions .....	109
Bibliography .....	111
Appendix A. Publication.....	122
Appendix B. Ethics Approval Documents.....	129
Appendix C. Paper-based Questionnaire Forms .....	134

## List of Tables

Table 5.1 Friedman test result of the questionnaire data .....	77
Table 5.2 Wilcoxon signed-rank test p-value of "effort" questionnaire.....	78
Table 5.3 Wilcoxon signed-rank test p-value of "robot follows demonstration" questionnaire .....	79
Table 5.4 Wilcoxon signed-rank test p-value of "Godspeed" questionnaire.....	80
Table 6.1 Results of significant difference tests .....	97
Table 6.2 Wilcoxon signed-rank test p-value and the average value of the questionnaires...	98

## List of Figures

Figure 4.1 The KASPAR robot for this study.....	43
Figure 4.2 System architecture .....	44
Figure 4.3 Compliance mechanism.....	46
Figure 4.4 Red and blue markers .....	47
Figure 4.5 The GUI for the first study .....	48
Figure 4.6 Experiment layout.....	50
Figure 4.7 Instruction sign.....	51
Figure 4.8 Activity flow of the participant in the “Modality Preferences” experiment.....	54
Figure 4.9 Activities in teaching the robot by moving the arms.....	55
Figure 4.10 Answer boxes.....	56
Figure 4.11 Age of the participants.....	57
Figure 4.12 Questionnaire result on human effortlessness .....	58
Figure 4.13 Questionnaire result on human enjoyment .....	58
Figure 4.14 Questionnaire result on different instruction modalities.....	58
Figure 5.1 Controller GUI .....	64
Figure 5.2 Participants’ GUI .....	65
Figure 5.3 Start/stop gesture.....	66
Figure 5.4 Experiment layout.....	69
Figure 5.5 Activities in the "Gesture Teacher" experiment .....	71
Figure 5.6 Interaction activities in gesture teaching (one gesture) .....	73
Figure 5.7 Answer boxes next to each gesture.....	74
Figure 5.8 Answer boxes of the IOS question .....	75
Figure 5.9 Age of the participants.....	76

Figure 5.10 Questionnaire result of "effort" .....	77
Figure 5.11 Questionnaire result of "enjoyment" .....	78
Figure 5.12 Questionnaire result of "robot follows demonstration" .....	79
Figure 5.13 Result of the IOS Questionnaire.....	79
Figure 5.14 Result of the Godspeed Questionnaire.....	80
Figure 6.1 Video rating interface: before playing .....	85
Figure 6.2 Video rating interface: playing.....	86
Figure 6.3 The video questionnaire within the GUI program .....	87
Figure 6.4 Experiment layout.....	88
Figure 6.5 Activities in the "Teacher Evaluator" experiment .....	91
Figure 6.6 The process of evaluating the video .....	93
Figure 6.7 Video questionnaire.....	94
Figure 6.8 Final questionnaire .....	95
Figure 6.9 Age of the participants.....	96
Figure 6.10 Pairs of evaluators and teachers.....	96
Figure 6.11 Results of QV1 (top), QV2 (mid), and QV3 (bottom) questionnaires .....	99
Figure 6.12 Results of QV4 (top), QV5 (mid), and FQ (bottom) questionnaires.....	100
Figure 6.13 Example plot of evaluation signals .....	102
Figure 6.14 Video annotation program .....	103

# Chapter 1. Introduction

Robots have been used widely in many areas including but not limited to: industrial works, such as in factories; education, being used in the classroom in order to induce motivation and increase knowledge in children; helping children and adults with special needs and disabilities such as autism develop and improve motor, social and other skills; assistants, ranging from personal assistant to helping doctors in the medical room, and many more. Robots are extremely helpful and current technology can be further developed to help humanity in even more ways.

## 1.1. Motivation

The study presented in this thesis was motivated by a study of how robots can be taught by humans, with an emphasis on allowing persons without programming skills to teach robots. Learning by demonstration was chosen specifically as the main interest as it is one of the main ways humans learn: by imitating. One of the advantages is that the human does not require a significant amount of knowledge about robotics and it can include people of all ages and skills as they only need to know how to demonstrate the task to the robot.

In imitation learning, there are five major questions, and one of them is “who to imitate”. In the case of a robot being taught by a human, this is how a robot can assess who is a good teacher i.e. what makes that person a worthwhile imitation subject. This thesis aimed to address this by evaluating *how a robot can assess who is a good teacher*.

In doing so, this study evaluated how human evaluators observe the teaching quality of other human teachers. In this case, the human evaluator is seen as the robot that evaluates the human teacher. The study then tried to use this to establish measurement metrics that can be utilised as rules or algorithms to determine what would be beneficial from a robot’s point of view in evaluating a human teacher.

## 1.2. Research Questions

The focus of this thesis was to investigate what criteria could or should be used by a robot to evaluate whether a human teacher is (or potentially could be) a good teacher. In effect, the robot should choose the teacher that can maximize the benefit to the robot using learning by imitation/demonstration.

To determine this, the study presented in this thesis mainly addressed the following research questions:

1. *What input modalities do humans prefer in teaching a robot? (RQ1)*

Running a fully autonomous robot in a human-robot interaction (HRI) experiment is very challenging and many studies bypass these difficulties by conducting the experiment using a Wizard of Oz (WoZ) approach. However, this essentially uses a human to evaluate the appropriate modalities. An autonomous robot would need to make the decision on the best modality to use by itself. This research question therefore tries to establish what input modalities are necessary to be provided by a robot to run an HRI experiment. For example, in order to teach a robot an action, is it possible to use only voice interaction? To use only one modality is beneficial as it minimises/reduces research preparation requirement. This research question is addressed in Chapter 4.

2. *How do humans evaluate the perception of engagement when evaluating a human teaching a robot? (RQ2)*

In investigating what criteria are needed to evaluate a good teacher this study evaluated the literature by looking at teacher-student relationship from a human-human interaction perspective. From the gathered literature two factors are considered to play an important role in defining a good teacher. Based on this, a main experiment was conducted to verify whether the same conditions applied to human-robot interaction. The main experiment was conducted in two sub-experiments. Chapter 4 discusses the first part of the experiment which records the data from teacher participants. Chapter 5 discusses the second part of the experiment where other humans evaluated the level of engagement of the teachers. Comparison between results from both experiments is presented at the end of Chapter 5.

3. *Can physical activity measured by the robot be used to measure the level of engagement? (RQ3)*

The study tried to link physical effort as a possible metric that could be used by the robot to measure the level of engagement of the teacher. This is discussed in Chapter 3. Based on this, the study checked whether the recorded physical activity data from the first part of the main experiment (Chapter 5) can be used to measure the level of engagement. The result is presented in Chapter 6.

### 1.3. Limitation of Work

This study wanted to see if a representative example of research laboratory robot technology is capable of assessing the quality of human teaching. In this case, the study used the KASPAR robot, which was developed by the University of Hertfordshire. However, the current technology may have some limitations that prevent the robot and the supporting elements from producing complete investigation results.

This study could have adopted another alternative method, such as the Wizard of Oz (WoZ) approach. The WoZ approach could simulate a human-like autonomous behaviour, which could be beneficial for the human-robot interaction (HRI) research. However, this study's purpose is to check whether this developed robot is useful in evaluating the human's teaching quality. Therefore, the robot in this study ran fully autonomously instead of using the WoZ approach.

For this study, a new software for the robot was developed to provide autonomous behaviours in the experiments. An investigatory experiment was conducted partly to validate whether these autonomous behaviours were sufficient for the main experiment to be carried out. However, some performance issues were encountered during the experiments. For example, the system failed to record the participants' voice properly in the main experiment. This was due to the limitations of the technology being used in the experiment. Consequently, this prevented some of the results expected from the study from being fully realised.

### 1.4. Summary of Contributions

The main contributions to knowledge of works described in this thesis are:

1. *Evaluation of input modality preference to teach a robot*

Several studies have been conducted in human-computer interaction to evaluate the user preference in input modality to interact with computers. The study in this thesis evaluated the preference in human-robot interaction. The study uses real-time autonomous behaviours for the robot (in contrast to the WoZ approach) to capture the dynamics of the interaction. Thus, the human behaviour is influenced by a real robot that has certain limitations and not a perfect imitation of human behaviour. However, the robot's behaviour is also consistent and controlled by the same program, not depending on a human that might be varied in controlling it.

## 2. *Addressing the “who to imitate” in imitation learning*

In the case of the robot being taught by a human, the study addressed the question by finding possible measurements to define a good teacher. Through literature research the study made an association to the engagement and immediacy of the teacher which is important in human-human interaction to be applied in a human-robot interaction context. The main experiment in this study was conducted to verify this association.

## 3. *Method to capture real-time video evaluation rating*

The study developed an original graphical user interface which allowed the participant to watch a video and to rate the event in the video, through mouse movements, while watching the video in real-time.

## 1.5. Overview of the Thesis

**Chapter 2** provides the background literature review that drove this study. It briefly discusses machine learning and introduces imitation learning and the benefit of this approach for allowing a robot to learn. It lists the five major questions as challenges in imitation learning. Some measurable elements or categories gathered from the literature to establish a possible metric for measuring a “good” teacher are discussed here.

**Chapter 3** gathers research from the literature to identify possible measurable physical activity attributes that may induce engagement. The chapter starts by revisiting the engagement topic in terms effort evaluation. The discussion also covers the communication modality from the perspective of human-robot interaction. Later, the chapter discusses how to possibly measure the physical attributes of human activity when the human communicates to teach a robot. Some measurement elements listed in this chapter were used to address RQ3.

**Chapter 4** describes the investigatory experiment that evaluates what input modality humans prefer in teaching a robot. This is to address RQ1. The participants in the experiment were asked to teach a robot five arm movements. The participants could teach the robot through three input modalities: (i) voice, (ii) gesture demonstration, and (iii) physical manipulation. At the end of the experiment, the participant filled in a questionnaire to rate their preference of those input modalities.

**Chapter 5** describes the first part of the main experiment. In the experiment, the participants (as teachers) were asked to teach a robot six arm gestures. The robot imitated

any arm movements the participants made. The data from the experiment was recorded to be analysed in the second part of the main experiment.

**Chapter 6** describes the second part of the main experiment. In the experiment, the participants (as evaluators) were asked to evaluate the teachers from main experiment part 1. The participants rate the level of engagement by watching the video of the teachers through a program interface that was specially developed for this study. Comparison between the result of main experiment part 1 and part 2 is discussed later in this chapter. This comparison was also used to address RQ3. Both Chapter 5 and Chapter 6 are to address RQ2.

**Chapter 7** concludes the study and discusses the future works.

## Chapter 2. Addressing the Who to Imitate in Imitation Learning

This chapter presents the background research that drives this study. The five major questions in imitation learning are explained in brief and one of them, namely the “who to imitate” question which implies answering the “what constitutes a ‘good’ teacher” question, is the major focus of this work. Some measurable elements or categories were gathered from literature to establish a possible metric for measuring a “good” teacher. These elements were then considered by conducting the experiments which are discussed later in other chapters.

### 2.1. Tailoring Robots to Task

Robots are usually tailored to perform their own individual functions by manufacturers. However, this tailoring is either not capable of handling new scenarios, or it is not able to derive something new from its specified functions. Humans can simply learn to do something once we are given the necessary information and set of instructions. On the other hand, most robots are only created for specific purposes. Hence, for example, a robot that operates as a museum tour guide may not be able to wash the dishes.

To carry out a certain set function designed to be of benefit to its user, it would be necessary to tailor the robot for specific purposes in the field. For example, an industrial robot has to be tailored differently to achieve its desired functions; e.g., one robot might be specifically tailored to be able to spray-paint, and another for fitting wheels. In tailoring these tasks both robot embodiment and how the robot carries out these functions are important, and it is therefore crucial for the robot to operate by using the full potential of the embodiment.

What if a robot can be tailored to learn a new behaviour? The important question then becomes “how can robots learn?” With ever-growing technology and research, there are now increasingly capable robots that can learn. Today’s robots are able to learn by exploiting numerous machine learning techniques and these are discussed more fully in the following section.

#### 2.1.1. Machine Learning

Russell and Norvig (2009) and Mohri et al. (2012) discuss three main categories on how machines learn, namely: supervised, unsupervised, and reinforcement learning. These methods are briefly discussed below.

In general, supervised learning is conducted when the machine learns a function after receiving a set of labelled instances as training data. In this case, labels are available for a certain amount of training data, but are missing and need to be predicted for other instances. This is in contrast to unsupervised learning, where instances are unlabelled (see also Kotsiantis (2007) and Ghahramani (2004)). In this case, the machine has to discover implicit relationships in the dataset. There is also semi-supervised learning, where the machine is provided with both labelled and unlabelled instances (Zhu, 2010). The final category, reinforcement learning, falls between these extremes of the supervised and unsupervised learning. In this case, there is some form of scalar feedback available for each predictive step or action, but no precise labelling.

In reinforcement learning, the machine would learn through trial and error interactions with its environment in order to learn (Kaelbling et al., 1996); the machine is given no instructions as to what actions to take, but must instead discover which action yields the most success by trying them (Sutton and Barto, 1998). Furthermore, in this category of learning, an agent would explore the range of possible strategies and receive feedback on the outcome of the actions performed, and from this information would deduce the optimal strategy (Kober et al., 2013).

There are more elements in machine learning that go beyond the scope of this thesis. To name some of them: classification, regression, ranking, clustering, and dimensionality reduction. Interested readers might look further on Russell and Norvig (2009) and Mohri et al. (2012) for more information on machine learning.

Nonetheless, there is a challenge to machine learning. Hoffmann (1990) summarizes that in machine learning, for any amount of information which is acquired, humans are generally required to do all the work; people have to write complex programs and, in some cases, provide large amounts of input data to programs.

The issue, of course, is that not necessarily everyone is familiar with robotics. As robots become popular with technology advancement and marketing popularity, there needs to be a way by which humans can teach robots by themselves without being particularly familiar with the field of robotics at all. It would be desirable for the user to simply teach the robot what behaviour it needs to exhibit, and the robot can be “taught” to do the specific behaviour. Teaching the robot as if it were human, with the addition of the fact that the robot is able to be taught functions by a user, would be beneficial.

The concept of robots being able to be taught is especially important, as robot use is increasing and even replacing what once only humans did. However, it is crucial to remember that not everybody will be proficient in robot programming and therefore to tailor robots to their own needs may prove a difficult task. It would be a much easier task for a typical person to teach the robot on how they could function, just as a teacher and a student do in human-human interaction.

## 2.2. Imitation Learning

As robots become more mainstream in daily use, it becomes crucial that these robots acquire the capability to be able to learn new skills in the most natural way possible as perceived by humans—even by those who have no experience in robotics, and, especially, the robot should be able to be taught by people across different ranges of communities and across all different kind of needs.

In robotics, imitation learning is also referred to as *robot programming by demonstration* (RPD) (Friedrich, Münch, et al., 1996) and *robot learning from demonstration* (LfD) (Argall et al., 2009). It is also referred to by some other names, such as *learning by demonstration* (LbD) and *programming by demonstration* (PbD) (Argall et al., 2009). Within the very heart of these terminologies, imitation learning has changed the way robots are programmed in order for the robots to exhibit specific or relatively complex behaviours. In imitation learning, the teacher demonstrates how to accomplish a task, and the robot then learns from the demonstration to accomplish the same task. From another perspective, imitation learning, in summary, is the user teaching the robot to accomplish a certain function, and it is advantageous as the users who may not be familiar with robotics are then able to teach the robot what to do.

We humans typically learn by imitation of others (Dautenhahn et al., 2003). According to Billard et al. (2008), imitation learning is a highly effective form to harvest the dataset for learning. It can also significantly accelerate the learning of sensory-motor links (Andry et al., 2001). Imitation learning would also help to diagnose problems of perception and action should they arise in developmental stages in children (Andry et al., 2001). On the other hand, evidence also suggests that the mimicry of posture and movements, even of “strangers”, significantly improves the relationship between interaction partners (Chartrand and Bargh, 1999), which benefits the quality of the interaction.

### 2.2.1. Importance of Imitation Learning

Compared to reinforcement learning, which requires a large search space in order to find a good solution and therefore may require a long time in order to learn, learning via imitation can be accelerated. This can be achieved by providing (demonstrating) good solutions in the first place. As imitation is also a natural way for a human to learn, the teaching method in imitation learning is more user-friendly. Non-robotic experts can intuitively teach the robot new skills.

Learning by demonstration is one of the main ways humans learn. As we already subconsciously use this method as human beings, we can use it when we are teaching a robot. Advantages include the user not requiring to know a significant amount about robotics (similar to a human teacher-student setting), and demonstrators can include people of all age and skill as they only need to know how to demonstrate the way to accomplish the task to the robot.

Learning by demonstration also eases the burden of having to explicitly program a machine by a human user, either minimizing it or eliminating it completely (Billard et al., 2008). It would also be a “natural” form of interacting with a machine (Billard et al., 2008), as it would be between a teacher and a student.

However, imitation learning has its own challenges. For example, animatronic devices that may correctly perform an imitation of a human limb may not respond to drastic movement changes nor alter its movements to correspond with new situations (Breazeal and Scassellati, 2002), and therefore may introduce problems such as unintended obstruction due to not being able to adapt to new situations when the environment is unfamiliar. The following section discusses the challenges in imitation learning.

### 2.2.2. Five Major Questions in Imitation

While imitation might accelerate how a robot learns complex behaviours, there are many aspects to consider. According to Dautenhahn and Nehaniv (2002), the imitating agent faces five major questions while doing the imitation. These major questions are discussed below.

### **2.2.2.1. Who to Imitate**

This question is about *who to imitate*: what makes a model<sup>1</sup> good to imitate? Certain criteria should be established to measure whether a model is (or could be) a good teacher. When there are several demonstrators available, the imitator should assess these criteria in order to choose which demonstrator to imitate. For example, a simple criterion might be one that maximises a benefit to the imitator.

### **2.2.2.2. When to Imitate**

How can a robot know when imitation should take place? Imitation might be just for play, or it might be in a teacher-student context. Some portions of the demonstration could be part of the interaction, thus the robot should be able to segment a beginning and an end of the behaviour to imitate. It also has to decide whether the imitation should be carried out immediately or after the demonstration. Immediate imitation allows the imitator to perceive the same environmental conditions perceived by the demonstrator, as the imitation occurred at the same time. Deferred imitation occurs after the demonstration, the demonstrator may no longer be present, and the environmental conditions may have changed. The robot should also decide when it is appropriate to imitate depending on the social context and the availability of a good model.

### **2.2.2.3. What to Imitate**

In assessing *what to imitate* the imitator faces a number of challenges: should it imitate all states, actions, effects, or goals of the observed behaviour, or could it be only some part of them? For example, if a robot wants to imitate someone playing a guitar, should it use the same exact guitar model, or could it use an acoustic instead or an electric type? In the case where the player is nodding his or her head, should it also replicate this action? In the case that it can copy the guitar playing perfectly, should it also copy the goal of expressing feeling and emotion?

---

<sup>1</sup> Note that the terms “demonstrator”, “model”, and “teacher” are terms used to describe the demonstrator in this document

#### **2.2.2.4. How to Imitate**

This question is related to the correspondence problem (Nehaniv and Dautenhahn, 2002): How does the imitator know which parts of the model's physical embodiment match those of its embodiment? How should it map the embodiment if there are differences such as in size, degrees of freedom and dynamic models between the imitator and the model? In the case of guitar playing above, there could be a mirroring problem regarding the facing position. How can the robot know which is left and which is right? A similar problem would arise if the human is left-handed when the robot wants to play right-handed.

#### **2.2.2.5. How to Evaluate the Imitation**

This is the question of how the matching of the behavioural function is made. This is to measure whether duplication of actions made by the agent, states of the body or effects on the environment result in an imitative function. Selection of an appropriate metric plays a significant role. It will be used to capture the notion of the difference between the performed and desired actions and to measure the difference between attained and desired states (Nehaniv and Dautenhahn, 2001; 2002). The evaluator of this measurement can be the imitator, the demonstrator, or an external observer.

#### **2.2.3. Existing Work on the "Who" and "When" Questions**

To the best of the author's knowledge, the "who to imitate" question has not been widely addressed. Many studies encompass this question by establishing roles, with one being the "teacher" and the other as the imitator (Jansen and Belpaeme, 2006). From the field of developmental psychology, it has been suggested that infants come to understand others because the infant perceives that the person they observe may have similar behaviours with themselves, that they are "like me" (Meltzoff, 2007). From the perspective of a social robot (Dautenhahn, 1994; 1995), this "like me" identification may allow the robot to engage in "meaningful" interactions with its social environment. This is because the imitator can understand based on its own perception, as both the imitator and the model share a common embodiment and therefore the imitator can understand what the demonstrator is experiencing.

Kaipa et al. (2010) addressed the who to imitate question by offering a method that selects an appropriate teacher by discovering the similarity of physical structure modelling. This method detects the similarity with no prior knowledge of both body structures. With regard to similarity measurement, Shen et al. (2008) offer a method for identifying similarity and

synchronous behaviour between a human and a robot while the imitation takes place. The method uses signal correlation techniques to measure periodic activity in a body position such as a hand waving.

Robots that are autonomously able to decide when to imitate are highly desirable in human-robot Interaction (HRI). Instead of pressing buttons to give instructions such as “start learning”, natural communication could be used, such as using voice or gesture or any other social cues. These might enrich the social interaction. Many researchers have been pursuing the idea of integrating social cues in the imitation learning (see Billard et al. (2008)). However, their works have mainly been addressing the “what” question, i.e. to highlight the important components of the demonstration.

The use of voice commands have been explored in (Nicolescu and Mataric, 2003; Lockerd and Breazeal, 2004; Clodict et al., 2007; Cakmak et al., 2010) to give commands such as to mark the “start” or “stop” of a learning mode and also to start a reproduction mode. The problem with these approaches is that the list of words or sentences to represent the commands is static and relatively close to a certain type of task. Static words are problematic, as the researchers have to carefully register the most common words to represent the command for related tasks. This problem will arise even more when multiple languages are considered.

Other researchers (Fritsch et al., 2005; Rohlfing et al., 2006; Nagai and Rohlfing, 2009) have been investigating whether the use of gesture has relevance in addressing the when to imitate question. Their works are based on research by Brand et al. (2002), which concludes that in action-demonstrations made by mothers to infants, mothers notably altered their movements, such as a wider range and variety of movements, in order for their infants to understand. Infants, like robots, have little knowledge about the context of their actions, their surroundings and environment, and even the partner they interact with. Through modifying their body movement, the parents direct the infants’ attention and help them know that the parents want to demonstrate something.

The study presented here attempts to move further in addressing the “who to imitate” question and the following section discusses the relationship between teacher and the student in relation to this concept.

## 2.3. The Relationship between Teacher and Student -- What Defines a Good Teacher?

“Teaching” in this thesis is not limited to teacher-student in a classroom setting. The relation of a teacher to a student could go beyond the scope of learning in terms of typical learning such as in a school. It could happen anywhere, for example, at home, where, as discussed by Maccoby (1992), parents are teachers and children are learners. However, this study focuses on evaluating the teacher in the formal relation of teacher to student.

### 2.3.1. Student Perspective

In generic educational systems in schools, there would be a teacher and a student. Generally, the teacher teaches a subject and the student follows it and learns it, and usually, the student will be tested on the taught subject, and their performance will be marked based on the grade they have on an assessment or exam.

Student engagement plays an important role in the outcome of the study. This is shown such in the research by Kuh et al. (2008) that investigated the effect of student engagement on the learning outcome. The research evaluated the engagement of students by measuring the time spent studying as well as the time spent in co-curricular and other educational activities (such as asked questions in class). The results were positive in that student engagement affected the score of academic outcome.

### 2.3.2. Teacher Perspective

While the student’s effort plays a role, is the teachers’ quality also an important factor? If it is, what are the relevant factors?

According to Klassen et al. (2013), engagement factors, in this case, “teacher engagement”, plays an important role. This is because fully engaged teachers deliver effective teaching. Furthermore, teacher engagement contributes to the school quality (see Rutter and Jacobson (1986), and Louis and Smith (1992)).

As a way to evaluate teacher engagement, Klassen et al. (2013) proposed a method that divided the engagement of the teacher into four categories. They are: (i) cognitive engagement, (ii) emotional engagement, (iii) social engagement to student, and (iv) social engagement to colleagues. The measurements were based on a questionnaire in which participants gave feedback to statements such as "I feel happy while teaching".

There are also some other factors that contribute to the quality of teaching from the teacher perspective. Chesebro and McCroskey (2001) investigated the relation of teacher clarity immediacy to student apprehension, motivation, affect, and cognitive learning. They found that clear and immediate teaching could improve the instructional outcome.

### 2.3.3. Reciprocal Relation

According to Louis and Smith (1992) and Skinner and Belmont (1993), student engagement and teacher engagement are interrelated. Teacher engagement could affect student engagement and vice versa. When the student viewed a teacher as not caring or making learning stimulating, they would be less engaging. On the other side, the teacher may care less if the student shows less effort.

### 2.3.4. “Bad” Teacher

Professional teachers require a certain level of knowledge and sometimes certifications to be able to teach. This certified knowledge is considered necessary to be proficient enough to teach. In this study, “human teachers” means any human being, and the definition of “teaching” in this study means any type of knowledge transfer, regardless of whether they are working in real life as a teacher or not. They might, or might not be good at “teaching”.

While imitation might accelerate the process of knowledge transfer, the human might provide information that might also degrade the knowledge attained by the robot. This is not necessarily that the teacher is “bad”, but it could happen accidentally. Friedrich et al. (1996) identifies these source of degradation as the following:

- Actions that are unnecessary, and do not contribute to achieving the final goal that is expected of the robot
- Incorrect actions/motions
- Unmotivated actions that cannot be learned by the system
- Choice of scenario—some conditions may strictly limit the robot from learning
- Wrong intention—the information is correct but in a different context

### 2.3.5. Summary

As robots become mainstream in society, it will then become crucial that these robots acquire the capability to be able to learn new skills in the most natural way possible as

perceived by humans—even by those who have no experience in robotics, and, especially, the robot should be able to be taught by people across different ranges of communities and across all different kind of needs.

The earlier sections have outlined a way to address the “who to imitate” question in imitation learning by looking at the relation between teachers-students in the learning process. In order to apply this in this study to the general human population, regardless of whether the individual is a professional teacher or not, the existing research noted above has identified briefly what factors should be considered if the teacher is “bad” (e.g. by accidentally giving incorrect knowledge to the learning system) as well as identifying qualities that are possessed by “good” teachers that could potentially deliver quality teaching. Among these qualities were teacher engagement and immediacy.

This study now investigates in more depth these “good” teaching aspects in the following sections.

## **2.4. Engagement and Immediacy**

Engagement and immediacy are interconnected, for example, in a conversation, a person might use immediacy to increase the engagement of another side of the conversation. These two factors are considered to play an important role for a teacher to give quality teaching to a student as discussed earlier in the previous section.

In order to address the “who to imitate” question in imitation learning, especially in identifying a “good” teacher, this study identified from gathered literature what constitutes these factors. The following discussion further explores these aspects in order to identify possible metrics for measuring these factors.

### **2.4.1. Definition of Engagement**

There are different wordings that define engagement. One of them, according to Goffman (2008), is of two or more people participating in a situation wherein they maintain a single focus on cognitive and visual attention. According to Sidner and Dzikovska (2002), engagement is a process where participants initiate contact, continue their interaction and decide when to conclude their connection. Bickmore et al. (2010) states, from a human-robot interaction perspective, that engagement is "the degree of involvement a person chooses to have with a system over time."

Another way to define engagement is the process whereby participants establish, maintain, develop and finish their perceived interaction. According to Sidner et al. (2004), the engagement process includes: the first contact, negotiating whether to participate, checking that others are still participating in the interaction, decide if they still desire to interact further and deciding when to conclude the interaction. Xu et al. (2013) considered engagement as “an emotional state linked to the participant’s goal of receiving and elaborating new and potentially useful knowledge.”

#### 2.4.2. Elements of Engagement

To translate the above definitions of engagement into possibly measurable elements, this study looked further into the literature to find what constitutes engagement. This study was keen to focus on the summary by Glas and Pelachaud (2015). They summarised the underlying elements that are considered fundamental to engagement, and sometimes used interchangeably to refer to engagement, as the following:

1. *Attention*: Peters et al. (2009) suggest attentional and emotional involvement as studies related to engagement, further stating that selective attention to a particular stimulus or a subject is necessary for a basic form of engagement. It is also important to assess the level of attention a listener may have before even establishing engagement (Peters et al., 2005).
2. *Involvement*: In terms of engagement defined as "being occupied with", it strongly suggests involvement, which is also synonymous with the term of being “occupied” or “involved” in something (Peters et al., 2009). It also has to do with subjects, stimuli or items that make the subject feel the sense of “immersion” (Lombard et al., 2000).
3. *Interest*: Interest relates to attention, as establishing engagement relies heavily on whether a person is interested in the subject of interaction or not. The person(s) may conduct an interaction wherein they hold no interest in the subject/stimuli whatsoever, but they may not be necessarily engaged with the topic at hand.
4. *Empathy/Rapport*: Empathy and rapport are somewhat synonymous, although in different contexts. Whilst rapport is “the feeling of being ‘in sync’ with your conversational partners” (Huang et al., 2011), empathy is the ability to understand and interpret another’s emotions or feelings, or what they are trying to convey (Decety and Jackson, 2004). As two separable elements, Glas and Pelachaud in (2015) identified these as:

- a. *Empathy*: Empathy is “a sense of similarity between the feelings one experiences and those expressed by others” (Decety and Jackson, 2004). Engagement, therefore, implies the empathic connection between participants (Glas and Pelachaud, 2015).
  - b. *Rapport*: Tickle-Degnen and Rosenthal (1990) emphasize the three essential aspects of rapport: *mutual attentiveness, positivity, and coordination*. *Mutual attentiveness* generates focused and cohesive interaction, *positivity* refers to mutual friendliness (however, it can also be negative) and *co-ordination* refers to being “in sync” between participants.
5. *Stance*: is related to inter-subjectivity, “an attitude which, for some time, is expressed and sustained interactively in communication in a unimodal or multimodal manner”.

Although the above list describes the elements of the engagement, they are concepts and relatively subjective, which is very challenging if the measurement is to be implemented in a robot. Nevertheless, this thesis further studies the literature to identify possible metrics for measuring the other factor, which is immediacy, and this is discussed in the following section.

### 2.4.3. Definition of Immediacy

Our engagement towards others indicates how our behaviour may be perceived. For example, if we do not engage with the person who is talking to us, they may think that we are not paying attention and therefore may perceive us as “uninterested” or, in certain cases, “rude”. There are probably some reasons why we do not engage with a person, and one of them could be the lack of immediacy factors.

According to the Oxford dictionary<sup>2</sup>, immediacy is “the quality of bringing one into direct and instant involvement with something, giving rise to a sense of urgency or excitement.” As an active behaviour, immediacy is actions demonstrated by speakers to decrease the psychological distance between themselves and their listeners (Mehrabian, 1966).

### 2.4.4. Elements of Immediacy

Szafir and Mutlu (2012) summarise that the reason why people are at more ease when sharing their thoughts and feelings is because they have high levels of immediacy, and

---

<sup>2</sup> <https://en.oxforddictionaries.com/definition/immediacy>

demonstrates such levels of immediacy through both verbal and nonverbal behaviour. They explained, regarding these channels, how immediacy cues are exhibited as the following:

1. *Verbal immediacy*

Verbal immediacy includes the content that is spoken, impassioned fortitude, and vocal signals such as tone of voice and volume of speaking, which may impact the listeners. These impacts may either affect the listener positively or negatively. Generically, increased immediacy can be sustained and accomplished if the content of the communication is perceived as friendly and empathic. Respect of listener's role and position and the appropriation that the topic spoken is to the listener's interest also helps to increase immediacy, and in *some* cases, display of positive emotion through words chosen.

2. *Nonverbal immediacy*

Nonverbal immediacy includes the cues of bodily language such as gestures and display of facial emotions, and it often must match verbal immediacy cues and the appropriate words to increase immediacy. In most settings, speakers who are more indicative in expression and incorporate gestures in their interaction create greater immediacy and attain higher levels of engagement by their listeners. Nonverbal immediacy cues often go hand-in-hand with verbal cues. The use of gesture along with speech goes popularly through daily communication.

In the context of teacher-student in learning scenarios, the literature also shows some action behaviour that related to immediacy give benefits in increasing the engagement of the student. Some of this these are:

1. *Vocal cues*. This includes the pitch and tone of voice, tempo, and loudness/volume. Speakers may use these vocal cues singularly or in combination in order to stress single words or statements, add emotion to their speech, and/or encourage an increase in listener engagement. This is also a way to convey underlying tones or emotions behind said speech to give a deep effect to the listener. Brown and Howard (2013) conducted a study that employed verbal cues in a robot that supported and gave feedback in a maths test, in a mode where the robot employed verbal cues, participants were less likely to get bored and a percentage of subjects felt much more motivated, and they also took less time in answering the maths questions.

2. *Increase volume* to increase instructor clarity, and also the emphasizing of utterance to make sure the words cannot be perceived another way/wrongly heard and maintain apprehension (Chesebro and McCroskey, 2001).
3. Increased *eye contact* with students. Eye contact has been reported to indicate a higher position in authority as well as mutuality, participation, and immediacy. Classroom research also suggests that eye gaze improves students' ability to recall information (Otteson and Otteson, 1980).
4. An instructor initiating *head nodding* has also been shown to have a positive effect on student reaction towards educators. Head nodding is frequently cited as an immediacy behaviour (Hale and Burgoon, 1984). Nodding is also a sign of reassurance, therefore giving the students a sense of understanding and developing a greater sense of trust and immediacy.

The last point above mentioned head nodding, which can be defined as a gesture. Regarding gesture, Cassell et al. (1994) identified four major categories of gesture in human-human interaction. They are as the following:

1. *Iconic*

Iconic gestures are such those that represent (or closely represent) the meaning of speech content or segment. It is one that exhibits a closely relevant meaning to simultaneous expressed word or phrase (Beattie and Shovelton, 1999). It also has formal relation to the semantic content of a linguistic unit (McNeill, 1985).

2. *Metaphoric*

Rather than direct gestures, metaphoric gestures are more often used in conceptual senses and are used to define ideas and concepts rather than exact definitions. Whilst both iconic gestures and metaphoric gestures similarly illustrate spoken sentences, iconic gestures represent physical features and metaphoric gestures tend to be more abstract, such as depicting an idea, rather than for depicting objects (Straube et al., 2011).

Straube et al. (2011) give an example of the difference: in the sentence "The politician builds a *bridge* to the next topic" and a subject depicts an arch with a hand, it is a *metaphoric* gesture as the "bridge" is a metaphoric sense. Whilst in "There is a *bridge* over the river" and the hand gesture is more or less the same, it is called an *iconic* gesture as the "bridge" being addressed is a physical one, not a metaphorical one.

### 3. *Deictic*

Deictic gestures are direct referents to the immediate environment, used in concrete or abstract pointing (Roth, 2001). Iverson and Goldin-Meadow (2005) states that children show three types of deictic gestures: *showing* (holding up an object), *index point* (directing an extended index finger at the referent) and *palm point* (extending a flat palm towards referent). Employed alongside deictic utterances, it can play an important role in classroom teaching (Roth, 2001).

### 4. *Beats*

Beat gestures are “simple, rhythmic gestures that do not convey semantic content” (Alibali et al., 2001). Beat gestures are basic patterning of usually hand movements, especially used when to emphasize a point. Beats and repetition can play a role in the context of what is being talked about.

## 2.5. Conclusion

This chapter presented related information from literature research to support this study. It briefly discussed machine learning and introduced imitation learning and the benefit of this approach for allowing a robot to learn. It listed the five major questions as challenges in imitation learning. It was noted that the “who to imitate” and “when to imitate” questions have not been addressed widely. The existing work in addressing both questions are listed and this study, which is focusing on the least addressed question: “who to imitate?”

This chapter also looked further in addressing “who to imitate” by considering the “good” teacher effect to the student, especially in classroom situations. It then considered two factors that are possibly playing an important role in defining a good teacher; the interrelated concepts of engagement and immediacy. The elements that constitute these two factors were gathered from literature and presented above.

Based on the elements of engagement and immediacy, more information on how to measure these elements are presented in the next chapter.

## Chapter 3. Linking Engagement to Effort in Physical Activities

This chapter discusses the physical activity that correlates to teaching, and how we can measure the physical activity that relates to cues that may induce engagement. The chapter starts by revisiting the engagement topic in terms of effort evaluation. Then, as humans need to communicate to teach robots, the discussion covers the communication modality from the perspective of human-robot interaction. Later, the chapter also discusses how to possibly measure the physical attributes of human activity when the human communicates to teach a robot.

### 3.1. Engagement, Revisited

From the discussion of the literature in the previous chapter, engagement emerges as one of the clear connotations of a better performing teacher. Rutter and Jacobson (1986) summarise that engaged teachers are ones who are enthusiastic in their subject/department of teaching, ones who commit themselves in student achievement and students' success, and do more than what they are expected to do. Self-confidence and overall uniqueness as an individual teacher is also an indicator that a teacher engages in their work life.

The engaged teacher will also exhibit emotions according to student performance, reflecting themselves onto student work and feel pride or disappointment of a student's work and behaviour. Engaged teachers are also more likely to show clarity and are more apprehensive (Chesebro and McCroskey, 2001).

In general, people who are dedicated and engaged with their work lives are more likely to have high levels of energy by working strongly (Bakker et al., 2008), and a study found that engagement is best predicted by both job and personal resources (Schaufeli and Bakker, 2004). They, therefore, expend more energy and put more effort into their profession.

People that are fully engaged in their work are more likely to be upbeat and positive about their work, and give their utmost potential and effort in their work (Loehr et al., 2005). People that engage more in their work and put more effort are more likely to be successful, productive and satisfied with their whole wellbeing.

#### 3.1.1. Effort in Teaching Perspective

Section 2.4.1 in the previous chapter lists the five elements of engagement as suggested by Glas and Pelachaud (2015). They are attention, involvement, interest, empathy/rapport, and

stance. This study was investigating a way for a robot to be able to measure the engagement of a human teacher. Although these elements could be used to evaluate engagement, they are concepts and very challenging to be measured, especially by a robot. In relation to this, this study considered the common thing discussed in the previous section: people put more effort when they are engaged. The study was investigating further by gathering literature related to the effort in teaching activity and also tried to link physical effort as a possible metric that could be used to measure the level of engagement of a teacher.

From the teaching perspective, teacher involvement with individual students is most likely to have an impact on students' perceptions of teachers, whilst lack of involvement means that the students experience teachers as "less consistent" (Skinner and Belmont, 1993). It is also important to notice what sort of "aspects" a teacher needs to make a student feel "involved"—aspects such as teacher clarity and immediacy, and their *effort* to make a student feel involved.

Human effort plays a large part in teaching, and depending on how much effort teachers exert to their students, it can largely affect how the student behaves or how the student performs. Evidence shows that teachers play a major role in how students learn (Di Gropello and Marshall, 2005). If a teacher puts more effort in their teaching, especially if the behaviours employed by teachers are similar to behaviours that effective parents exhibit, they are more likely to form and maintain both a positive and a productive relationship with their student(s) (Meltzer et al., 2001).

Usually, the more effort a teacher expends to engage with their classroom and make the students feel involved, the better a student will perform—not only in aspects of performance, but also their state of relationship with their teacher (Christophel, 1990).

Teacher immediacy is also viewed by students as positive and favourable (Christophel, 1990). Teachers who thereby use more "verbal items" to engage and involve their students, such as telling of their own experiences outside of class or using humour in class, going an extra mile and exerting more effort necessary to engage their students.

Meltzer et al. (2001) conducted a study wherein they investigated whether teacher effort committed to individual children "varies as a function of children's personal characteristics". They found that the teacher has to exert more effort in time and energy to treat children with challenging behaviour in order to deliver the same curriculum. Therefore, it may be the

case that less effort would be needed when teaching children without behaviour challenges when teaching the same curriculum.

### 3.1.2. Brief Summary

The discussion above shows that effort is needed as part the quality elements of the engaged teacher. In some senses, the effort is related to the energy exerted by the teacher in terms of engaging. Therefore, in this thesis, the hypothesis is that this effort may be a metric that can be used to measure the level of engagement of the teacher.

The following section discusses how to measure the effort with a literature survey about communication modalities in human-robot interaction. This is because the human needs to communicate, through a communication channel, to the robot in order to teach. Later, Section 3.3 discusses possible ways to measure the effort.

## 3.2. Communication Modality in Human-Robot Interaction

Since Sheridan (1992) studied the teleoperation of industrial robotic platforms as human-robot interaction (HRI), the research in HRI has expanded into several different research areas (Goodrich and Schultz, 2007). One of the areas of particular interest is multimodal interfaces for multimodal interactions, where a human can interact with the robot more closely to how they would with another human being.

In daily life, humans naturally interact with others using multimodal interaction. Humans have the amazing ability to effectively exchange information and convey feelings through eye contact (Hugot, 2007). We use facial expression, posture and small head movements (Knapp et al., 2013). Children start to use the modality of gesture to go along with their undeveloped speech at an age as young as around ten months old (Goldin-Meadow, 1999). Even deaf children have been shown to use gestures in place of verbal communication at an early age (Goldin-Meadow and Morford, 1985).

However, humans would not usually interact with machines in the same way as they would with other humans. For certain machines, people do communicate in ways similar to ways during a human-human interaction, anthropomorphising with the machine (Perzanowski et al., 2001). On the other hand, with recent advances in technology, it is now quite common for humans to interact and speak with more machines. Consumer products like smartphones and similar devices have virtual assistants to obtain information and have been developed to

have enough computing power to capture human speech, such as Siri<sup>3</sup>, Cortana<sup>4</sup>, and Google Now<sup>5</sup>. Such systems are able to obtain information when asked by voice or text input, such as asking about the weather or when a flight will leave, or simple commands that they are able to do within the device such as calling someone or putting up a reminder for a dentist appointment. Language-learning programs such as Duolingo<sup>6</sup> also prompt users to input sentences as answers to questions.

Stiefelbogen et al. (2004) suggested that in order facilitate natural communication between a human and a robot, multi-modal interfaces are necessary. One of the objectives of HRI is to make human-robot interaction easier, much more intuitive and certainly user-friendly. By providing a multimodal interface, it would greatly improve on the user's engagement with the robot and interact with them with a more familiar and natural manner, similar (if not completely) to the way they may interact with other human beings.

As humans, we can hear, see, touch, smell, and taste and they are input channels. As output channels, we can speak, make gestures (including body poses), touch something, show facial expressions, and gaze with our eyes. The research presented in this thesis, in particular, was interested in using voice, gesture, and tactile modality in investigating how humans teach a robot, especially in measuring the level of engagement. The following section expands on these three modalities.

### 3.2.1. Voice

Human speech is arguably the most common and most prominent interaction modality. We use speech to converse with other human beings and perhaps other artefacts or animals, and we commonly use speech including but not limited to: exchanging information, socialising, conveying our feelings and/or to give commands. The tone and volume of our speech can also tell a lot about our emotions and how we feel (e.g. a loud shout followed by fast-spoken words may mean that the person is angry). The words we say also play a major

---

<sup>3</sup> <http://www.apple.com/uk/ios/siri/>

<sup>4</sup> <http://www.microsoft.com/en-us/windows/cortana>

<sup>5</sup> <http://www.google.co.uk/landing/now/>

<sup>6</sup> <https://www.duolingo.com/>

role in how we are perceived (e.g. the prominent use of swear words may give the impression that we are vulgar or rude). Speech is also a way identifying other people. Therefore, for a human-computer interaction that is natural, it is significant for the robot to be able to comprehend and understand spontaneous speech (Stiefelhagen et al., 2004).

As robots become more prominent in society, they must also make their behaviour more “natural” to humans, not only appeal to different categories of persons but also how they convey themselves (Breazeal and Aryananda, 2002). Therefore, detecting how a person might feel just through their speech poses a challenge. In many cases, close-talking microphones are also necessary in HRI experiments, as remote microphones/microphones at a distance presents a problem since we want a robot that can operate without people having to force themselves to wear the “necessary” type of microphone in order for the robot just to work (Stiefelhagen et al., 2004).

However, speech interaction, although the most crucial and most prominent of the modalities, is also highly problematic. According to Zuo (2011), these are the problems that voice interaction currently faces:

1. Accuracy of speech recognition
2. Strategy of dialogue
3. Out-of-vocabulary words (words that the speech interfaces have not yet prepared)
4. Utterance target

To add to the factors above, the interface should also be able to face possible problems that human speech may pose such as interruptions, mumbling, or irrelevant background noise (Perzanowski et al., 2001).

### 3.2.2. Gesture

Gesture is an important feature of social interaction and is naturally and most of the time used by humans to convey a message that speech cannot perhaps show on its own such as describing something like iconic information. According to Salem et al. (2011), gestures often go hand-in-hand with speech and serve as something that somewhat “elaborates” human speech, and it can be a minimum finger shift or a large arm movement, depending on what we are talking about and how much gesture we need to “elaborate” our speech. Sometimes, the gesture itself takes over some words of speech (such as generic pointing to show that something is “there”). Gestures are also somewhat an indicator of enthusiasm on the

subject that is being talked about. Gestures are defined as natural nonverbal human-robot interaction (Obaid et al., 2012).

Gesture is a subject researched for many years by many researchers from different fields of research and especially social sciences. Hand, arm and body movements can be classified as gestures, although definitions vary very widely and a great deal of research has aimed to describe different types of gestures (Salem et al., 2011). For example, these modalities may include pointing gestures, eye contact, and emotions expressed (Stiefelhagen et al., 2004). Hand signals, sign language, and other forms of movements from the head, the arm and the body are considered gestures.

Research has also been undertaken on the use of algorithms to identify gestures in real time (Obaid et al., 2012). Gestures constructed by robots had a positive impact on human participants (Kose-Bagci et al., 2009). Gesture is also very closely related to learning by imitation.

Gestures by themselves, however, are a limited interaction system. A gesture by itself can be interpreted the wrong way, or can only be vaguely interpreted instead of having a clear meaning (Chao et al., 2010). A gesture must then retain its accuracy on what it wants to convey. An ambiguous gesture with no other modalities can also lead to inaccuracy and confusion on what to interpret, and gesture must, therefore, be also assisted by another modality—a particular one in mind that goes hand-in-hand with it is speech.

### 3.2.3. Tactile

Tactile interaction is one of the non-verbal types of interaction. Physical touch is divided into cutaneous and kinaesthetic, the former focusing on smaller details (such as skin stretch or vibrations) and the latter for larger details, like basic shapes (Robins et al., 2010). Both complete the basics of human tactile interaction. Physical contact is especially used by children as a form of communication and trust building (Robins et al., 2010). According to Argall et al. (2010), physical robot-human contact may be anticipated. If the contact is unexpected by the robot, it may either help with the robot interaction, or disturb it completely.

Tactile interaction in HRI ranges from strictly industrial robots to one with social interactions with humans (Argall and Billard, 2010). In programming by demonstration, the demonstrator/user should be able to teach the robot by direct touching and moving (Grunwald et al., 2003). This type of method is similar to the form of programming of

industrial robots (Grunwald et al., 2003). A survey by Argall et al. (2010) suggests that tactile interaction with humans must at least consider one of the following:

1. Performance with humans: how the robot performs through an interaction with a human in a conducted session, and possibilities of it behaving unexpectedly (such as unintended physical obstruction)
2. Necessary tactile implementation for behaviour execution: obligatory robot-human or human-robot tactile interaction, examples including but not limited to:
  - a.) The human guiding/demonstrating the robot
  - b.) Human-robot team task
  - c.) Human-robot contact being the sole point of interaction, such as robot-assisted touch therapy)
3. Necessary tactile implementation for behaviour development: the robot is dependent on tactile contact from a human to form or improve a behaviour

One of the examples of a robot with tactile interaction capability is KASPAR, a child-sized robot with a purposefully minimally expressive face that focuses on utilising tactile play scenarios using bodily expressions (Dautenhahn et al., 2009). KASPAR interactions primarily focus on tactile and gesture, with the KASPAR studies aiming to facilitate interaction with autistic children—who commonly have difficulties in understanding gestures and facial expressions, both verbal and non-verbal communication, and impaired in understanding one’s intentions—interacting with a robot (Robins et al., 2010).

Another example includes a robot developed by DLR’s Robotics Laboratory, with kinematics and sensory feedback potentials similar to a human arm. Its main focus was to perform functionally in a previous unknown terrain, and one of its criteria is to be able to secure the safety of humans interacting with the whole robot structure (Grunwald et al., 2003). Its torque sensors in each joint allow stiffness and impedance control.

### 3.3. Measuring the Effort

This research addressed the who to imitate question in imitation learning (see Dautenhahn and Nehaniv (2002)) with the goal of “who” as:

“To assess whether a humanoid robot is capable of deciding who is a good teacher by imitating the movements of human teachers in an interaction game.”

The outcome of this goal is to offer a measurement of teaching performance of the teachers during the interaction. This measurement can be useful for the robots to focus on one partner (human) when facing more than one partner in an interaction session.

To achieve the goal, this thesis investigates the elements that can be measured by the robot. Later on, through an experiment, the measurement was compared to the perceived engagement seen by a human.

### 3.3.1. Definitions of Effort

The Cambridge English dictionary defines effort as the “physical or mental activity needed to achieve something”<sup>7</sup>. Merriam-Webster determines it as the “energy to do something”<sup>8</sup>, whilst the Oxford dictionary specifies effort as “a vigorous or determined attempt”<sup>9</sup>. Therefore effort, as we can define, is the *measurement* of how much *energy* we exert to an attempt to achieve a (certain) result, and “a lot of effort” is commonly perceived as something done through hard work with a satisfying result. From various definitions, effort then relates to how much energy is expended.

According to Feldenkrais (1972), the effort is related to how familiar the person is to a task. If a person “masters” a certain action or skill, then conscious efforts needed to do that certain action decreases. Therefore similarly, if someone is relatively new to a task they’ve never done before, more effort is required to achieve said task. Usually, the more effort is required, the more energy the person has to exert to achieve a wanted result. If a person is doing a task they are rather familiar with and requires less energy to achieve, they may use less effort. Vice versa, if a person is doing a task they’ve newly encountered and it also requires a large amount of energy to be exerted, the effort needed to achieve said task then sharply increases.

On the other side, people also relate effort to the reward. In the real world, it is not only for purposes of self-satisfaction, it is also employed as a metric in work. Large efforts at work are often exchanged for societal rewards: money, esteem and status (Siegrist, 1996). In societal perspectives, high-effort exchange with low rewards (such as inadequate payment)

---

<sup>7</sup> <http://dictionary.cambridge.org/dictionary/english/effort>

<sup>8</sup> <http://www.merriam-webster.com/dictionary/effort>

<sup>9</sup> <http://www.oxforddictionaries.com/definition/english/effort>

is often met with recurrent feelings like anger or frustration. It, therefore, shows that, usually to humans, if a person exerts an amount of effort that is not worth the result or the reward, it may result to the individual thinking that the effort exerted was not “worth it”. Especially as high-effort and low-reward imbalances greatly (in terms of the reward “not worth” the effort exerted) and may be particularly stressful.

### **3.3.2. Speech**

By definition, “speech” is the auditory/vocal medium that humans typically use in order to project language, and “language” is, therefore, the system that represents communicating conceptual structures (Fitch, 2000).

As previously discussed, human speech is usually accompanied by vocal cues, head movements, gestures and facial expressions. These behaviours are usually to add effect to the speaker’s context and to further aid understanding of the speaker’s spoken text (Graf et al., 2002).

#### **3.3.2.1. Energy Element in Speech Production**

Speaking typically involves moving the jaw and tongue and is usually accompanied by acts such as gestures during speaking. According to Searle (1969), speech production incorporates the usage of vocal cues such as tempo, pitch, tone and volume, and they may be accompanied by gestures, body postures, head nodding or eye gaze. How speech is perceived is dependent on vocal cues and nonverbal cues displayed, and often people may take account of what sort of relationship they hold with the speaker, as well as their status (such as formality), how they choose and present their words, and what context they are in.

Fadiga et al. (2002) present a study of how speech is perceived using transcranial magnetic stimulation, with the hypothesis that the “listener only understands the speaker when their articulatory gestures are activated”. They demonstrated that there is an increase of motor-evoked potentials recorded from the listeners’ tongue muscles when they are listening to a speech, especially when the words involved strong tongue movements when pronounced.

Graf et al. (2002) studied recordings of several hours of speech and measured their main facial features. The study found that, although it largely varies from person to person, head and facial movement patterns strongly correlate with the text’s prosodic structure. Angles and amplitudes of the speaker’s movements vary widely—however, the timing of these movements are surprisingly consistent. During prosodic utterances, eyebrows often rise, sometimes accompanied by head nods.

In a study of speech motor development (Smith, 2006) the authors discuss speech motor performance in terms of linguistic goals. Would there be a correlation between speech movements with actions such as chewing or breathing—which also involves movements of the jaw?

Studies have led to the conclusion that speech movement is not just a simple charting of distinct units of sound, phrases, or syllables. Producing a given sound or syllable may not necessarily mean that the acoustic characters that are present when spoken output is the same as another (Smith, 2006). However, infants aren't born with these complex mappings and it takes years for the adult systems to fully develop (Smith, 2006).

Head movements are also strongly correlated with speech. In terms of perception, watching a speaker's lips can influence a perfectly audible speech. Munhall et al. (2004) presents a study on participants that viewed animations of talking heads that had movement recordings of speech-in-noise, where they have to recognise as many words as possible in noisy listening conditions. All twelve participants that were recruited had no problems with hearing, speech, language nor vision. Given to the participants were recorded motions of the head and the face, produced by a male accompanied by recorded acoustics. The animation is synchronised with speech. The study concluded that participants recognised more syllables with natural head motion depicted by the animation than when it is either altered or eliminated completely. This means that head motions play a rather direct role in the perception of speech and is therefore also an effort when producing speech.

### 3.3.3. Gesture

This thesis focuses on arm movement to be used as gesture interaction with the robot. Arms are the upper limbs of the body and comprise the elbow joint and the shoulder joint. The arm extends to the hand in common usage and can be categorised into the upper arm, forearm and hand, and the corresponding bones and muscles in the shoulder are counted as part of the arm<sup>10</sup>. The hand especially is the most intricate mechanism of complexity, with capabilities to function to do simple actions we take for granted in our day-to-day life. The arm correlates very much to hand movement and gestures that a teacher may display.

---

<sup>10</sup> <https://en.wikipedia.org/wiki/Arm>

However, besides the hand, the arms seem to have no emphasized role in locomotion (Collins et al., 2009). Though to understand locomotion specifically, especially with movement of the legs, arm movement is a growing research interest (Donker et al., 2001). Collins et al. (2009) cited that arm movement may have been the result of shoulder movements during motion, therefore seen as a “pendulum” effect.

Pontzer et al. (2009) presented a study wherein they investigated control function of arms in human walking and running in order to test and clarify the proposition as to whether arms motions in walking/running are passive movements powered by the movement of the lower part of the body rather than just movements that are driven by shoulder muscles actively. The results of their study showed that “the arms act as passive mass dampers which reduce torso and head rotation, and upper body movement is primarily powered by lower body movement”.

### **3.3.4. Methods to Measure Arm Activity**

We can utilise the same measurements and tools to measure arm movement in the same way as we measure common physical activity such as walking or running. Here are some methods that can possibly be used to measure arm movement.

#### **3.3.4.1. Kinect**

3D depth cameras were first introduced and built to revolutionise gaming and how people perceive and experience entertainment, enabling them to interact with their games with their physical bodies instead of a controller (Zhang, 2012). The Kinect sensor lets the computer see the physical world in 3D and interpret person movements.

Kinect sensors notably incorporate a depth sensor, a colour camera and a four-microphone array. Skeletal tracking by the Kinect is presented as joints for body parts (e.g. arms, legs, head, and shoulders). 3D coordinates represent each joint.

The advantages of the Kinect is that it is comfortable and does not require any device attached to a part of the body, therefore allowing the person free access without the hassle of putting on something and maintaining its position whilst they are moving.

#### **3.3.4.2. Accelerometer**

A validation study evaluated an accelerometer’s reliability to measure upper-extremity rehabilitation outcomes by monitoring arm movements of persons with chronic stroke. Participants of a study by Uswatte et al. (2005) wore accelerometers on each arm, the chest and the more affected leg. The validation study reports on a method to measure the

rehabilitation of arm use and the reliability of the accelerometers that measure constraint-induced therapy; a method used to improve and enhance arm usage for persons suffering a chronic stroke.

No participant in the study reported any injury and the study concluded that just two accelerometers alone are appropriate enough to assess whether the rehabilitation had any effect on arm function with those people who suffer from a chronic stroke.

Accelerometer-based pens have been developed to recognise handwritten digit and gesture trajectory. Wang and Chuang (2012) present an accelerometer-based pen that can measure accelerations of hand motions (like writing digits or making hand gestures) and wirelessly transmit the measurements for online trajectory recognition.

#### **3.3.4.2.1. Commercial e-Health Tracker**

The commercial e-Health Tracker, such as the Fitbit<sup>11</sup>, has a three-dimensional accelerometer that can measure activity data, the advantage of it being wireless and fairly easy to utilise. Takacs et al. (2014) present a study where they test the reliability of Fitbit One and has concluded that it is a valid and reliable device with error percentage below 1.3%. Whilst it also has the capabilities of wireless interfacing with mobile devices as platforms/apps grow in numbers, it also means that it is a relatively efficient tool for researchers to track physical activity in studies as its reliability and validity is relatively high (Diaz et al., 2015).

#### **3.3.4.3. Gyroscope**

Whilst it is much more of an electronic companion than something to measure physical activity, usage of gyroscopes may come useful in measuring not overall “arm movement”, but specific angular placements wherein a research needs to specifically map a hand/arm angular position when exhibiting a certain behaviour, like a pose, as the usage of gyroscopes are more precise in angle measurements (Motoi et al., 2003).

##### **3.3.4.3.1. Alongside accelerometers**

Frequently a pair, it is suggested that accelerometers are used alongside gyroscopes to increase accuracy in measuring (Luinge and Veltink, 2005). Orientation can be estimated from the combined efforts of using both signals of gyroscopes and accelerometers (Luinge

---

<sup>11</sup> <https://www.fitbit.com/uk>

and Veltink, 2005). When coupled with gyroscopes, they can reduce both false positives and false negatives (Li et al., 2009).

Wang et al. (2010) present a pen that includes a triaxial accelerometer sensor, where the generated signals produced from writing can be transmitted to a computer wirelessly. Experimental results of the pen (IMUPEN) proves its advantages of its portability without limitations and without needing any external device and can reduce integral errors and reconstruct movement trajectory. The device has been proven to be effective and has valid results.

#### **3.3.4.3.2. Alongside Kinect**

As Kinect focuses on skeleton recordings and orientation, movement and placement of the body, gyroscopes are utilised to record angular velocity, especially the limbs (and therefore ultimately the arms) (Gabel et al., 2012).

### **3.3.5. Categories to Take Account When Measuring Physical Activity**

#### **3.3.5.1. Age**

Physical activity declines with age. Sallis (2000) summarises that physical activity declines especially in the teenage years (13-18), and was emphasised as the most prominent decline in physical activity. Physical activity declines more slowly at the stage of adulthood. Trost et al. (2002) evaluate the differences in age and gender differences, and their results support the fact that physical activity sharply declines during childhood and adolescence.

It is then possible to take age into account when predicting levels of physical activity, e.g. teachers of older age may exert lesser physical activity. If a person is older and therefore exerts lesser physical activity during teaching, then teachers of older age may induce less engagement as they display lesser engagement-inducing cues.

#### **3.3.5.2. Gender**

Gender is an important factor to take into account when measuring physical activity. Not only in terms of how much physical activity is expended but also take account body types and psychological effects. During elementary school, boys are more active than girls, and for the most part, boys are more consistent in physical activities (Trost et al., 2002). A surprising find discovers that males decline in physical activity more than females, particularly in youth (Sallis, 2000) although males are said to be more physically active.

In a study by Antoniou et al. (2006) in a school environment, female teachers were more likely to report higher degrees of burnout than male teachers. However, in a work engagement study by Langelaan et al. (2006), gender only slightly contributes to predicting work engagement.

### **3.3.5.3. Body Size/Type**

Body size/type becomes important when measuring the likelihood of someone having to do physical activity, and how much activity they typically carry out. For example, a certain body size or a body type may exert more physical activity in order to induce engagement than another body size/type.

Whilst it is also important to take account that the body size/type is considered in results, it is also important that how the activity is “measured” is taken into consideration. For example, where to place measuring equipment. Pedometers used in counting steps may fail to recognise the difference in height and leg length and step count may be influenced in a stride of someone with shorter/longer legs (Hills et al., 2014), and could cause an incorrect measurement in measuring physical activity.

## **3.4. Summary and Next Action**

The discussion presented here gathered research from the literature to identify possible measurable physical activity attributes. It paved the way by considering the effort as an inherited quality of an engaged teacher. Modalities such as voice, gesture, and tactile modality were explained for the evaluation where a teacher communicates in teaching a robot.

The chapter has described the complexity of physical elements regarding speech reproduction. On the other hand, the movement of arms is relatively easily measured by using some off-the-shelf sensors.

This study tried to link physical effort as a possible metric that could be used to measure the level of engagement of the teacher. The main experiment presented in this thesis recorded physical activity data in the first part of the main experiment (Chapter 5). The second part of the main experiment (Chapter 6) evaluated the data whether it can be used to measure the level of engagement of the teacher.

Before conducting the main experiment, another experiment was conducted to evaluate the preferences of a teacher using a communication channel in teaching to a robot. The result

was used for selecting what modality to use to teach the robot in the main experiment. The experiment that evaluates that modality is described in the following chapter.

## Chapter 4. Evaluating Modality Preferences

This chapter presents an evaluation of human participants' preferences of input modalities obtained from teaching a robot to mime to a rhyme. The robot will be taught by participants to move its arms using speech, visual gesture, and by physical manipulation. The results of this evaluation were used to select a particular modality to study in further detail in order to evaluate the level of engagement of the human teacher.

### 4.1. Background

It is desirable that in HRI research the robots feature multi-modal communication interfaces in order to facilitate natural communication. Humans often use multi-modal interaction in daily life as a way to communicate with each other—typically through speech, physical gesture, and eye gaze. We exchange information, we express our feelings, and we socialise using one or more of these modalities. As indicated by Perzanowski et al. (2001), in interaction with machines which harbour features and characteristics similar to that of a human, we also have a tendency to communicate in ways similar to that of human-human interaction, anthropomorphising with the machine.

Despite the benefits of natural interaction, developing a robot for HRI research which features multi-modal communication interfaces presents many difficulties. In order to be able to facilitate processing a range of inputs including but not limited to visual, audible and gestural cues, the robot's system needs a considerable amount of computing power. It also needs robust integration algorithms to make decisions in real-time about which inputs to consider for outputting an appropriate response through the robot's actuators. Integrating social queues to flow naturally throughout the interaction would also expend further processing power.

Developing such system would also require a substantial amount of time, which overall, could slow down the progress of research that is required for the robot to have a certain set of (tailored) autonomous behaviours. These problems can be minimized by using the Wizard of Oz (WoZ) (Steinfeld et al., 2009) approach to conduct experiments. With this approach, the limitation of the robot can be set aside and replaced by behind-the-scene controllers to produce behaviours for the robot which are perceived by participants as autonomous. However, the studies presented below *do not* take the WoZ approach.

In this research, the aim is to consistently evaluate the level of engagement of a human when teaching a robot. In doing this, the study uses real-time *autonomous* behaviours for the robot (in contrast to the WoZ approach) to capture the dynamics of the interaction. Thus the human behaviour is influenced by a real robot that has certain limitations and not a perfect, but typically much closer to, imitation of human behaviour. However, the robot's behaviour is also consistent and controlled by the same program, not depending on a human that might be varied in controlling it. In doing so, the study in this chapter will evaluate the user preferences of input modality for the teachers to teach the robot. The research will then continue to evaluate the level of engagement of the human teacher by using a particular modality that is considerably preferred by the human.

## 4.2. Related Work

To the best of the author's knowledge, there are few studies in HRI that investigate users' preference of modalities in interacting with robots. There are, however, some studies in HCI (Human-Computer Interaction) that measure the modality preferences of the users when interacting with computers. As suggested by Kiesler and Hinds (2004), and Breazeal (2004), existing studies in HCI provide a large source of information and inspirations for research in HRI. As such, the study discussed here has taken related research from HCI into consideration. The discussion below presents firstly the existing work in the HCI domain and then is later followed by related works in the HRI domain.

The experiment "Put That There" by Bolt (1980) is widely considered a pioneering demonstration that first showed the value and opportunity of multi-modal interfaces over uni-modal interfaces in HCI. The experiment was conducted using speech and pointing gestures, and its goal was to use both modes as command cues to draw a map. The research was one of the first to take the use of "pronouns" into consideration (e.g. this and that) for use alongside a set of commands where both voice and gesture were necessary.

The multi-modal interface gave rise to the important question of when the system is capable of multi-modal interactions, will the users utilise this ability to interact multi-modally? Oviatt (1999) indicates the answers by discussing the ten myths surrounding multimodal interactions that gave a useful guidance in building multimodal systems for researchers. The conclusion: while a system may be capable of multimodal interaction the users did not always interact multi-modally. Instead, users have the tendency to switch between uni-modal and multi-modal interaction depending on the type of action being performed.

The answer above was supported by an earlier study by Oviatt et al. (1997). They found that participants used 86% of their time to give multi-modal commands when giving spatial location (navigating a map in order to move, add, modify or calculate the distance between objects). Inversely, for tasks which did not require selection or spatial information, such as printing maps, participants interacted multi-modally less than 1% of the time. This data indicates that multimodal interactivity might not always be necessary and depends on the type of task the participants are engaging in.

Later, Oviatt et al. (2004) conducted another map navigation experiment with a Wizard of Oz approach and summarised that the users would interact uni-modally or multi-modally depending on the complexity of the task. Users tended to interact multi-modality for more challenging tasks, such as placing an object where the location was intersected with multiple objects. However, once they were familiar with the task, they would begin to use one particular modality.

Schüssel et al. (2013) conducted an experiment to compare speech, gesture, and touch modalities. This experiment measured what modality was used and combined by the users to complete a task. The experiment was also conducted in a Wizard of Oz approach and the task was to select graphical icons on a computer monitor. The usage results of the modalities were: touch (63.2%), speech (21.6%), gesture (11.2%) speech+gesture (3.6%), speech+touch (0.5%). None of the participants used speech+gesture+touch at the same time.

Carbini et al. (2006) observed users' preferences in a storytelling game. The task for each user was to compose a coherent story from a set of objects on a computer screen. The research involved adult and child participants and concluded that children could easily interact using speech and gesture compared to adults. This accounted for the users' preferences on modalities depending on the age range of participants, with children preferring to use gesture and speech. The results of the full dataset were: gesture (45%), speech (5%), and gesture+speech (50%).

All of the research cited above was conducted in HCI domains where users interacted with computers. This current research is focused on the interaction between humans and robots. The studies presented below are more closely related to the research concerning HRI.

Research by Khan (1998) surveyed 134 participants on their preferences of interaction modalities with a robot. One of the questions in the survey asked the participants their

preferred methods of communicating with a service robot, for example, to take care of clothes on a couch or when the robot is to inform a user when a task is completed. The participants could choose more than one preference; the following statistics show the percentage of participants who chose them for each modality. Results showed that speech was predominantly the most chosen modality (82%), followed by touch screen (63%), gestures (51%) and typing commands (45%). However, the results of the conducted survey were taken without the participants actually having to interact with a robot at all, where the participant did not experience the actual interaction with the robot. However, this may help with further studies if multimodal service robots were to be built.

Salem et al. (2012) conducted a study which compared the preference of modalities in HRI and used a real robot. Their research differs with the current research as they investigated the robot's output-side of the multi-modal interface, instead of the input-side. They examined the participants' perspectives regarding a robot when the robot provides information to the human uni-modally (voice only) and multi-modally (voice+gesture). The study discovered that the robot was judged more positively if it displayed non-verbal behaviours such as hand or arm gestures alongside speech, even if they did not match the speech utterances semantically. This showed that users had a greater preference for interacting with robots that are much more "like them" because of the likeliness and familiarity of human communication between the human and the robot.

Humphrey and Adams (2008) also conducted research which is relevant to the current study. The study focused on how different visualisations of a navigation compass of a tele-operated robot affected the interaction between a human and the robot. In the experiment, the participants' preference of the compass visualisation was one of the things which were measured. The two different compass visualisations were compared: top-down (from the top of the robot) and world-aligned (parallel). The results showed that the world-aligned compass provided faster task performance compared to the top-down. However, the top-down compass provided perceived situational awareness and was seen to be easier to work with, and was preferred by the users. However, there was no significant difference in which visualisation had a higher preference.

Profanter et al. (2015) conducted an experiment using the Wizard of Oz approach with thirty participants. The participants programmed an industrial robot system and were given the following input modalities for defining task parameters: touchscreen, gesture, speech and 3D pen input. After the experiment, the participants filled out a questionnaire which

included questions about their opinions on the different input modalities. The results showed that most users preferred touch and gesture input over the 3D tracking device input. Speech was the least preferred input modality.

From the research cited some conclusions were derived and used to guide this study. They are discussed in the following sections.

### 4.3. The Study

This study is inspired by the challenge of creating a multi-modal interactive robotic system to investigate users' preferences of input modalities in teaching a robot. The study was designed to ask users to experience three different modalities whilst delivering the same instructions in teaching a robot: voice, gesture, and moving the arms.

The study presented in this paper is built from two main observations from the related works. These related works come from the HCI research domain, where humans interact with computers. Our research puts them in the perspective of HRI, where humans interact with robots, in order to explore and develop the possibility of them being applied to the domain of HRI.

The primary observation comes from the work of Oviatt (2004). He concluded that simple interaction tasks can be conducted sufficiently and effectively by using a unimodal system alone. Based on this, our research investigated further the modality comparison by conducting an experiment that asked users to do a simple task. Furthermore, the experimental outcomes will be used to select what modality will be chosen for onward development in this research.

The secondary observation comes from the prior studies by Schüssel et al. (2013), Carbini et al. (2006), Khan (1998), and Profanter et al. (2015) in evaluating input modalities. Their works indicated that preferences will vary depending on the task. Inspired by those studies, our study will compare the preference of different input modality in doing one particular task.

This research is intended for developing an autonomous humanoid robot that has the ability to perform in real-time, live, multi-modal interaction. The developed system has the capacity and capability to detect voice commands and interpret gestures and touch. Every process ran together in real-time. In the discussion section, this chapter presents the

comparison of user preferences for the three different input channel modalities when they are interacting with and instructing the robot a set of arm movements.

The basic idea of the experiment for this study was to develop a robot that is able to learn and be taught to create a set of movements (dance) while following a music. However, the idea had a disadvantage because it required a robot that can move fast enough to perform the dance movements. In order to overcome this challenge, the dance was simplified to a simple mime task and the music was reduced to a nursery rhyme. With these alterations, the experiment transformed into teaching the robot to mime following a nursery rhyme. The robot could be taught to move its arms using voice commands, by the users' gestures, and by physically moving the robot's arms.

The experiment was designed to run non-intrusively, so the participants were not required to wear special apparatus, such as gloves or markers. The participants were also not required to wear microphones or headphones for speech and voice commands. Instead, the voice command system used a speaker-independent system, so the participants were not required to be trained before commencing the experiment.

#### 4.4. The Task

In designing the experiment, we imagined a task to facilitate an imitation game that allows for comparison of the input modality preferences. The idea was implemented by teaching a robot to mime a nursery rhyme. The rhyme was "Hickory Dickory Dock". The lyrics are:

*Hickory dickory dock*

*The mouse ran up the clock*

*The clock struck one*

*The mouse ran down*

*Hickory dickory dock*

The rhyme has five lines, which can be translated into a task to teach five movements. Using this rhyme, the participants were given a task to teach the robot to make one movement for each line. Furthermore, the task was designed to enable the participants to instruct the robot to move through different modalities. For this study, three modalities discussed in the sub-sections of Section 3.2 were chosen as input modalities to teach the robot. The modalities were:

1. Voice
2. Gesture
3. Tactile

We envisaged robot movements that can be taught by using those modalities. Moving the arms, as part of the tactile modality (Argall and Billard, 2010), inspired the activity to be taught. Any of these modalities can be used to teach arm movements. With this factor as a consideration, this research chooses the following five arm movements to be taught for comparing the input modality preference. The movements were:

1. Open both arms
2. Move left arm up
3. Move left arm down
4. Move right arm up
5. Move right arm down

The experiment was designed with the participant teaching the robot by only using one modality in a session. With this design, the user utilised different modalities in different sessions (within the same experiment). After the experiment, the participants were given a questionnaire to compare the preferences of different input modalities.

#### **4.5. The Robot**

This research uses KASPAR (Dautenhahn et al., 2009), a child-like humanoid robot (Figure 4.1). The KASPAR robot was developed by the Adaptive System Research Group at the University of Hertfordshire. The particular robot that was used for this study is the third of generation KASPAR. The robot had 17 degrees of freedom (DoFs) and an internal PC to run the robot autonomously.

The KASPAR robots have mainly been used in research for children with autism. The robot was designed intentionally to have minimal face features that allow children with autism understand the facial expressions. Nevertheless, the KASPAR robots have also been used for typically developing children in other application areas (Wood et al., 2013).



**Figure 4.1** The KASPAR robot for this study

#### 4.5.1. Software Development

This section and the following sub-sections discuss the development of the robot to support this study, and mainly discuss the software aspects of the robotic software. The software was developed from scratch and used some off-the-shelf robotics software libraries to support the required features.

Originally the robot had its own software to operate the robot. It served different features and purposes than that of this study. The internal PC has a relatively low amount of computing power and is only able to run basic autonomous behaviours. The study outlined in this chapter required robotics software that would be able to recognize voice, track human body movements, and allow the robot to comply with external physical movements by humans moving the robot's arms.

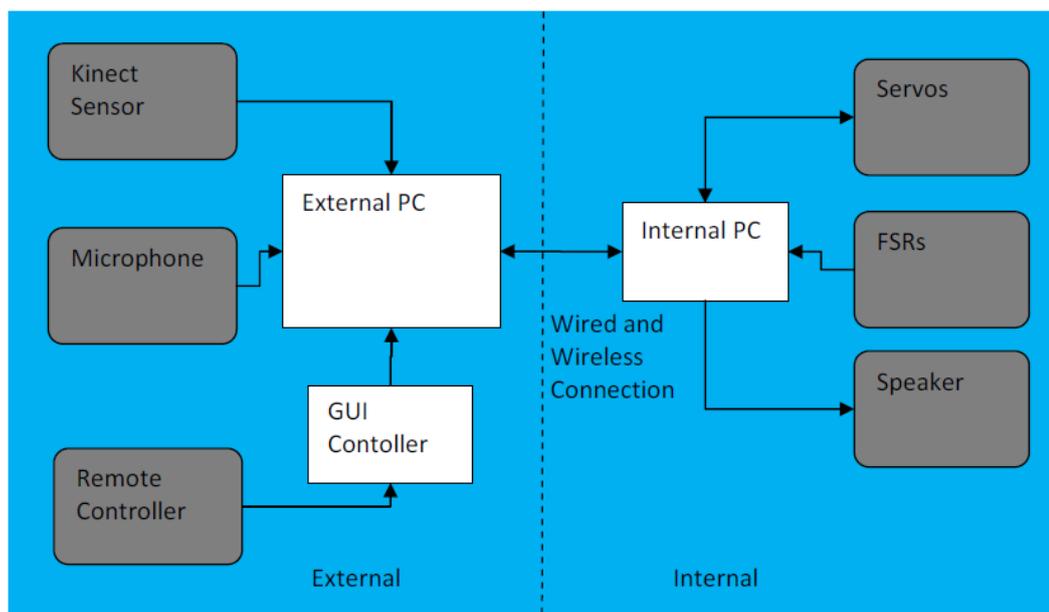
For this study, the software was developed to suit the hardware specifications of the robot, thus eliminating the need for hardware development. The author started by looking at the existing hardware-software connection which has two connection modes: (i) direct connection, and (ii) Ethernet.

The first connection mode, direct connection, was a legacy feature from the earlier generation of KASPAR which used an external PC to control the robot. In this mode, two

physical serial port connections are needed. The first port is used to control the servos i.e. for moving the arms. The second serial connection is used to deliver touch pressure data from the skin patches. In addition to these 2 serial cables, an audio cable is needed to connect an internal speaker to the external PC. In this direct connection mode, the internal PC on the robot is not used.

The second connection mode was a new feature to the KASPAR generation that is used in this study. In this mode, the original software uses an internal PC to control the servos and stream the skin data through an Ethernet connection. The sound is produced by sending commands to a handling program in the internal PC.

In terms of computing flexibility, the direct connection mode allows the use of an external PC with enough computing power for any required capabilities. In terms of connectivity, the Ethernet mode is more flexible as this mode only needs one Ethernet cable connection. Furthermore, the robot has a built-in WiFi feature that also allows for a wireless connection.



**Figure 4.2 System architecture**

As introduced earlier, the experiment in this study requires a robotics software that is capable of speech recognition, visual tracking, and comply with external physical movement. With the two potential connection modes, it is possible to develop new software that runs on an external PC to provide the input modalities stated. Based on the above condition, the software needs to run several processes on more than one computer. In brief, the architecture of the system is shown in Figure 4.2.

The robotics software utilised a robotic middleware to establish a data communication link between each of the software modules running on different computers. Initially, ROS (Quigley et al., 2009) was used as the middleware. However, because of the limited resources of the internal PC, the author utilised the YARP middleware (Fitzpatrick et al., 2014) and subsequently used it to develop the robotics software for this research.

The internal PC runs the Linux operating system (OS) and the software modules that run on it are developed in C and C++. It uses the eSpeak<sup>12</sup> text-to-speech engine for speaking. The software runs eSpeak with options to set the pitch adjustment to 60, the word speed to 120 words per minute, and using voice "en+f4" (a standard British English variant).

The external PC runs the Windows OS. Connection to YARP from C# based modules in this external PC side is provided by making a library from the YARP source code using SWIG<sup>13</sup> to support C#.

The initial development of the software established a platform that supports the development of independent software modules to implement the required features. With the support of robotics middleware, the software can run on several computers to cope with the computing requirements, especially for the visual tracking and voice recognition system. After establishing this platform, the software development then moved further to fine tune the features of the system and providing the graphical user interface (GUI) for the experiment such as discussed later in Section 4.5.5.

#### 4.5.2. Physical Compliance

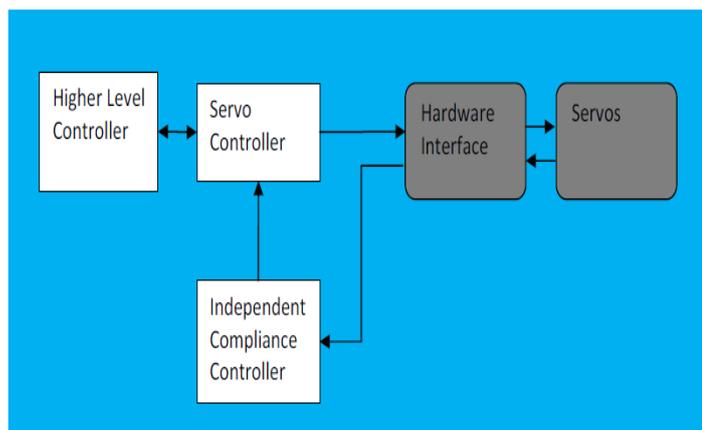
A software module that processes touch sensing was developed in this study to be part of the robot's features. The robot has a number of sensor patches in several locations that can detect contact forces. These are force sensitive resistors (FSRs). These can be used to detect when a user touches a certain location where the patch is located and how hard the force is. However, this particular feature was actually not used because it needed the participants to touch the specific location where the patches were located. Instead, another compliance system was developed to allow the users to move the arm at any location during a live interaction.

---

<sup>12</sup> <http://espeak.sourceforge.net/>

<sup>13</sup> Simplified Wrapper and Interface Generator (SWIG), <http://www.swig.org/>

A physical compliance module was developed to allow the human user to physically move the robot's arms during the interaction without causing damage to the motors. Initially, the software utilised the existing sensor patches on the robot to sense the physical touch. Later on, it was considered limiting for the interaction as the participants have to touch the exact location where the sensors are located. Therefore, a compliance module was developed to read external force sensed by the servos. By using this method, the participants could hold any part of the arm in order to move it, e.g. the users could move the arms by moving the upper arm or moving the hand. The latter requires smaller force because it is further away from the shoulder joint.



**Figure 4.3 Compliance mechanism**

The block diagram of the compliance system is shown in Figure 4.3. The software module detects the external force by measuring the servo's load value. If it exceeds a certain threshold, the module assumes that the robot's limbs are being physically manipulated externally by the user. The module then adjusts the servo position to match the external force. The system works independently and can override any arm movement commands sent by the higher level controller.

Tan et al. (1994) summarized that human force control bandwidth needs to be around 20 Hz to have a smooth compliance. Unfortunately, in the current implementation, due to the limitation of the hardware, there was a significant time delay in the compliance controller's loop path. For each servo, it took around one second to respond to the external force. This meant that the control bandwidth of the servos only achieved 1 Hz. In this case, the robot's arms could be perceived as being slightly stiff to move. This limitation could bring negative effects to the human-robot interaction that it could, for example, make the participants hesitate to move the arms as they might be afraid to break the arms.

### 4.5.3. Voice Recognition

The developed system used the Microsoft speech recognition engine. It is a speaker-independent system so did not require training prior to the experiment. The experiment is prepared with a non-intrusive interaction in mind so that the users do not have to wear a microphone or headphones. For this, the system uses a directional microphone to listen to the user's voice. The microphone location was adjusted so the sounds coming from the robot (voice and mechanical servo movements) were less likely to disturb the user's voice.



**Figure 4.4 Red and blue markers**

Colour markers were placed on the robots' fingers (see Figure 4.4) to refer to the arms by colour instead of left and right. The markers' colours are red and blue. This was done because labelling the arms by colour was deemed to be easier for participants to use when they faced the robot.

For the experiment, the system was programmed to detect five different voice commands to instruct the robot to move its arms. The commands were:

1. "red up"
2. "blue up"
3. "arms open"
4. "red down"
5. "blue down"

As suggested in the name, "up" and "down" commands will instruct the corresponding red or blue arm, as stated by the given command, to go up or down. The "arms open" instruction will prompt both of the robot's arms to open wide to the robot's left and right.

The system was only able to detect one particular command at a time. After saying a command, the user was expected to wait for the robot to respond before giving another command.

#### 4.5.4. Visual Tracking

To track the user's arm movements, the developed system uses the Microsoft Kinect sensor. The built-in Kinect software development kit (SDK) from Microsoft provided a skeleton representation of the user's position and pose.

The positions of the wrists were measured and interpreted as commands to move the robot's arms. The system was programmed so that it only detected five positions, which were equivalent to the five voice commands.

#### 4.5.5. Graphical User Interface

The robot's autonomous behaviour was controlled via a C# based GUI. It had several tabbed-panel interfaces to control and test the robot. The main interface, which is shown in Figure 4.5 was used in the experiment to display the automated progress of the interaction session. This interface also allowed testing of the interaction session manually.

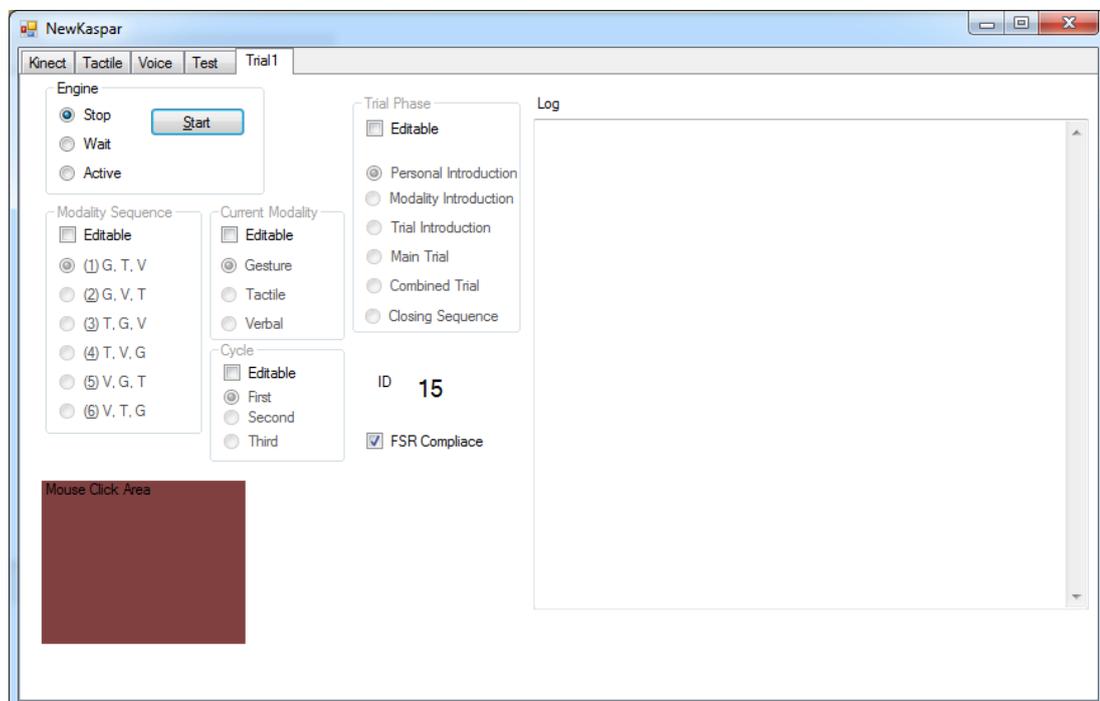


Figure 4.5 The GUI for the first study

## 4.6. Experiment Method

This section discusses the method used for conducting the first experiment in this study to measure the preferences of the human participants whilst teaching a robot.

### 4.6.1. Ethics Approval

Prior conducting the experiment, an ethics application document was prepared and submitted to the Ethics Committee at the University of Hertfordshire. The Ethics Committee approved this experiment with ethics approval number 1213/10. Later on, an ethics application extension was submitted and received the approval number a1213/10.

### 4.6.2. Target Participants

Initially, the experiment was designed to have two groups of participants. The first group was children aged 7 to 9 and the second was adult participants. The plan was to compare the difference between child and adult participants. Originally the author planned to run the experiment in a mainstream school near the University of Hertfordshire. This was so the children did not have to travel to the robotics lab in the university. The author had obtained a Criminal Records Bureau (CRB) certificate (now is called Disclosure and Barring Service, DBS) to work with children.

As all the preparations were made to conduct the experiment, it was later realised that due to the equipment being used in the experiment it was considered impractical to take the setup to a school. For example, the system needed a quiet environment for the voice recognition system to work properly. Therefore the plan to run the experiment with child participants was discontinued. The experiment was then carried with adult participants only and run in the robotics lab at the University of Hertfordshire.

### 4.6.3. Tester Participants

As part of the preparation, two colleagues in the lab were invited as alpha-tester participants. This was to get initial feedback about the experimental setup. After testing it, the setup was considered suitable to run the experiment and then the experiment was carried out with beta-tester participants to fine tune the setup and ensure that everything was ready.

Six more colleagues in the lab were invited as beta-tester participants. This was to check the performances of the voice recognition, visual tracking, and the robot's physical compliance. The result of each trial was used to improve the robustness of the system performance.

#### 4.6.4. Equipment and Setting

Figure 4.6 shows the physical layout of the experiment. The robot was located on a table, "sitting". The participants were facing the robot and sitting in a chair that could be moved closer to or further away from the robot.



**Figure 4.6 Experiment layout**

The KASPAR robot and the robotics software discussed in section 4.5 were used to run this experiment. The external PC was a laptop with an Intel Core i5-2450M processor.

A Kinect sensor was placed next to the robot and was used to visually track the human participants' movements. The sensor used an additional zoom lens (Nyco<sup>14</sup>) to reduce the required capture area of the sensor. This lens was used to make sure that if the human participants stood up, it would still be within the capture range of the sensor.

---

<sup>14</sup> <http://nyko.com/products/xbox-360-zoom-for-kinect>

A unidirectional microphone was used as input for the voice recognition system. The microphone was arranged in a location to ensure the human was still within the capture area of the microphone while the robot was outside the capture area. This is so that the sound from the robot would have less impact on the voice recognition system.

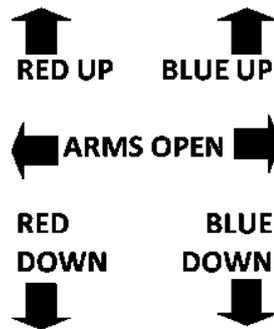


Figure 4.7 Instruction sign

Next to the robot, there was also an instruction sign for the voice commands. An image of the sign is shown in Figure 4.7. It was placed on the table to help the users to remember the five instructions to move the robot's arms. The instruction sign indicated arrows to help the users and they reflected the direction of the arms movements.

Video cameras were used to record the activities while the experiment sessions took place. One camera was placed near the robot, mainly to record the human participants. Another camera was placed on the right-hand side behind the participant, mainly to record the robot and also to record the human from behind. The recording from this camera could be used to show the activity in the session without revealing the face of the human participant.

#### 4.6.5. Interaction Scenario

In the main trial session, the participant interacted one-to-one with the robot, without using any device or gadget to interact with the robot. The robot asked the human participant to teach the robot to mime to a rhyme. The rhyme being used in this experiment was called "Hickory Dickory Dock" (see section 4.4). The robot says the rhyme, one line at a time. The participant then had to instruct the robot to move its arms for each line of the rhyme. There were a total of five lines in the rhyme, with five movements to be taught to the robot.

There were three robot-input modalities for teaching the robot. They were voice (spoken instruction), gesture (show the movement), and by physically moving the robot's arms. The experiment was designed that the participants used only one modality in one particular sub-

session. This meant there were three unimodal sub-sessions wherein each sub-session they used only one particular modality.

In the voice sub-session, the robot needed to hear the complete instructions. Therefore, the robot would follow the instruction after the participant finished stating the instruction.

In the gesture sub-session, the robot could track the movement of the participant at any time provided the participant was within the tracking area. The sensor was prepared so that the participant would still be within the tracking area regardless of whether they were sitting or standing up. In this sub-session, the robot followed the instructions as they happened (immediately) when the participant moved their arms.

In the physical movement sub-session, the participants moved the arms of the robot by physically manipulating them. The robot was developed to have a compliance mechanism that allowed the participants to move the arms without breaking the robot's mechanics. Since the participants had to physically move the robot's arm, in this sub-session the participants had to move close to the robot. The chair was free to move so the participants could still sit in the chair while interacting with the robot.

In addition to the three unimodal sub-sessions, there was another the sub-session. The fourth sub-session used the only possible combination of the modalities which were of voice and gesture. This was because when the participant moved close to the robot in order to move the arms, the Kinect sensor could not track the user. Also, when close to the robot, the voice of the robot would interfere with the voice from the user which would cause the voice recognition system to detect invalid commands.

In the fourth session, the robot asked the participant to teach either by telling, showing, or in combination. The robot would follow any instructions given by the participant.

At the end of each sub-session, the robot displayed the whole movement that had been taught using the related modality. It would say the rhyme and move its arms at the same time to each line of the rhyme.

#### **4.6.6. Procedure**

In the experiment, the trial was run individually with a single participant for each trial session. The experiment session was expected to be completed in 30 minutes.

Overall, each participant took part in four main sessions. The flow of participant's activities in these sessions is shown in Figure 4.8. The following list explains each of the sessions one by one.

### 1. Pre-trial session

Before starting the experiment, the participants signed a consent form, and completed a pre-trial questionnaire which requested demographic data and the participants' familiarity and expertise on robotics.

### 2. Introduction session

After filling the pre-trial questionnaire, the participant was then invited to sit on the chair in front of the robot. After that, the investigator introduced the robot to the participant. The robot then raised its right arm (or "red arm") and asked the participant to come near to the robot and shake its hand gently. This was to make the participants aware that they could physically move the robot arms, even though it felt slightly stiff.

Next, the participants were introduced to the nursery rhyme and told what to do in the main trial session. After that, the participants were introduced to the input-modalities provided on the robot. Then, the participants were asked to practice making the robot move using each modality.

During this introduction session, the robot was operated semi-autonomously using a wireless clicker. The clicker would advance the sequence of the sub-sessions within the introduction session. At the end of the introduction session, the participants were told that the following session is the main trial and the robot would run fully autonomously.

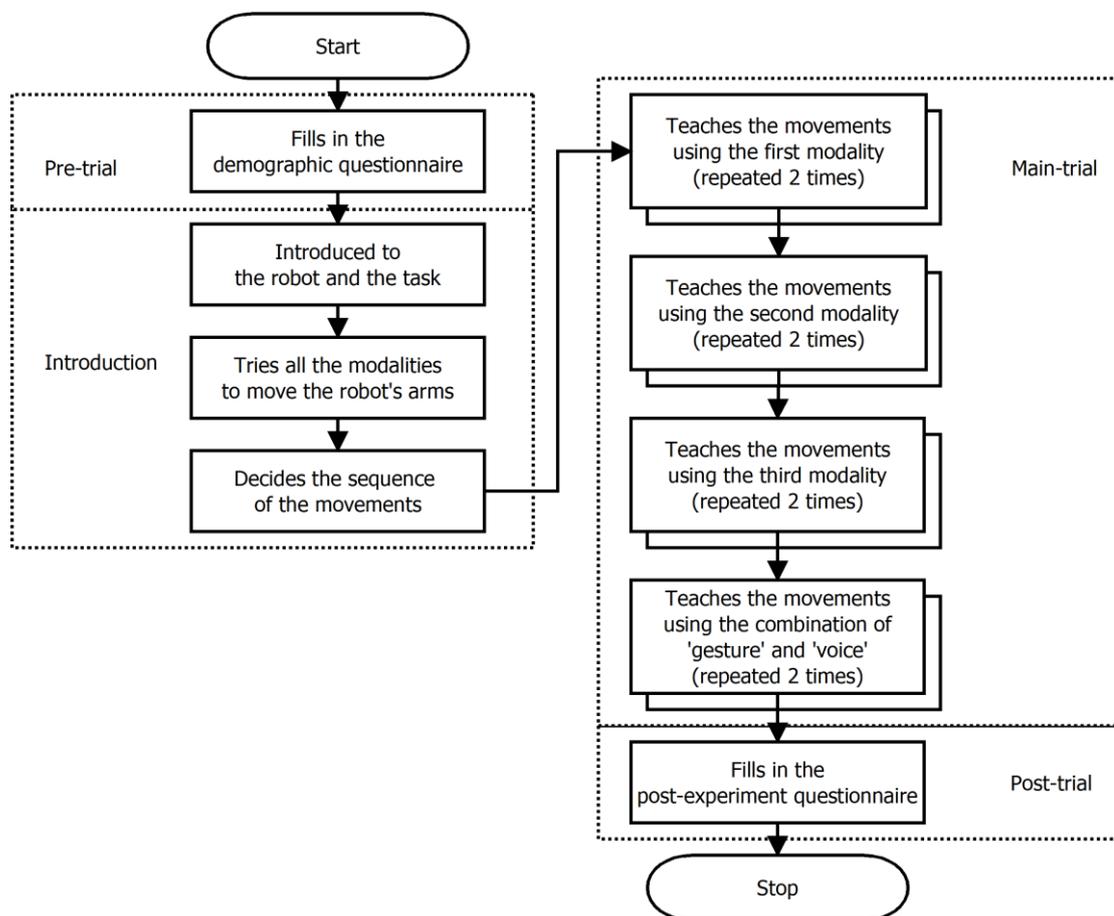
### 3. Main trial session

In the main trial session, the participants were left alone to interact with the robot one-to-one. The robot told the participants what they should teach, one movement for each line of the nursery rhyme. The participants used only one modality for one complete rhyme in one sub-session. In total, three uni-modal sessions were conducted one after another in a prearranged sequence so the next participant would have a different sequence. All the participants had the same fourth sub-session which was using a combination of voice and gesture to teach the robot. The detail of the process of teaching using one particular modality is shown in Figure 4.9. In this case, it is about the "moving the arms" modality. Similar patterns also apply for other modalities.

The robot ran fully autonomously in this session. After each sub-session above, the robot demonstrated the whole movement that had been taught using the related modality. The robot said the rhyme and moved its arms at the same time to each line of the rhyme. The investigator stayed in the same room reading a book and sat back-facing the participants at a table without any computer or electronics devices. The participants were told that in case of emergency or if they wanted to stop, they could notify the investigator at any time.

#### 4. Post-trial session

After completing the main trial session, the participants filled in a post-trial questionnaire. The participants were then asked verbally whether they had any comments they wanted to express regarding the experiment.



**Figure 4.8 Activity flow of the participant in the “Modality Preferences” experiment**

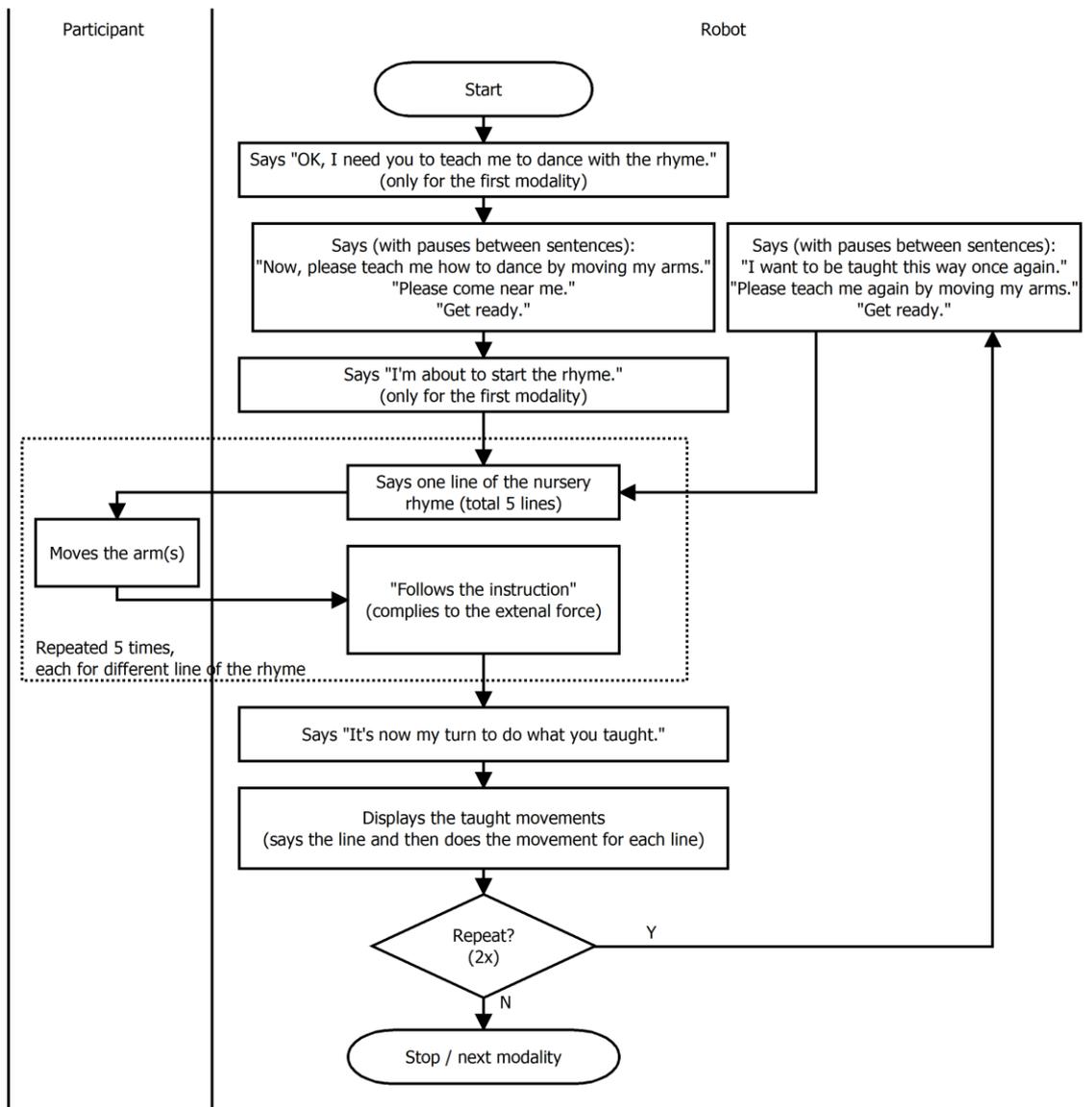


Figure 4.9 Activities in teaching the robot by moving the arms

#### 4.6.7. Dependent Measurements

All the participants completed a questionnaire once they had finished interacting with the robot. This post-trial questionnaire asked four sets of questions. Each question used the Likert scale for the participants to rate their answer on a scale from 1 to 5 as shown in Figure 4.10.

1	2	3	4	5
---	---	---	---	---

**Figure 4.10 Answer boxes**

The questions were as the following:

1. Did you fully understand what instructions KASPAR said during the main session? (1 = not very well, 5 = very well)
2. In terms of effort, how did you feel about the different methods to teach KASPAR to dance? (1 = very hard, 5 = very easy)
3. In terms of enjoyment, how did you feel about the different methods to teach KASPAR to dance? (1 = least enjoyable, 5 = most enjoyable)
4. When KASPAR showed what it had learned, how well did you feel KASPAR followed your instructions? (1 = not very well, 5 = very well)

Every question from 2 to 4 had separate answer boxes for each interaction modality which were:

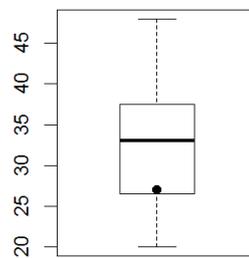
- A. "By speaking to it:"
- B. "By demonstrating using my arms:"
- C. "By moving the robot's arms:"
- D. "By a combination of speaking and demonstrating:"

## 4.7. Results

### 4.7.1. Participants

The participants were recruited from the university staff and students. The invitation was advertised verbally and they were given a link to an online scheduler (Doodle<sup>15</sup>) to pick the available time slots that were suitable for them.

The experiment was conducted with 16 participants. They were six females and 10 males aged 20 to 48 years old. In each gender category, one person was very familiar with robotic systems, while none had a prior knowledge of the robot setup that was used in this experiment. The boxplot of the age of the participants is shown in Figure 4.11.



**Figure 4.11 Age of the participants**

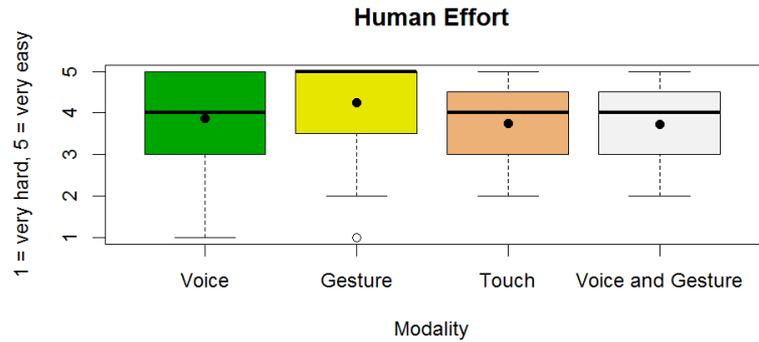
### 4.7.2. Data Analysis

For the first question of the questionnaire, that asked whether the participants fully understood what the robot said during the experiment, no participant selected a value lower than 4. The mean score was 4.56 (SD = 0.51).

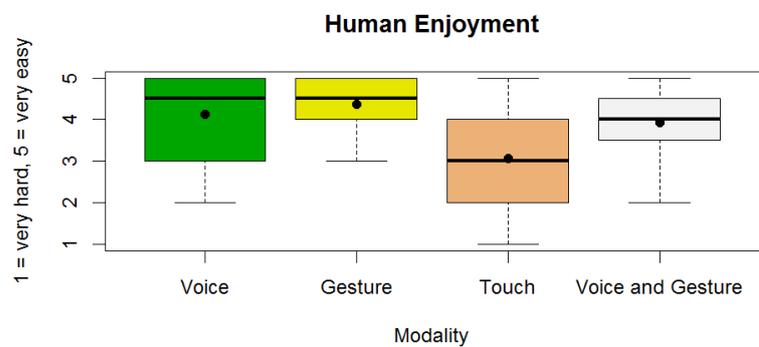
The questionnaire result on the effort to teach the robot to dance is shown in Figure 4.12. The data was checked using one-way repeated-measures ANOVA. The result was  $F(3,42) = 0.848$ ,  $p = 0.476$ , which meant there was no significance. The result suggests that no particular modality is perceived as harder than the others.

---

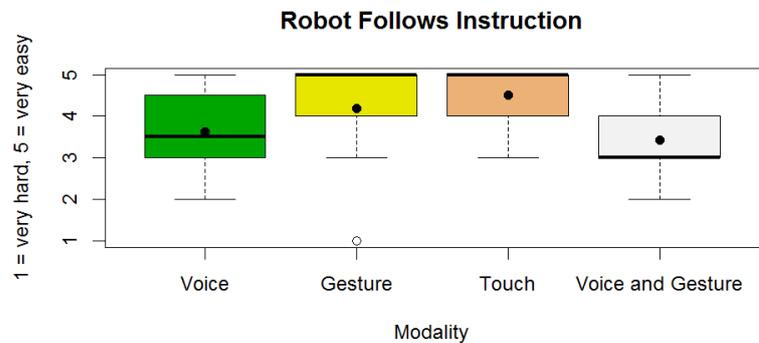
<sup>15</sup> <http://doodle.com/>



**Figure 4.12** Questionnaire result on human effortlessness



**Figure 4.13** Questionnaire result on human enjoyment



**Figure 4.14** Questionnaire result on different instruction modalities

The result that is shown in Figure 4.13 shows participants' perceived enjoyment of conducting the task for each modality. The touch modality received the least enjoyable rating. The statistical analyses indicated a significant difference in preferences,  $F(3,42) = 6.461$ ,  $p = 0.001$ . The pairwise comparisons results indicated that there was a significant difference ( $p = 0.008$ ) between participants ratings for gesture ( $M = 4.4$ ,  $SD = 0.74$ ) and touch ( $M = 3.07$ ,  $SD = 1.28$ ) interaction modalities.

Finally, Figure 4.14 shows the participants' perception of the robot's ability to follow instructions. The difference was marginally significant,  $F(3,39) = 2.56$ ,  $p = 0.069$ . The pairwise comparisons showed a preference ( $p = 0.011$ ) for touch ( $M = 4.43$ ,  $SD = 0.65$ ) over voice+gesture ( $M = 3.43$ ,  $SD = 0.85$ ).

## 4.8. Discussion and Conclusion

This research has investigated a robotic system that can be taught movements to follow a nursery rhyme. The robotics software was developed from scratch and using some off-the-shelf program libraries. The software development is only presented briefly above in order to illustrate its necessity as part of the experiment; however, the developments may be better to be presented as a detailed technical publication. Three modalities were provided as input channels to give information to the robot as commands to move its arms. They were voice, gesture, and touch. Two modalities were provided as output channels: voice and gesture. The robot operated autonomously during one-to-one sessions. The robot had touch-compliance which allows humans to physically move its arms into the desired pose. The system supported an integration of multiple modalities through a TCP/IP-based inter-process communication mechanism. The experiment was conducted with adult participants.

### 4.8.1. Summary of Findings

The research findings indicated that being given a task which was to teach a robot to mime actions that follow a nursery rhyme, there was no statistically significant difference in preference ratings regarding human effort.

In contrast, there were favourable preferences regarding the human enjoyment. The touch modality was the least preferred and the gesture modality was rated the highest. The author argues that the touch modality scored lowest due to the participants worrying about breaking the arms of the robot. This was because the compliance only controlled the arms compliance at a 1 Hz cycle rate instead of 20 Hz (see Tan et al. (1994)).

For the robot's perceived ability to follow instructions, touch modality received the highest rating. The combined voice+gesture modalities received the lowest. This could be due to the robot only performing the instructed action after the voice command had completed, while the action after the gesture mode interaction was followed immediately. However, they were not statistically significant at the 5% level, and only indicated a trend towards a higher mean preference to the touch modality.

#### 4.8.2. Relation to Literature

The study presented in this chapter evolved around the observations from several studies from HCI and HRI. While this current research focuses on HRI, it derived two main observations from both HCI and HRI research as discussed earlier in Section 4.3.

In general, without considering the task, the results are in contrast to the result of research by Schussel et al. (2013), Carbini et al. (2006), Khan (1998), and Profanter et al. (2015) that in this current study the participants did not have a significant preference for one particular modality. However, this contrast indicates an agreement with Oviatt et al. (2004), namely that for certain tasks humans can communicate with robots effectively using a uni-modal communication channel.

#### 4.8.3. Limitation

This study used only a relatively small sample population. It also only involved adult participants. If conducted with participants with different age ranges, especially children, there is a possibility that the result could be different.

## Chapter 5. Main Experiment Part 1: Robot Teaching

The research presented in this thesis consists of three experiments, a first “investigatory” experiment and subsequently a “main” experiment comprising of two sub-experiments. The first experiment aimed to select a particular modality to be used in the subsequent experiments and is presented in the previous chapter. That experiment compared user preferences of input modality in order to teach a robot to mime. The findings suggested that the users appeared to have no preference in terms of human effort for completing the task. However, there was a significant difference in human enjoyment preferences of input modality and a marginal difference in the robot’s perceived ability to imitate.

The two further sub-experiments form part of the main experiment. The study of the first sub-experiment is presented in this chapter. The second sub-experiment is discussed in the next chapter. Both experiments were conducted in order to establish measurement metrics to evaluate human teaching engagement from a robot’s point of view.

### 5.1. Background

Rather than directly programming robots to carry out tasks, this study investigated how robots could learn these tasks by imitating humans. By using an imitation learning method, a human user can teach the robot by demonstrating how to accomplish a certain task. The robot will then learn the new behaviour by imitating the movements that the teacher demonstrates. In relation to this imitation learning, especially to address the “who to imitate” question, this study investigated what features are needed for a robot to evaluate support for this question when learning from a human teacher.

As discussed in Chapter 2, while imitation could accelerate how a robot learns complex behaviours, there are many aspects to consider. Among these factors is the question of what makes the demonstrator a good “model” to learn from and what criteria should be established to measure whether a model (the demonstrator) is (or could be) a good teacher. When there are several demonstrators available, the imitator (a robot in this case) needs to assess the examples against a number of criteria in order to choose which demonstrator is best to imitate. For example, a simple criterion might be one that maximises a benefit to the imitator, such as maximising the knowledge transfer, or getting the possible most correct movements. This chapter describes the first part of the main study that investigated the

assessment criteria to assess what makes a good model, by capturing the activity of teachers' when teaching arm gestures to a robot.

## 5.2. The Task

Two sub-experiments were conducted as parts of the main study. A brief overview of both experiments is presented below, before focusing on the discussion for the first sub-experiment. The second sub-experiment is discussed in the next chapter.

This main study aims to measure the criteria that can be used by the robot to detect a good teacher. The main study was separated into two experiments. In the first experiment, each participant acted as a teacher to teach the robot a number of arm gestures. In the second experiment, each participant acted as an evaluator of engagement of the participants in the first experiment. To allow evaluation, the first experiment was video recorded from the robot's viewpoint and the videos were then shown to participants who graded the original participant's engagement.

The whole main study involved two groups. The first group taught a humanoid robot to perform arm gestures from a nursery rhyme. The second group then evaluated the first group's teaching activities. As such, the first group was called "gesture teachers" and the second group was called "gesture evaluators" to match their activities. The works presented here describes the first group only, which were the "gesture teachers".

Similar to the experiment discussed in the previous chapter, the main study uses a nursery rhyme in the experiment. This time, the nursery rhyme was "Wind the Bobbin up". The lyrics are as follows:

*Wind the bobbin up*

*Wind the bobbin up*

*Pull, pull, clap, clap, clap.*

*Wind it back again*

*Wind it back again*

*Pull, pull, clap, clap, clap.*

*Point to the ceiling*

*Point to the floor*

*Point to the window*

*Point to the door.*

*Clap your hands together*

*One, two, three*

*Put your hands upon your knees.*

Six gestures were selected to be taught to the robot. They were:

1. Wind the bobbin up
2. Pull, pull
3. Clap, clap, clap
4. Point to the ceiling
5. Point to the floor
6. Put your hands upon your knees

These gestures were selected as they did not depend on the physical location of the robot nor the participants. For comparison, pointing to a door or to a window would require a robotic algorithm to make the robot recognize those objects. Or, at least the robot would need to be able to detect when the human participant pointed to a certain object.

The gesture teacher participants taught the robot by demonstrating the gestures that are relevant to each line of the nursery rhyme. For each gesture, the robot would imitate the demonstration while the participant was giving the demonstration. The participant signalled the start and the end of the demonstration using a special start/stop gesture. There were no special descriptions of the actual arm movements given regarding the gesture. They were open to the interpretation of the participants.

### **5.3. Robotics Software**

For this study, the robotics software that had been developed and used for the earlier experiment (described in the previous chapter) was developed further. This time the development was focused on the software on the external PC only. The software on the robotics side was unmodified and left as it was. The visual tracking module was re-used and tailored to fit the scenario in this experiment. The following sections describe the development of the system.

### 5.3.1. Controller GUI

This experiment uses two GUIs. The first one was the participants' GUI, and the second one was the controller GUI. The GUI for participants is discussed later in Section 5.3.2. The controller GUI is discussed below.

The controller GUI was used to control the robot which was beneficial during the testing and development of the system. During the experiment, the controller GUI was used to display real-time session information. The robot in the experiment was developed to interact fully autonomously during one-to-one interaction sessions with the human participants. To keep this impression, the computer monitor that displayed this GUI could be turned off. The appearance of the controller GUI is shown in Figure 5.1.

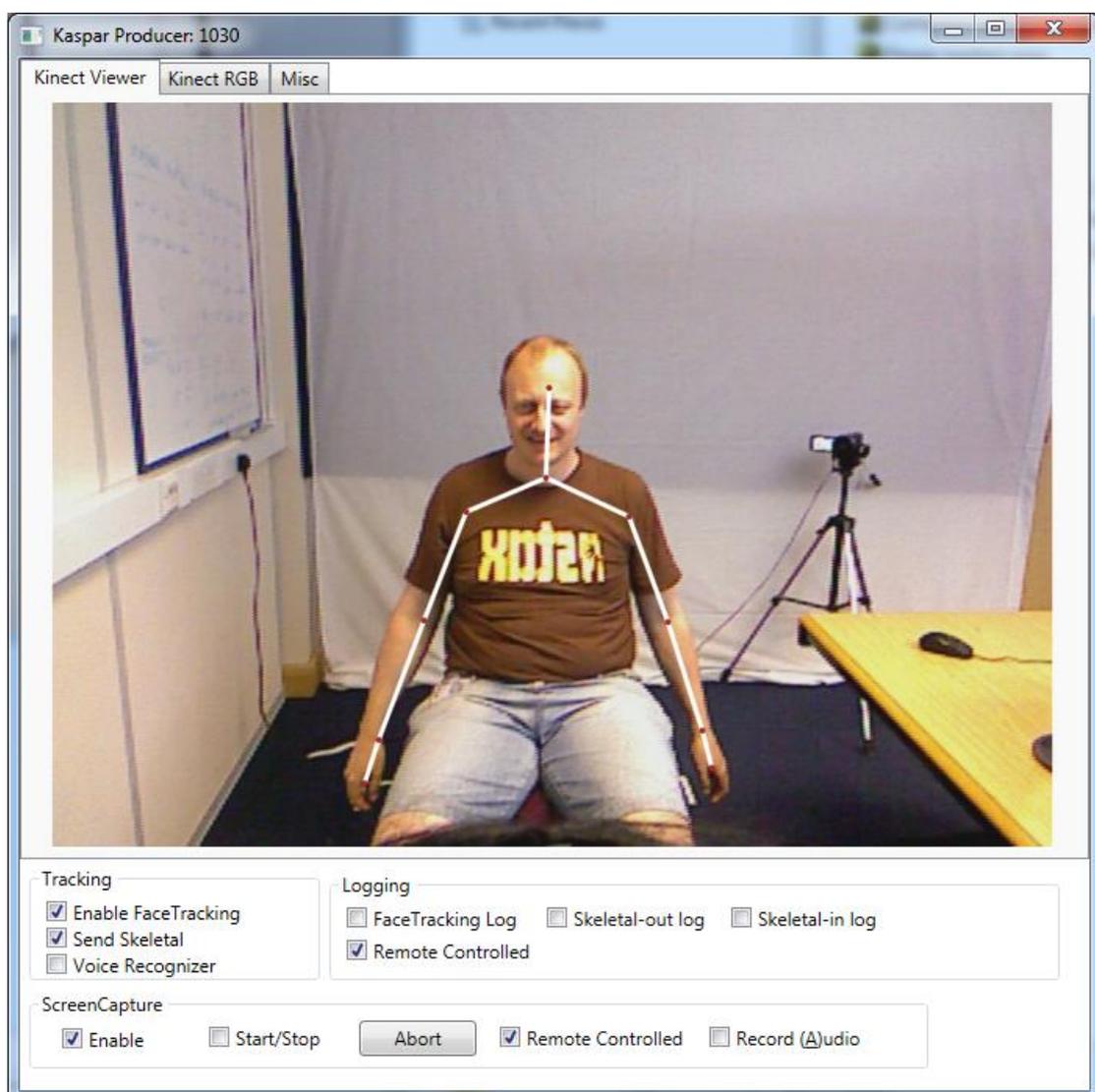


Figure 5.1 Controller GUI

### 5.3.2. Participants' GUI

The participants' GUI was used to display instructions of what gestures were to be taught to the robot. It had several tabbed panels. The participants accessed the panels one by one from the leftmost side and moved progressively to the right. Some of the panels were used to introduce the task and how to interact with the robot. One of the panels contained an interface to play an animation video of the nursery rhyme. The video was aimed at participants that were not familiar with the nursery rhyme. The appearance of the controller GUI is shown in Figure 5.2.

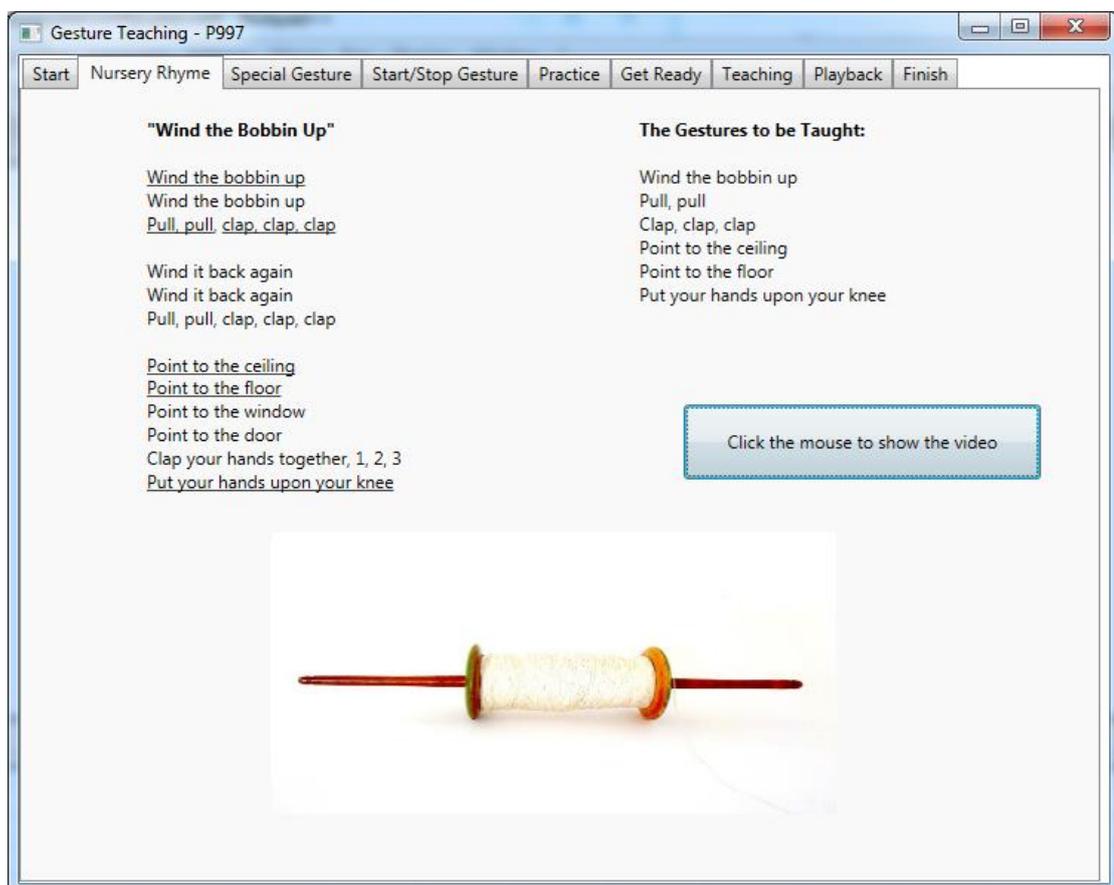


Figure 5.2 Participants' GUI

### 5.3.3. Start/Stop Gesture Detection

The participants taught the arm gestures by demonstrating the gestures to the robot. Before the demonstration, the participants needed to perform a special gesture to make the robot start imitating their gestures. After completing the demonstration, the participants again needed to perform a special gesture to make the robot stop imitating.

In this experiment, the same special gesture was used for both the start and stop of the gestural demonstration. The pose of this special gesture was that the participants put both arms straight next to the body. The posture is illustrated in Figure 5.3.

The software detected the special gesture using a finite machine state mechanism. It generates a software event when the participant made an arm movement in the special gesture pose from any of the other poses. When the participants made the special gesture prior to starting the demonstration, the system interpreted it as a start gesture. The following special gesture which came after that would be interpreted as a stop gesture.



**Figure 5.3 Start/stop gesture**

#### 5.3.4. Imitation Behaviour

The robot imitated the human gestures by visually tracking the arm movements. The software mapped the position of human's joints to move the arms of the robot with mirror mapping where the left arm of the participants was imitated by the right arm of the robot. The robot carried out the imitation instantly. Every movement the participants made was imitated by the robot.

The software had physical protection limits in place. If the movement was considered outside the boundary of a safe range of movement, the particular part of the arms that

reached the boundary would stop at the boundary. When the movement was within the allowable range again, the movement would resume.

### 5.3.5. Limitation

The robot had some limitations in the animation behaviours. Two came from the limitations of the hardware and one was intentionally placed in the software. The following describes the reasons for this:

The first limitation was that the visual tracking sensor (Kinect) could not detect the position of the arms when they were crossing in making a circular movement of winding a bobbin (“wind the bobbin up”). When the arms overlapped horizontally the sensor could not distinguish which one was in front of the other. This research did not use a special algorithm that could be used to overcome this limitation. Instead, the program used the data stream coming from the sensor as it was (without pre-processing) to be mapped to the robot’s arms.

The second limitation was that the robot’s physical embodiment could not make a hand clapping gesture when the hands were close to the chest. The hands would not come close enough to make a perceived “hand clapping”. Further to this, the fingers and the palms were not moveable, so they would never make a proper hand clapping gesture regardless of the distance to the chest.

The third limitation was because the robot was set to have slow movements. The speed was set to 200 (the value for servos). Originally, without slowing down, the robot movements were perceived as jerky. This was because there was no acceleration control applied to the servos. So for every movement, the robot would move instantly and when the movement stopped, it stopped suddenly. This is like a car making sudden moves and stops, hence the jerky movements.

### 5.3.6. Tester Participants

The robotics software, especially the participants’ GUI part, was developed to match the interaction scenario. The software was tailored to specific interaction flows required in the experiment. The software’s relation to the interaction was reciprocal in some ways in that the interaction scenario, mainly the introduction session, was partly adjusted to the developed software.

To test the software and also the experiment scenario, five tester participants were invited to evaluate the system. The testers were colleagues within the robotics labs. They participated in the test experiment in the same way as real participants. The result and the feedback from the test participants were used to improve the software and to fine tune the experimental scenario. These test experiments were conducted after the study had received ethics approval which is discussed in the following section.

## 5.4. Experiment Setup

The experiment was conducted in a robotics lab at the University of Hertfordshire. The following section discusses the setup of the experiment.

### 5.4.1. Ethics Approval

As the experiment involved human participants, an ethics application was submitted. The Ethics Committee of the University of Hertfordshire approved the submission by protocol number COM/PGR/UH/02024.

Later on, an amendment to the ethics application was submitted. It contained some minor modifications. One of them was to add a statement in the consent form to allow the investigator use the recorded video for scientific publication. This amendment was approved by the Ethics Committee protocol number aCOM/PGR/UH/02024(1).

### 5.4.2. Inviting Participants

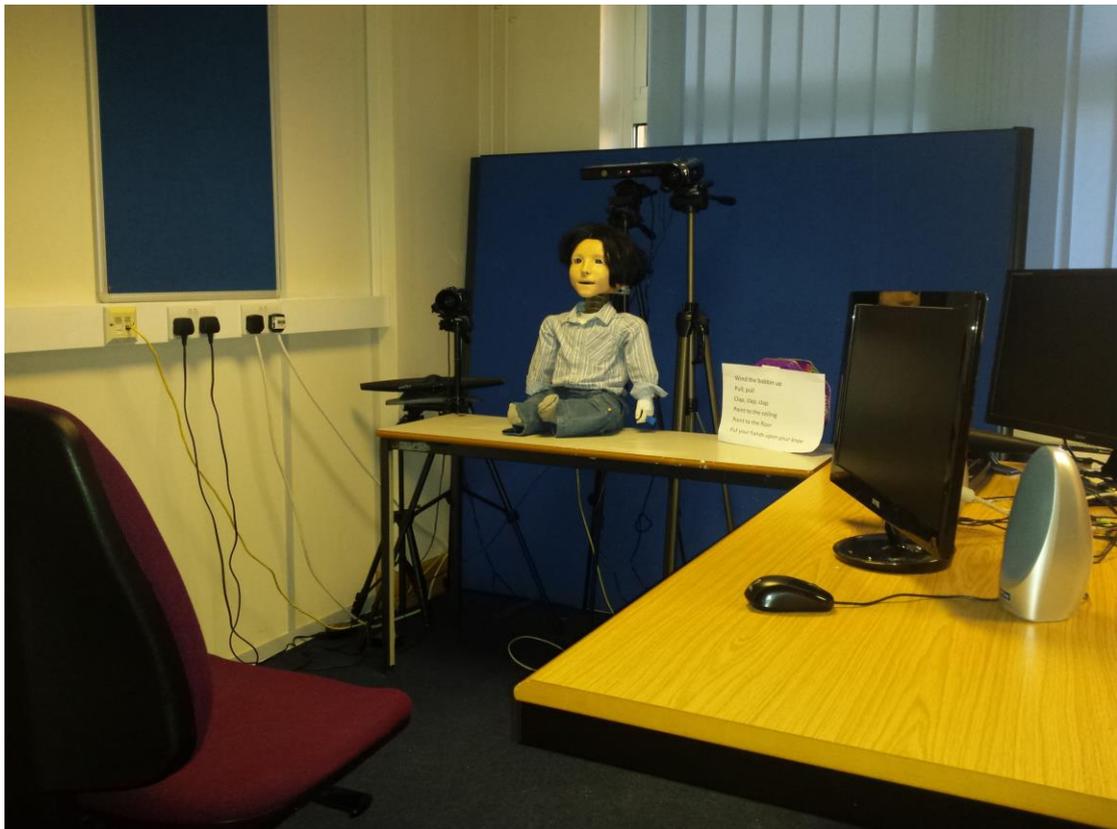
The study targeted the adult population to participate in the experiment. The participants were invited from the university and consisted of students and staff. The participation in the study was entirely voluntary.

Invitation posters were created and placed on several bulletin boards located within the university. Smaller size posters were provided as flyers. The locations of posters and flyers were stated within the ethics application. Permission was obtained verbally prior to the erection of the posters.

Direct invitation, verbally and through email, was also carried out to invite the participants. The smaller size posters were given when inviting the participants verbally. Compared to the invitation posters or flyers, only the direct invitations were successful in getting people to take part in this study.

### 5.4.3. Equipment and layout

The layout of the experiment is shown in Figure 5.4. It uses the same KASPAR robot used in the previous experiment discussed in Chapter 4. This time a more powerful external PC was used because of the computational requirements of the system. This was a desktop PC powered by an Intel XEON E5-1650 processor with 32 GB of RAM.



**Figure 5.4 Experiment layout**

A Kinect sensor was used to visually track the human participants. The sensor was located above and behind the head of the robot. The location of the sensor was chosen to record videos of the participants as if seen from the robot's eyes. Later on, the videos were kept as backups because the quality of videos from another external video camera was better than the ones which were captured by the camera of Kinect sensor. However, the position of the sensor above the head gave a perspective of the participant from above. It was decided that for the gesture evaluator experiment (the second sub-experiment) the recordings from a camera located on the front left of the participant had better video quality and recordings of the participants from the same height level (centred vertically about the chest).

Two other cameras were used to record the session. One camera recorded the session from above and behind the head of the robot. This camera was located next to the Kinect sensor. The other camera was located towards the rear right of the participant. This camera captured the activity of the participants from the back and also captured the robot's movements.

Three microphones were used by the software to record the sound during the experiment. The first one was the built-in Kinect microphone. The second one was a regular omnidirectional microphone. The third one was a unidirectional microphone. The recording was activated by the software when the robot was imitating the human participants. In addition to these, the audio was also recorded by the three video cameras. Compared to the software controlled recording, the video camera recorded the audio during the whole session from the beginning of the experiment until they were turned off manually at the end of the experiment.

## 5.5. Experiment Procedure

The participants took part in each experiment individually. Each session lasted approximately 30 minutes. The activities in the experiment are shown in Figure 5.5 on the next page. The procedure for the experiment is described below.

### 5.5.1. Pre-trial Part

Firstly, the participants were given an information sheet regarding the experiment. The participant then signed a consent form. After that, the participant filled out a pre-trial questionnaire. The questionnaire contained a question that asked whether the participant was familiar with the "Wind the Bobbin up" nursery rhyme.

### 5.5.2. Introduction Session

The information sheet given in the pre-trial part contained information about the activities in the experiment. The introduction session gave more detailed information.

After completing the pre-trial questionnaire the participants were invited to sit in the interaction area. The participants were then introduced to the robot. After that, the participants were told that the participants' GUI (see Section 5.3.2) would give information about the experiment and the investigator would explain information alongside. They only used a computer mouse to access the GUI.

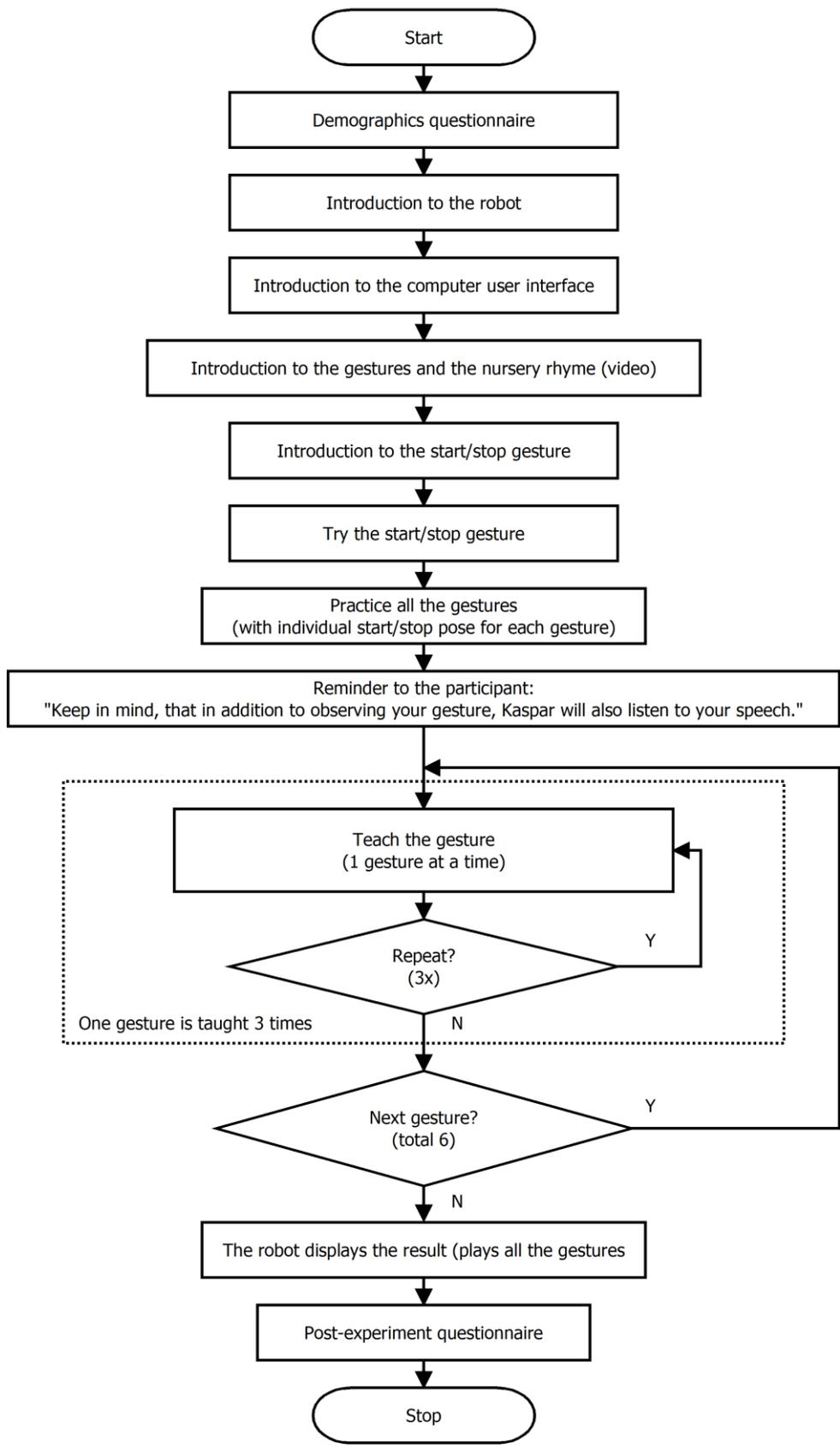


Figure 5.5 Activities in the "Gesture Teacher" experiment

The participants' GUI contained tabbed panels, one of the panels introduced the lyrics of the nursery rhyme. Within the same panel, the GUI provided a button that would open another window to play an animated video of the nursery rhyme. This was provided mostly to introduce the nursery rhyme to the participant if they were not familiar with it. The animation lasted approximately 2.5 minutes. Regardless of their familiarity, all the participants were asked to watch the animation. They were informed that they did not have to watch the whole animation.

The participants' GUI has several panels that in sequences introduced the participants to the experiment. One of them was used to introduce the start/stop gesture (see Section 5.3.3). The participants were asked to perform three successful start/stop gestures. The participants faced toward the robot and the robot acknowledged the gestures by saying "ready" for the start gesture, and "stop" for the stop gesture. Following this introduction, the participants were then asked to practice teaching the robot all the six gestures discussed in Section 4.4. In this sub-session, the robot would imitate the arm gestures made by the participants after detecting the start gesture. The participants were told to teach one gesture at a time and make a stop gesture before teaching another gesture.

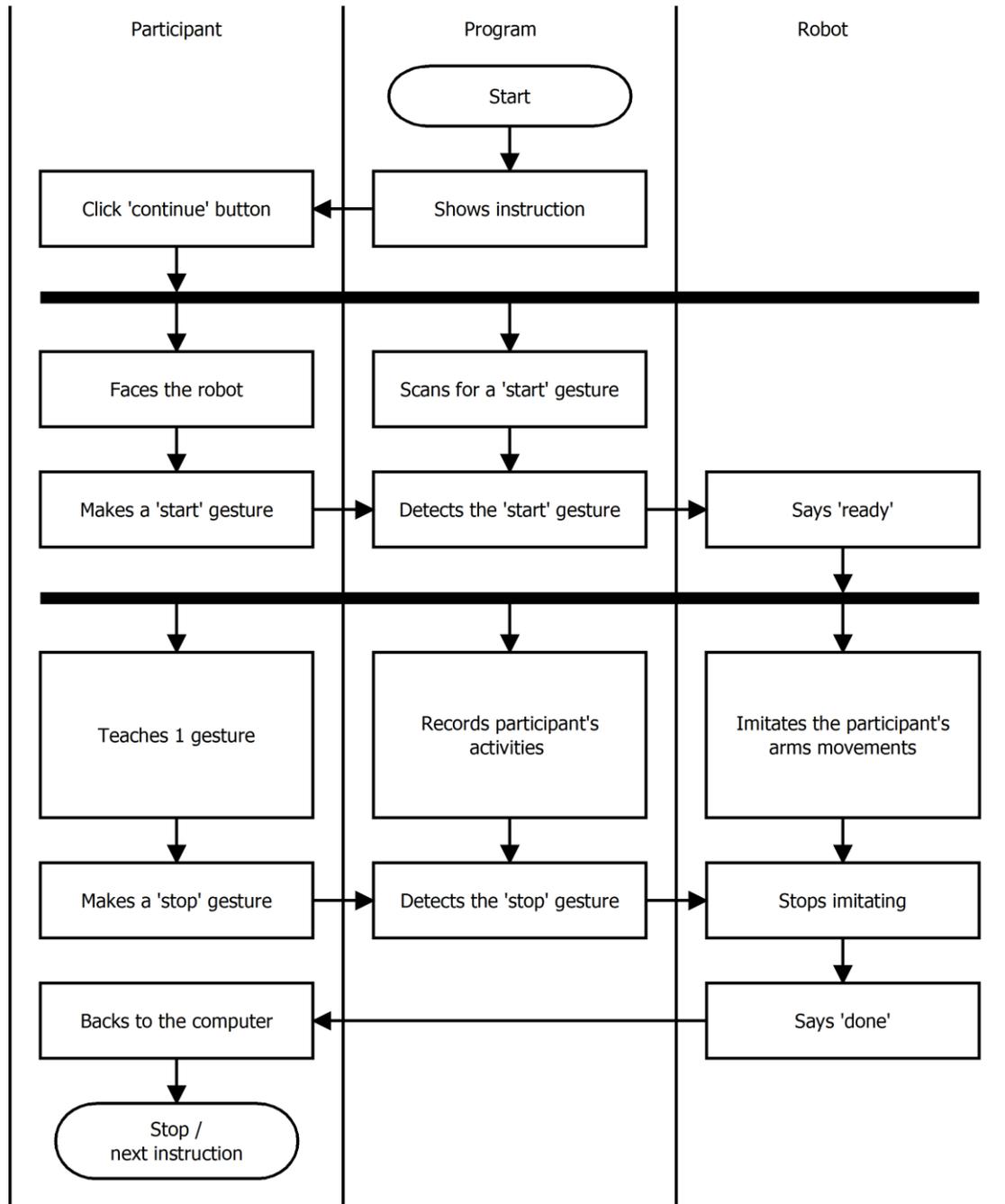
After completing the practice, the GUI then displayed a question on whether the participant was ready to teach the robot. The investigator told the participants that the introduction had been completed. After that, the participants were left alone to follow the instructions in the participants' GUI to teach the robot. The investigator stayed in the same room reading a book or other papers.

Before leaving the participants to progress, the investigator reiterated some information from the GUI to the participants and said "Keep in mind, that in addition to observing your gesture, KASPAR will also listen to your speech."

### 5.5.3. Teaching Session

In the teaching session, the participants' GUI was displayed on a computer monitor in close proximity to the participants and prompted what gesture the participants should teach. The participants then clicked a "continue" button to teach the robot. The participant faced the robot to teach it. To actually begin teaching the robot, the participants made a start gesture. The robot acknowledged it by saying "ready" when the gesture was detected and started to imitate the arm movements the participants made. After teaching a gesture, the participants made a stop gesture and the robot acknowledged it by saying "done". In this case "done",

instead of “stop”, was used to indicate that the robot has done learning the particular movement. The participants then came back to the computer monitor and clicked a “continue” button to continue with the next instruction. The flow of the process, for one gesture only, is shown in Figure 5.6.



**Figure 5.6 Interaction activities in gesture teaching (one gesture)**

The GUI repeatedly prompted three times for each gesture. After one gesture was asked for three times, the GUI then continued to prompt for another gesture. So, in total, there were six gestures and each one was repeated three times.

After completing the training with the robot, the participants' GUI then moved to another panel to show what the robot had learned. After clicking a "continue" button, the robot would start showing the gestures learnt. The robot said the name of the gesture first before displaying the arm movements. The robot repeated this pattern to display all of the learnt gestures.

At the end of this main session, the GUI then moved to the last panel, which informed the participant that the session was finished.

#### 5.5.4. Post-trial

When finished teaching the robot, each participant was asked to complete a post-study questionnaire. After completing the questionnaire, the participant was asked verbally if there were any comments or feedback regarding the experiment. The feedback was used to record if there was more information that could support the analysis.

### 5.6. Data Collection

#### 5.6.1. Dependent Measurement

The post-questionnaire was provided over two pages. The first page asked four questions. The first three questions were as follows:

1. In terms of effort, how did you find teaching KASPAR the gestures? (1 = very hard, 5 = very easy)
2. In terms of enjoyment, how much did you enjoy teaching KASPAR the gestures? (1 = least enjoyable, 5 = most enjoyable)
3. When KASPAR showed you what it had learned, how well did you feel KASPAR followed your demonstration? (1 = not very well, 5 = very well)

1	2	3	4	5
---	---	---	---	---

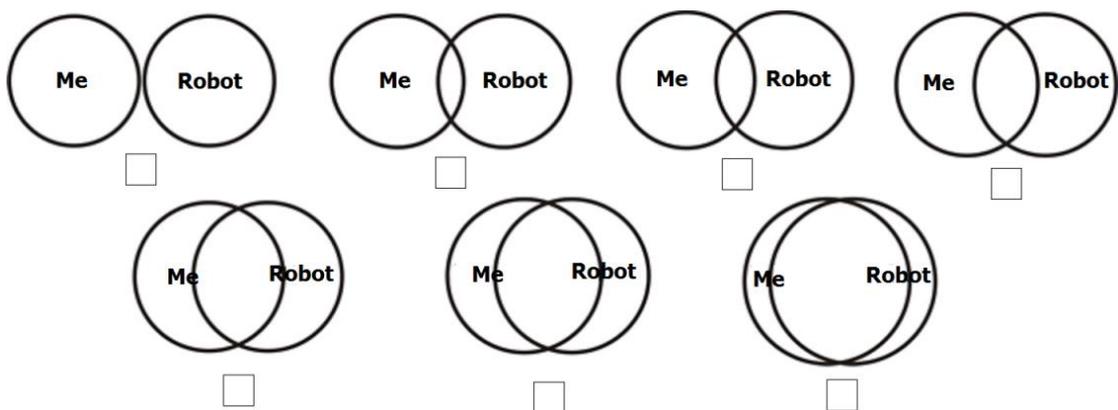
**Figure 5.7 Answer boxes next to each gesture**

Within each question above the participants answered the questions, marking the answer box (Figure 5.7) located next to the name of the gestures being taught in the session, which were:

- A. "Wind the bobbin up"

- B. "Pull, pull"
- C. "Clap, clap, clap"
- D. "Point to the ceiling"
- E. "Point to the floor"
- F. "Put your hands upon your knee"

The fourth question asked the inclusion of other in the self (IOS, (Aron et al., 1992)) through this instruction: "Please tick the picture that best describes your relationship to the robot during the experiment session." The available answers are shown in Figure 5.8. The IOS questionnaire is useful to find out how much the subjects related themselves to the robot. In this case, this is related to how willing was the participant to share information (i.e. the teaching) to the robot. This question was provided for evaluation of whether it has any relation to the engagement of the teacher.



**Figure 5.8 Answer boxes of the IOS question**

The second page contained the fifth and final question, which was provided for evaluation of whether the robot had any effect on the engagement of the teacher. This was actually a series of questions to measure the users' perception of the robot. It used the questions from the Godspeed questionnaire series (Bartneck et al., 2009). The Godspeed questionnaire is useful for assessing the participant's impression of the robot. It asked the participant to rate their perception of the robot in several questions within five categories (anthropomorphism, animacy, likeability, perceived intelligence). It also asked the emotional state of the participants (perceived safety category). For example, the participant could state their feeling in scale from 1 to 5 between anxious and relaxed (the complete questionnaire can be

seen in Appendix C). Both the IOS and Godspeed questionnaires aim to support the analysis with additional data.

### 5.6.2. Method

During the main session, the robotics software recorded the activities of the participants. The recorded data included voice, skeletal joints, and face tracking. The software recorded this data in a database and gave timestamps to every software event recorded. For streamed data such as the skeletal joints and face tracking, each entry was timestamped individually.

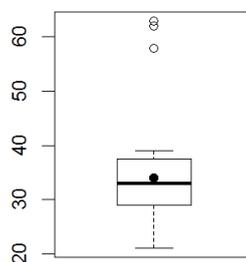
The entire duration of the sessions was also recorded using three video cameras. They recorded the whole session from the beginning until they were turned off at the end of the experiment.

Paper-based questionnaires also recorded the participants' responses to the questionnaires. These answers were later inputted to a spreadsheet in a computer to make an electronic copy of the data. The data then was analysed further and discussed in the data analysis section.

## 5.7. Results

### 5.7.1. Participants

The experiment was conducted with 9 male and 7 female participants (total 16). The youngest participant was aged 21 and the oldest one was 63. The boxplot of the age of the participants is shown in Figure 5.9.



**Figure 5.9 Age of the participants**

### 5.7.2. Data Analysis

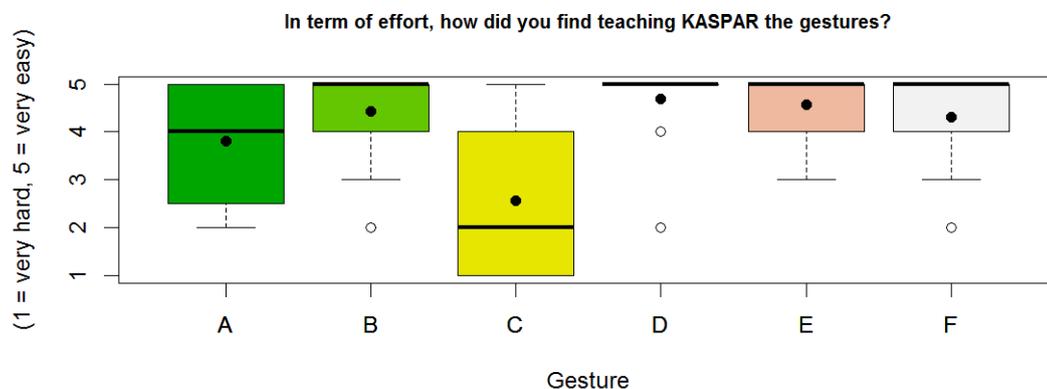
The result of Friedman test of the questionnaire data is shown in Table 5.1. All the tests showed significant differences, except Q2.

**Table 5.1 Friedman test result of the questionnaire data**

Alias	Question	Friedman chi-squared	dF	p-value
Q1	In terms of effort, how did you find teaching KASPAR the gestures?	31.229	5	8.44E-06
Q2	In terms of enjoyment, how much did you enjoy teaching KASPAR the gestures?	7.8327	5	0.1657
Q3	When KASPAR showed you what it had learned, how well did you feel KASPAR followed your demonstration?	35.322	5	1.30E-06
GS	Godspeed Questionnaire set	17.879	4	0.001303

The result of the first question of the questionnaire that asked about participants' effort in teaching the gestures is shown in Figure 5.10. The result shows that Gesture C ("Clap, clap, clap") was perceived by many participants as hard.

The p-value of Wilcoxon signed-rank test between each gesture (shown in Table 5.2) suggests that Gesture C is significantly different to all other gestures in terms of perceived effort. Gesture D ("Point to the ceiling") was deemed to be the easiest one in terms of effort but only significantly different to Gesture A ("Wind the bobbin up") and Gesture C. In this case, Gesture A is considered statistically less easy than Gesture D.

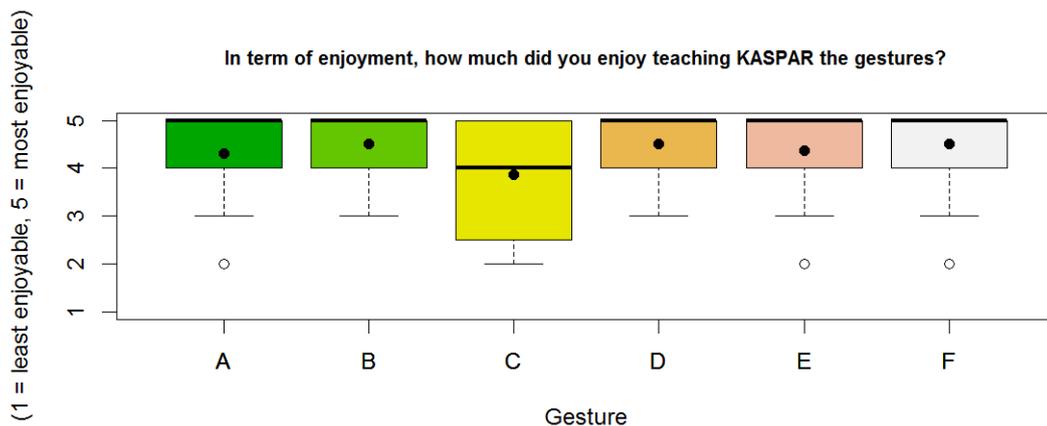


**Figure 5.10 Questionnaire result of "effort"**

**Table 5.2 Wilcoxon signed-rank test p-value of "effort" questionnaire**

	A	B	C	D	E	F
A		0.132438	0.027951	0.026324	0.08267	0.24868
B	0.132438		0.001781	0.412668	0.890128	0.737739
C	0.027951	0.001781		0.000478	0.000954	0.003807
D	0.026324	0.412668	0.000478		0.455787	0.242672
E	0.08267	0.890128	0.000954	0.455787		0.62281
F	0.24868	0.737739	0.003807	0.242672	0.62281	

In terms of enjoyment, from the second question, all the participants seem mostly to enjoy teaching the robot. The result is shown in Figure 5.11. The p-value of the Friedman test is 0.1657, which indicates no statistical difference in the level of enjoyment among all the gestures.



**Figure 5.11 Questionnaire result of "enjoyment"**

From the third question, Gesture C also received the least score in terms of how well the robot followed the demonstration. The result is shown in Figure 5.12. The paired test result shows that this gesture was significantly different compared to almost all other gestures except Gesture A.

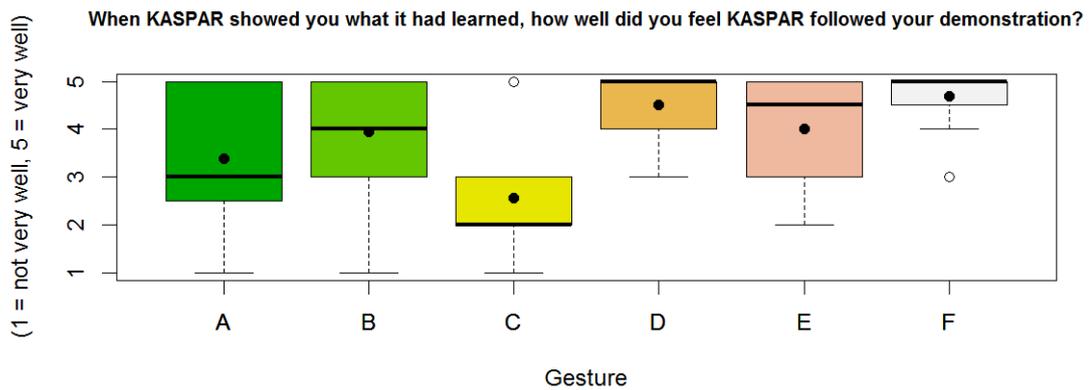


Figure 5.12 Questionnaire result of "robot follows demonstration"

Table 5.3 Wilcoxon signed-rank test p-value of "robot follows demonstration" questionnaire

	A	B	C	D	E	F
A		0.247495	0.091418	0.016845	0.214688	0.003995
B	0.247495		0.008845	0.192902	0.87244	0.049492
C	0.091418	0.008845		0.000316	0.006832	0.000119
D	0.016845	0.192902	0.000316		0.285458	0.445509
E	0.214688	0.87244	0.006832	0.285458		0.089295
F	0.003995	0.049492	0.000119	0.445509	0.089295	

The questionnaire result of the inclusion of other in the self is shown in Figure 5.13. The result shows that the participants felt a moderately close relationship between them and the robot during the experiment session.

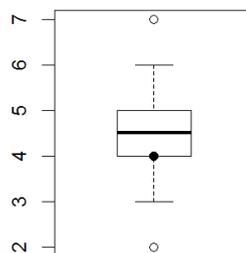
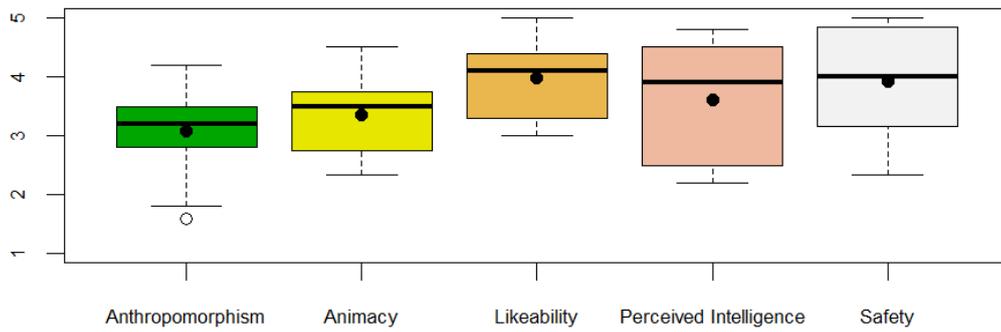


Figure 5.13 Result of the IOS Questionnaire

The result of the last part of the questionnaire, which was the Godspeed questionnaire, is shown in Figure 5.14. The anthropomorphism received the least score among other categories. The p-value of the paired test indicates that anthropomorphism score is significantly different to the likeability and safety categories.



**Figure 5.14 Result of the Godspeed Questionnaire**

**Table 5.4 Wilcoxon signed-rank test p-value of "Godspeed" questionnaire**

	Anthropomorphism	Animacy	Likeability	Perceived Intelligence	Safety
Anthropomorphism		0.21255	0.00396	0.11655	0.01327
Animacy	0.21255		0.02717	0.33572	0.06093
Likeability	0.00396	0.02717		0.4262	0.87963
Perceived Intelligence	0.11655	0.33572	0.4262		0.36435
Safety	0.01327	0.06093	0.87963	0.36435	

### 5.7.3. Data for the Teacher Evaluator Experiment

As mentioned earlier in this chapter, the experiment presented was the first part the main experiment which consisted of two sub-experiments. The second sub-experiment is presented in the next chapter. The data recorded in this experiment was compared against the data from the second sub-experiment, which was the "teacher evaluator" experiment.

Almost all the data in this experiment was successfully recorded. The video recording from external cameras successfully recorded the session from the beginning to the end of each participant's session. The Kinect data, skeletal and face data, were stored in SQLite database<sup>16</sup>, using a timestamp, and can be retrieved using the SQL language.

<sup>16</sup> <https://sqlite.org/>

One problem arose from the audio recording. The experiment set up used a unidirectional microphone which was initially expected to record only the voice from the participant (and not recording the sound from the robot). The problem was that the microphone had an automatic "muting" feature that required the participant to speak loud enough for the microphone internal circuitry to open the mute. From the recording result, not every participant spoke loud enough according to the microphone, therefore while the participants did speak, the recording did not give any audible sound (other than the baseline noise).

This research conducted the experiment by using a snapshot of research laboratory robot technology. The robot and the developed software by themselves were considered working as expected. The problem came from a support device that supposed to record clean sounds from the participants. There were other non-directional microphones used in the experiment, such as on the cameras. But they also recorded the voice and mechanical noise from the robot. As mentioned earlier, there were some other interaction data successfully recorded. Nevertheless, this caused the experiment to miss one interaction data that might be useful to evaluate the level of engagement of the human teachers.

## 5.8. Conclusion

For this experiment, the study developed software that allowed participants to teach a robot some arm gestures. In general, the software worked as expected to control the robot and to record the data of the participants.

From the questionnaire result, the robot worked as expected. Most of the participants gave a positive score to the robot performance on almost all gestures. Two gestures were indicated to have problems which were the "Clap, clap, clap" and "Wind the bobbin up". These were already anticipated in the beginning because of the limitation of the robot. The limitation on the "Clap, clap, clap" gesture was due to the robot's physical embodiment that could not make a hand clapping gesture when the hands were close to the chest. The limitation on the "Wind the bobbin up" gesture was caused by the visual sensor, Kinect, that could not detect the position of the arms when they were crossing in making a circular movement of winding a bobbin.

In general, except the audio recording, the experiment presented in this chapter had successfully recorded the "teacher data" from 16 participants. This data was to be used in

the experiment discussed in the next chapter to compare how other people evaluate the level of engagement for participants in this experiment while teaching the robot.

## Chapter 6. Main Experiment Part 2: Teacher Evaluator

This chapter presents the second phase of the experiment and evaluates the behaviour of human gesture-teachers as seen from another human perspective. In this case, one acts as a teacher and the other acts as the evaluator. The purpose of the experiment was to measure human perceptions of an engaged teacher. This was to get insight from human evaluators on how a robot can evaluate the human teachers. In this case, based on the result from the human evaluators, the study tried to establish the measurement metrics that can be useful from the robot perspective to evaluate the human teachers.

### 6.1. Background

Chapter 2 discussed what constitutes a good teacher based on the level of engagement, using references from the literature. Chapter 3 expands on this by discussing what physical attributes are exhibited by engaging teachers. Based on this, Chapter 5 discussed the first part of the main experiment (sub-experiment 1), which captured and analysed the behaviour of the teachers when they taught the arm gestures. The interaction modalities of that experiment were chosen by reference to the experiment discussed in Chapter 4. The data captured from the experiment described in Chapter 5 included the intensity of the voice of the human teacher and the acceleration of arm movements of the human teacher. The intensity of the voice and the acceleration of arm movements were analysed in measuring the effort of the teacher in relation to evaluating the level of engagement of the teacher. In this chapter, we investigate further by evaluating if humans shared the same perceived level of engagement.

Not only to assess on how a human evaluates another human teacher, the experiment in this chapter was also to relate the assessment results to the perceived effort sensed by the robot. In this case, the study tried to compare the data recorded by the robot sensors in the previous experiment to the data in the following experiment. The findings of this analysis aimed to suggest indicators that could be used to develop a learning strategy, which would be useful for assisting robots to learn more effectively from human teachers, especially in relation to learning from imitation.

## 6.2. The Task

In the first part of the main experiment (sub-experiment 1) video recordings of the participants, as gesture teachers, were taken while they were teaching the robot. The participants taught arm gestures from the “Wind the Bobbin up” nursery rhyme to the robot. In the second part of the main experiment (sub-experiment 2), the teacher evaluator participants watched the video and their perception of the gesture teachers’ behaviour was captured.

The task of participants in sub-experiment 2 was to watch videos of people (teachers) teaching arm gestures to the robot. The video was shot from the front-left of the teacher.

The robot was not shown in the video. The audio contained sounds from both the teacher and the robot.

The videos from the sub-experiment 1 were divided into segments. Each segment contained a short session of one teacher-taught arm gesture. In total, there were six segments per teacher. The duration of the videos varied from 6.9 to 34.3 seconds depending on the gesture and the teacher.

The participants in this experiment evaluated the behaviour of the teacher using an interface that can be used to rate in real-time while watching the video. After rating the video in real-time, the participants completed a questionnaire to evaluate the behaviour of the teacher they watched. To allow the participant to fill in the questionnaire for individual videos, they were prompted at the end of each video before moving to evaluate another video.

## 6.3. Software

This time, the study did not use a robot in the experiment for interaction. The robot was only used to show the actual robot being taught in the video.

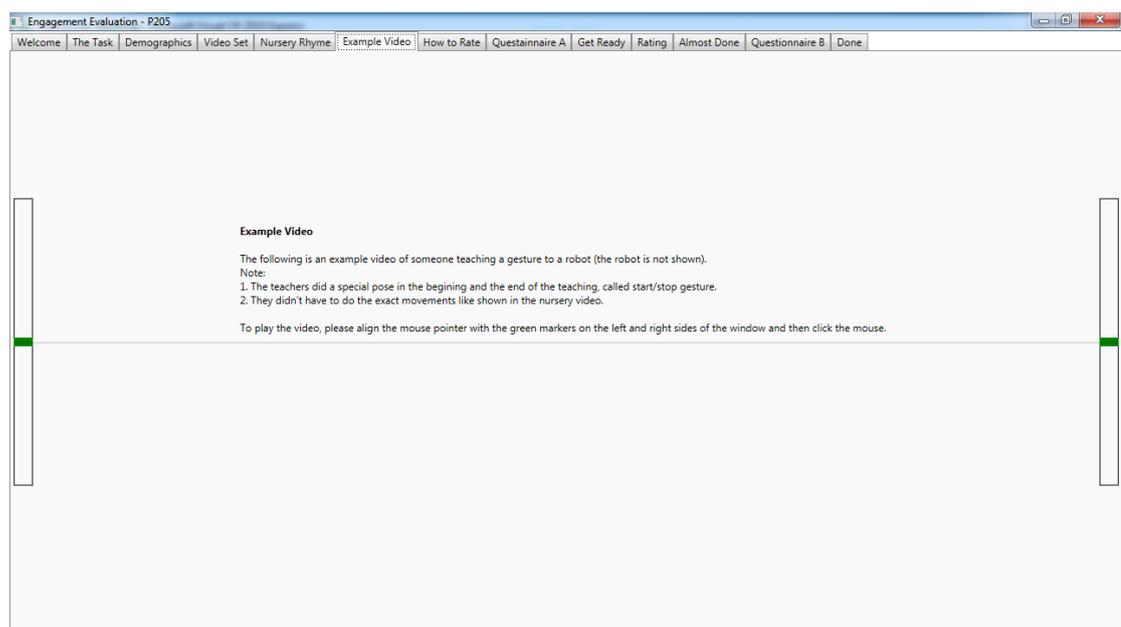
Another program was developed for this experiment. The participants used the program to play the video of the teachers (from sub-experiment 1) on a computer monitor. The experiment was designed to have the participants use the same program to rate the teacher in real-time. Instead of using a separate device to rate the video, the program provided a graphical based interface to allow the participants to rate the video in real-time, while simultaneously watching the video. The following section describes the program.

In addition to the program used by the participants to rate the video, one other program was also developed to evaluate the result of the experiment. This program was used by the author to video annotate in comparing the result of sub-experiment 1 to the result of sub-experiment 2. This program is discussed later in Section 6.8.2

### 6.3.1. Video Rating Interface

An experiment which used an interface to rate a video in real-time while watching it was conducted by Wood (2015). That study used a slider shown on the right side of the video. The evaluators rated the video by moving the slider up and down by moving the mouse up and down respectively.

There was a significant difference from the videos in this experiment to the study by Wood: while the videos in the current study lasted for a matter of seconds, the videos in Wood's study lasted for minutes. With short videos, it would be distracting if the participants had to constantly look at the edge (in this case, the right edge) to track the rating being given.



**Figure 6.1 Video rating interface: before playing**

For this study, the author developed a new, enhanced version of a real-time video rating interface which used a vertical line across the video screen that moved up and down, following the mouse movements. This way, the participants can visually track the mouse movement while watching the video at the same time. The horizontal line could move up and down within a boundary that covered half of the screen in the middle of the display. Figure 6.1 shows the appearance of the interface before playing the video. In this case, it

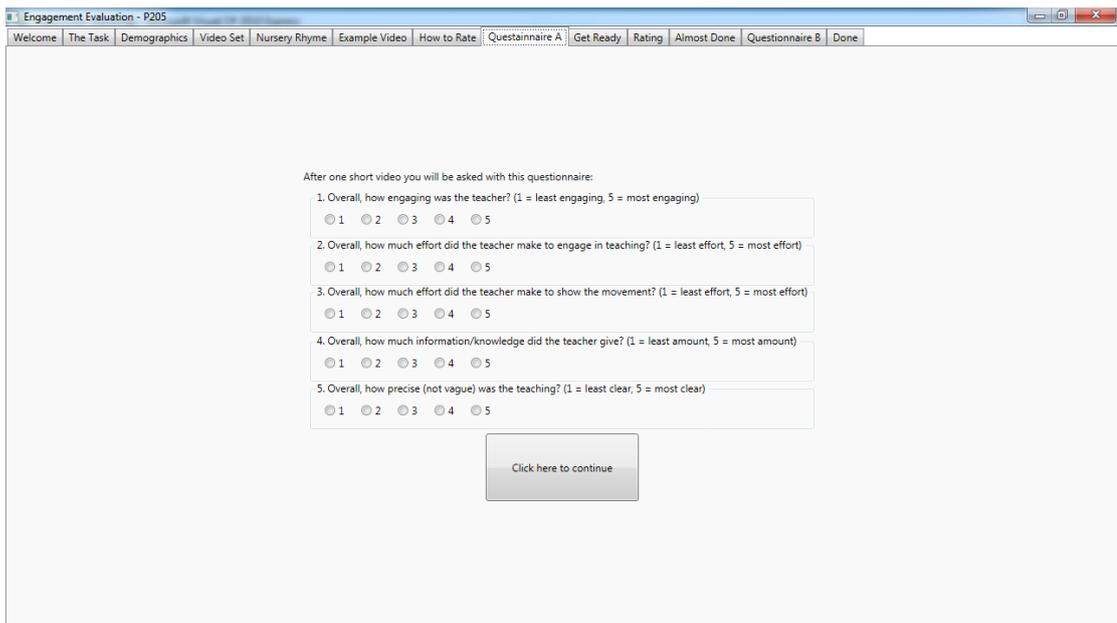
shows the message for showing an example video. To play the video, the participants needed to align the horizontal line to be in the middle of the screen. After that, the participants could play the video by clicking the button of the computer's mouse.

Figure 6.2 shows the appearance of the interface while playing the video. There was no pause, stop or rewind control interface in the video player. Once the video started, it would play for the complete duration.



**Figure 6.2 Video rating interface: playing**

The program (the rating GUI) would prompt the user to complete a questionnaire after playing each video. After playing the whole set of videos, the program would also prompt a final questionnaire, which was different to the questions given previously. The program prompted a questionnaire within the panel of the GUI. The appearance of the video questionnaire is shown in Figure 6.3. The zoomed-in appearance of video and final questionnaires, which are showing the questions more clearly, are shown later in Section 6.6.1.



**Figure 6.3 The video questionnaire within the GUI program**

### 6.3.2. Tester Participants

A similar pattern to the previous experiment was carried out to test the program. An initial concept of the GUI was designed and implemented. This implementation was then tested by the author to evaluate the program. After some iterations, the author then invited some tester participants to test the program and the interaction scenario between the investigator and the participants, particularly the introduction session.

The tester participants were colleagues from the robotics lab. They participated in the test experiment in the same manner as real participants. The results and feedback from the participants were used to improve the software and to fine tune the experimental setup. After two reiterations (with two participants), the software was considered ready to be used for the actual experiment.

## 6.4. Experiment Setup

The experiment was conducted in a robotics lab at the University of Hertfordshire in the same location as the previous experiment. The following section discusses the setup of the experiment.

### 6.4.1. Ethics Approval

This experiment was one of the two-part experiments of the main study. Both parts were interconnected and the ethics application was submitted as a single application. As

mentioned in the previous chapter, the Ethics Committee of the University of Hertfordshire approved the submission by protocol number COM/PGR/UH/02024. Later on, an ethics amendment application was submitted and was approved by the Ethics Committee protocol number aCOM/PGR/UH/02024(1).

#### 6.4.2. Equipment and Layout



**Figure 6.4 Experiment layout**

The layout of the experiment is illustrated in Figure 6.4. The robot was located near the participants. It was off and was only used for the purpose of showing the actual robot which was taught by the gesture teacher participants group.

The participants used a C# based GUI program displayed on a computer monitor. The computer monitor size was 21.5 with a native resolution of 1920x1080. For the experiment, the screen resolution was set to 1280x720.

The participants mainly used a computer mouse to control the program. At the beginning, the participants also used a keyboard to input some data, such as their demographics data. External speakers were used to deliver the audio.

## 6.5. Experiment Procedure

In this experiment, each evaluator participant watched two sets of videos from two different teacher participants. Some of the teacher participants were reinvited as evaluator participants. The time gap to the previous experiment for the reinvited participants was at least 9 weeks, giving them reasonable time not to remember too much about the previous experiment. More information about the evaluator participants is discussed in Section 6.7.2. The experiment lasted approximately 20 minutes. The flow of the activity in the experiment is shown in Figure 6.5. The procedure of the experiment is described below.

### 6.5.1. Pre-trial Part

Before beginning the actual experiment, the participants were given an information sheet regarding the experiment. After reading the information sheet, the participant signed a consent form. In contrast to the previous experiment, they completed a demographics data form electronically through the same program that was used to evaluate the behaviour of the teacher participants. They also indicated on the form in the program if they were involved as teacher-participants in the previous experiment.

The participants randomly selected two sets of videos (teachers) to be evaluated. The selection was provided as lottery papers that drawn randomly by the participants. The participants then typed the code on the paper into the computer. The program would play the video sets according to the code.

The first participant had the opportunity to select two sets from all of the possible video sets. The next participant could select another couple of video sets from the remaining video sets. As the number of participants that took part increased, the number of choices they had for selecting videos decreased until the final participant was left with the last set of videos. The pairs came from the earlier random selection which picked two sets of videos, thus making them a pair.

The results of the above random selection were checked by the investigator. If they picked their own videos, they were asked to draw another one or another set.

### 6.5.2. Introduction Session

The introduction, through the same GUI program, was started by introducing the “Wind the Bobbin up” nursery rhyme. The program had a button that, when clicked, would open a

video player to play an animated video of the nursery rhyme. The participants were allowed to skip the video if they were familiar with the nursery rhyme.

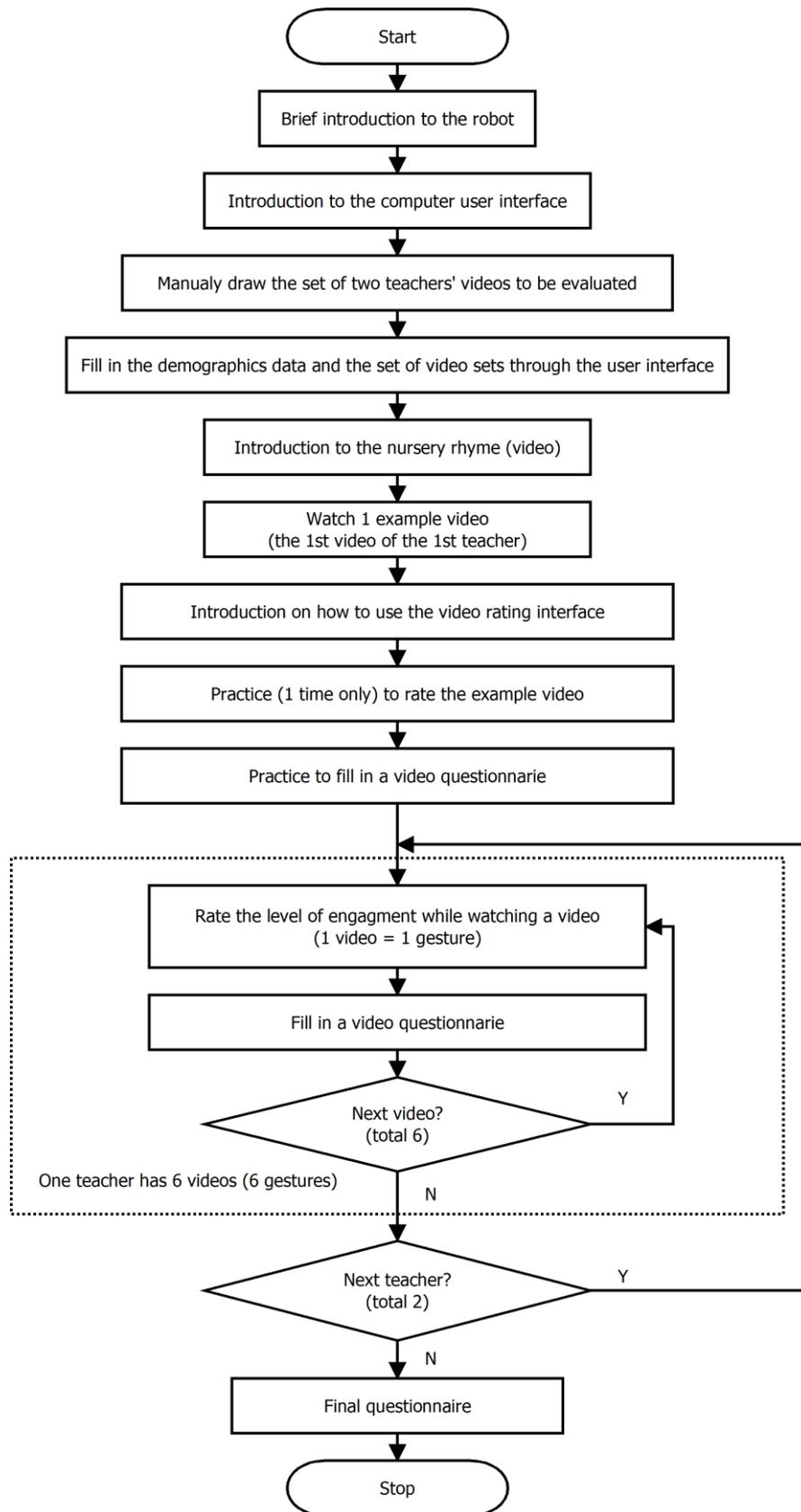


Figure 6.5 Activities in the "Teacher Evaluator" experiment

Next, the participants were shown an example video of a teacher participant teaching one of the gestures. This was an introduction to the interface of the video player and what kind of video that they would watch. As discussed in more detail earlier in Section 6.3.1., they needed to align the visible horizontal line in the screen to be in the middle of the screen and click the mouse to start playing the video.

After that, the participants were shown how to rate the video in real-time while watching the video. The participants were then asked to practice rating a video using the same video shown as the example video.

Following the practice, the participants were prompted with a mock-up of the video questionnaire. The participants were asked to read the questions and were asked if they had any question about the questionnaire. Afterwards, they were asked to practice filling out the questionnaire based on the example video.

Filling in the mock questionnaire was the last part of the introduction session. The participants were then asked if they had any questions before moving on to the main evaluation session.

### 6.5.3. Evaluation Session

Before starting the evaluation of the videos, the participants were told that they would only watch each video once. The investigator reiterated that they needed to rate the video simultaneously while watching the video.

The same example video was played by the GUI program as the first video to be evaluated by the participants. As discussed earlier, the participants had to align the mouse before clicking its button to start the video. This time, the real-time stream of mouse movements while the video was being played was recorded as the rating data. After the video had finished playing, the program prompted the video questionnaire for the participant to complete using the mouse. The flow of the activity in the evaluation session where the participant evaluated one video is shown in Figure 6.6.

The GUI program repeated the above sequences to show the whole set of videos of one teacher participant. After that, the GUI program moved to the second teacher video sets and played the videos with the same sequence of gestures as the first teacher. The selection of the pairs of teachers was randomly drawn by the participants as previously mentioned.

All participants evaluated the videos with the same sequence of gesture for every teacher. Each of them watched the same example video from their respective first teacher of their pair of video sets.

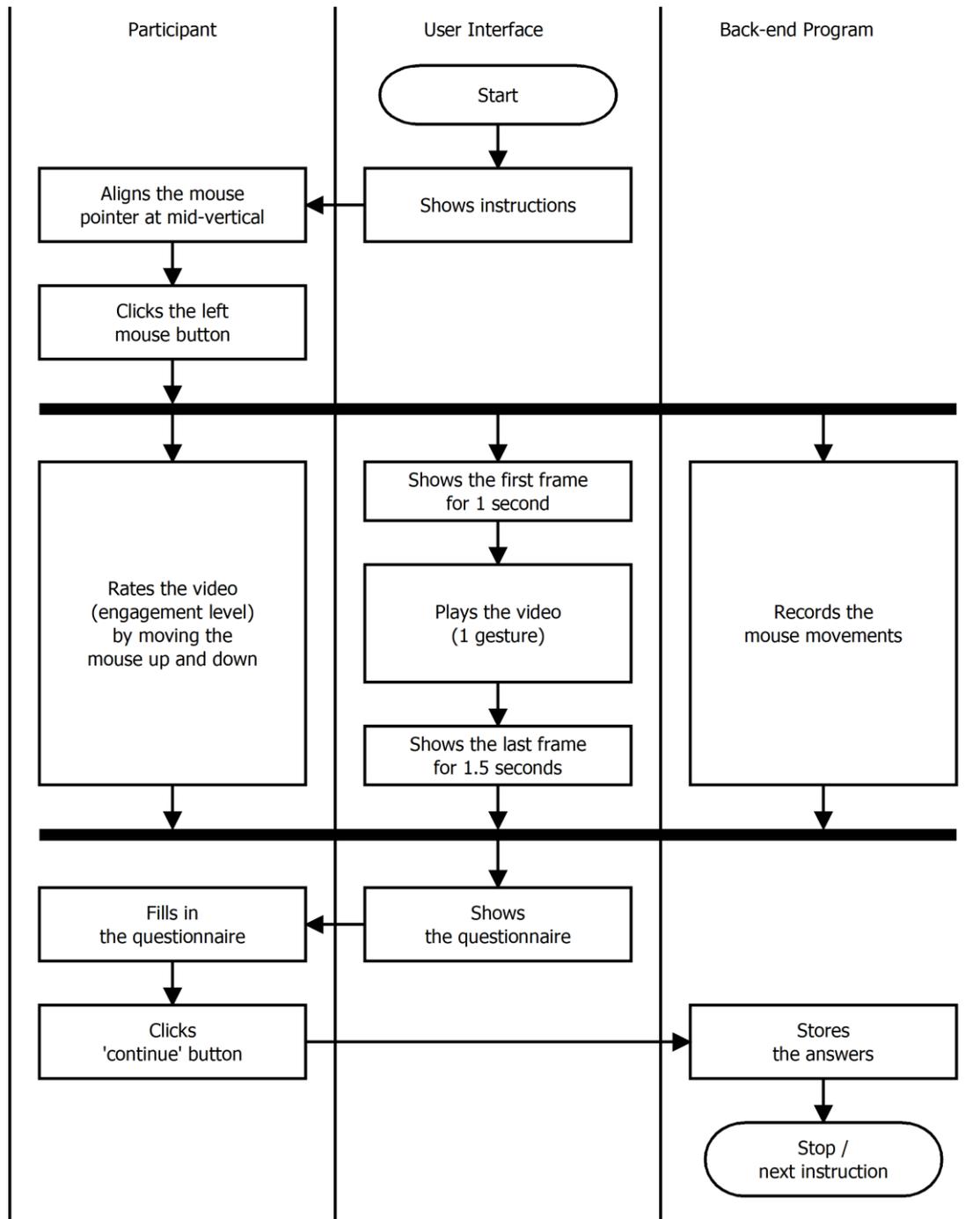


Figure 6.6 The process of evaluating the video

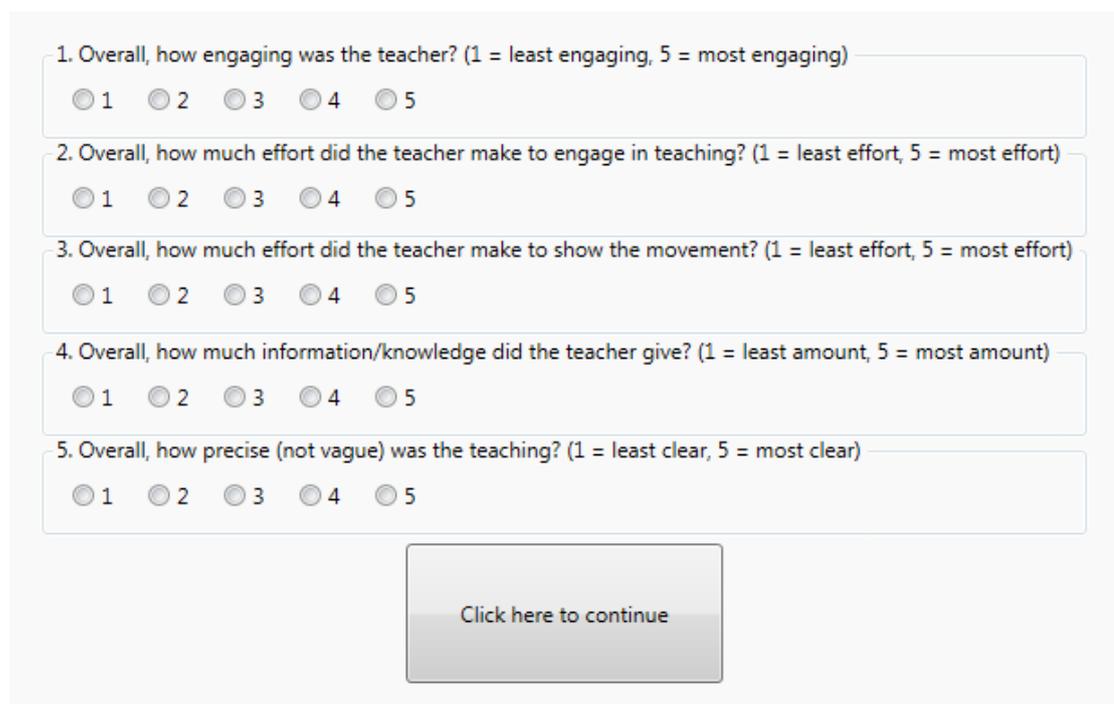
## 6.6. Data Collection

### 6.6.1. Post-trial Part

Using the same GUI program for rating the engagement level of the teachers, the participants were prompted with a post-trial questionnaire to complete. Once finished, the GUI panel in the program showed that the task was complete. The participants were then asked verbally if they had any comments or feedback regarding the experiment.

### 6.6.2. Dependent Measurement

Two sets of questionnaires were provided in the GUI program for the participants to complete. The first set of questions was the video questionnaire. The questionnaire was prompted by the program after playing each video. In total, each participant filled in 12 video questionnaires. The appearance of the video questionnaire is shown in Figure 6.7.



The image shows a screenshot of a video questionnaire interface. It consists of five vertically stacked question boxes, each containing a question and five radio button options labeled 1 through 5. Below the questions is a large rectangular button with the text "Click here to continue".

1. Overall, how engaging was the teacher? (1 = least engaging, 5 = most engaging)

1  2  3  4  5

2. Overall, how much effort did the teacher make to engage in teaching? (1 = least effort, 5 = most effort)

1  2  3  4  5

3. Overall, how much effort did the teacher make to show the movement? (1 = least effort, 5 = most effort)

1  2  3  4  5

4. Overall, how much information/knowledge did the teacher give? (1 = least amount, 5 = most amount)

1  2  3  4  5

5. Overall, how precise (not vague) was the teaching? (1 = least clear, 5 = most clear)

1  2  3  4  5

Click here to continue

**Figure 6.7 Video questionnaire**

In relation to the video questionnaire, the participants also input real-time rating data when watching the videos. The participants were asked to rate the level of the engagement of the teacher shown in the videos. They moved the computer mouse up and down to rate the engagement level. The level was indicated by a visible horizontal line that followed the mouse pointer movement. The horizontal line could be moved within a boundary which occupied the half of the screen. The participants were told that reaching the uppermost

position meant giving the maximum rating and thus, the lowermost position meant giving the minimum rating. To start evaluating a video, the participants needed to align the horizontal line in the middle of the screen.

The second questionnaire (final questionnaire) was prompted after the participants evaluated the whole set of videos. The appearance of the questionnaire is shown in Figure 6.8.

Final Questionnaire

For each of the gestures, how much effort do you think it should take to do the teachings?  
(1 = least effort, 5 = most effort)

A. 'Wind the bobbin up'  
 1  2  3  4  5

B. 'Pull pull'  
 1  2  3  4  5

C. 'Clap, clap, clap'  
 1  2  3  4  5

D. 'Point to the ceiling'  
 1  2  3  4  5

E. 'Point to the floor'  
 1  2  3  4  5

F. 'Put your hands upon your knee'  
 1  2  3  4  5

[Click here to continue](#)

**Figure 6.8 Final questionnaire**

### 6.6.3. Method

The data contributed by the participant was recorded electronically by the GUI program. When the participants completed a questionnaire, the program stored it onto a database-backed storage. The data included participants' demographics, video questionnaires and final questionnaires. The program recorded each value with a time stamp.

Mouse movements were recorded within a video playtime when the participants evaluated a video. The up and down movements were used by the participant to rate the level of the engagement of the teacher being shown in the video. Each recording had an ID that uniquely paired the recording with the video. The recording also covered the part where the video

player inserted time delays in the beginning before fully playing the video and at the end before closing the video.

## 6.7. Results

### 6.7.1. Participant

The experiment was conducted with 16 participants. They consisted of 11 male and 5 female participants. The age of the youngest participant was 21 and the oldest one was 63. The box plot of the age of the participants is shown in Figure 6.9.

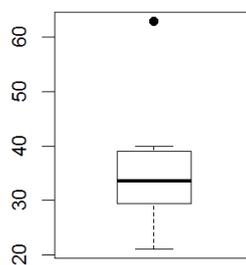
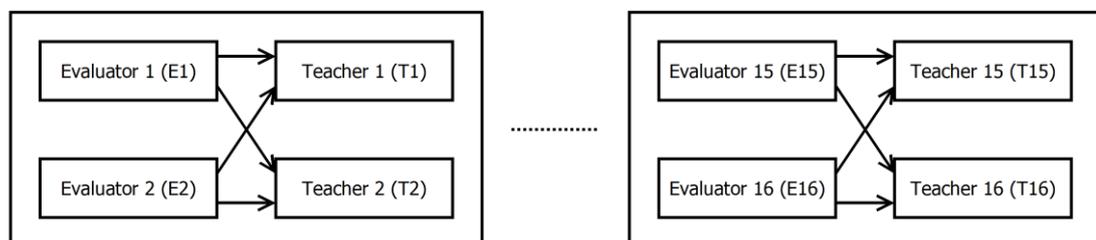


Figure 6.9 Age of the participants

From the 16 participants, 13 of them were involved as participants in the main experiment part 1 as "teachers". There was a 6-week interval from the end of part 1 to the start of part 2 and the time interval for the repeating participants was at least 9 weeks. The arrangement of the experiment assigned two participants (as "evaluators") to evaluate two teachers (one pair). In total, there were eight unique pairs. This pairing between the participants in this experiment as "evaluators" and the evaluated teachers is illustrated in Figure 6.10.



Note: Teachers were selected randomly by evaluators

Figure 6.10 Pairs of evaluators and teachers

## 6.7.2. Questionnaire Data Analysis

The experiment mainly recorded two data categories: (i) questionnaire and (ii) video rating. The length/amount of stored record of the questionnaire data is fixed, which reflects the number of questions. On the other hand, the video rating data has dynamic length. It depends on the duration of the video and how dynamically the participant moved the mouse while rating the teacher. This section discusses the analysis of the questionnaire data. The result of video rating data is discussed separately in Section 6.8 which compares the rating with the data from part 1 of the experiment.

Seven sets of questions were asked in the experiment to compare between the six gestures being evaluated the video. These sets come from six questions that were asked after the evaluator watched one gesture video, and one final question at the end of the experiment. The Friedman test was used to check whether there is any significant difference between gestures for each question. The results are shown in Table 6.1. All the p-values from the test are less than 0.05 which indicate that in each question there is one or more significant difference between gestures.

**Table 6.1 Results of significant difference tests**

Alias	Question	Friedman chi-squared	dF	p-value
QV1	Overall, how engaging was the teacher?	13.68098	5	0.017768
QV2	Overall, how much effort did the teacher make to engage in teaching?	27.83762	5	3.92E-05
QV3	Overall, how much effort did the teacher make to show the movement?	13.98698	5	0.015692
QV4	Overall, how much information/knowledge did the teacher give?	13.70166	5	0.01762
QV5	Overall, how precise (not vague) was the teaching?	18.13936	5	0.002777
FQ	For each of the gestures, how much effort do you think it should take to do the teachings?	23.47368	5	0.000274

To find which pairs have a significant difference, the Wilcoxon test was used to make paired tests between gestures in each question. All the results are shown in Table 6.2.

**Table 6.2 Wilcoxon signed-rank test p-value and the average value of the questionnaires**

QV1	A	B	C	D	E	Average
A						3.4375
B	0.166934					3.75
C	0.103156	0.714896				3.8125
D	0.401245	0.714829	0.511275			3.65625
E	0.402785	0.040388	0.025359	0.115873		3.25
F	0.570414	0.535536	0.364218	0.78059	0.200781	3.59375

QV2	A	B	C	D	E	Average
A						3.46875
B	0.598477					3.59375
C	0.025452	0.060023				4.03125
D	0.466627	0.875917	0.083928			3.65625
E	0.215239	0.055581	0.000316	0.03164		3.15625
F	0.779388	0.819701	0.040419	0.68702	0.107343	3.53125

QV3	A	B	C	D	E	Average
A						3.875
B	0.714906					3.96875
C	0.320578	0.515683				4.125
D	0.77384	0.981901	0.518016			4
E	0.118247	0.060681	0.014772	0.061623		3.483871
F	0.6323	0.427753	0.164842	0.399555	0.371525	3.65625

QV4	A	B	C	D	E	Average
A						3.4375
B	0.542557					3.59375
C	0.050286	0.156661				3.96875
D	0.06873	0.222714	0.721088			3.90625
E	0.949168	0.389241	0.017496	0.017067		3.46875
F	0.224901	0.549837	0.347525	0.517343	0.098395	3.75

QV5	A	B	C	D	E	Average
A						3.8125
B	0.321952					3.5625
C	0.62522	0.109863				4.03125
D	0.698436	0.471448	0.327568			3.8125
E	0.035218	0.298425	0.004516	0.055209		3.3125
F	0.831111	0.317513	0.465305	0.73848	0.019116	3.875

FQ	A	B	C	D	E	Average
A						3.75
B	0.951327					3.875
C	0.211227	0.157831				3.375
D	0.005009	0.001722	0.042833			2.625
E	0.005575	0.002239	0.061759	0.808008		2.6875
F	0.037689	0.020224	0.280486	0.324791	0.457167	3

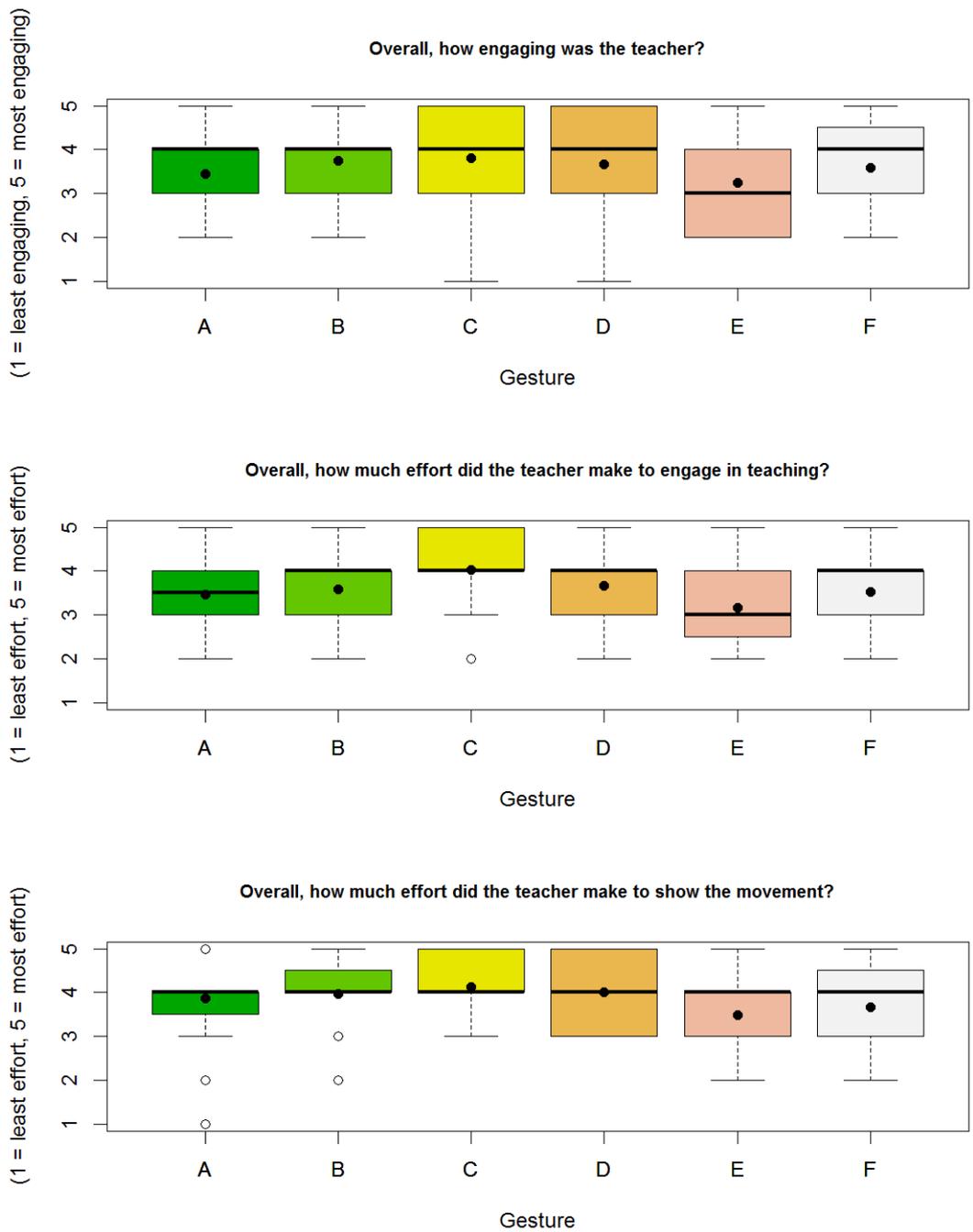
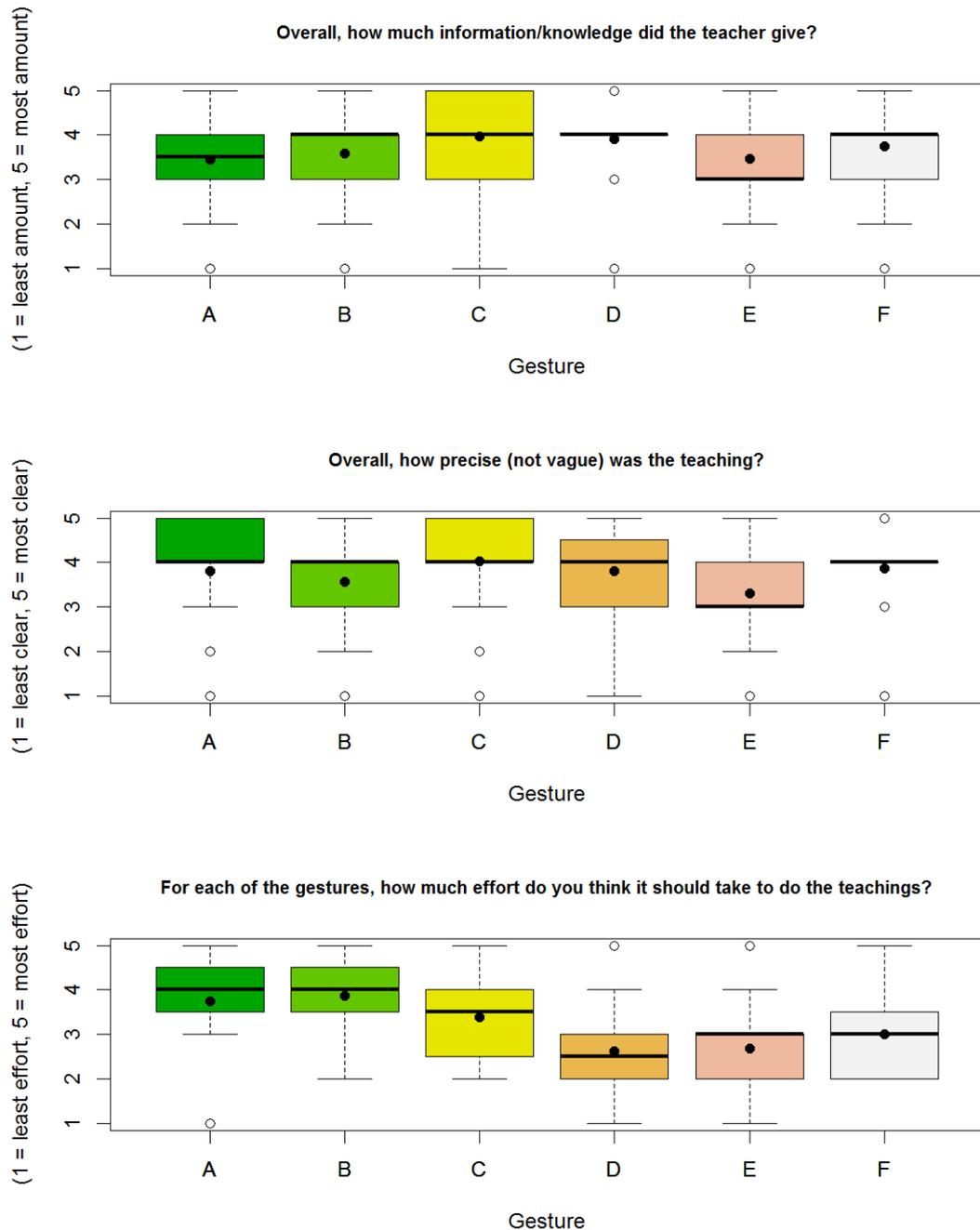


Figure 6.11 Results of QV1 (top), QV2 (mid), and QV3 (bottom) questionnaires



**Figure 6.12 Results of QV4 (top), QV5 (mid), and FQ (bottom) questionnaires**

From the results of average values and paired tests in Table 6.2 and the boxplot graphics as shown in Figure 6.11 and Figure 6.12, the following discusses the result of each question:

1. Overall, how engaging was the teacher? (QV1)

Gesture E ("Point to the floor") was perceived by the evaluators as where the teachers were less engaging in comparison to others. This gesture is particularly different statistically to Gesture B ("Pull, pull") and C ("Clap, clap, clap").

2. Overall, how much effort did the teacher make to engage in teaching? (QV2)

In terms of effort to engage, Gesture E received the lowest average score. In this case, it had a significant difference to Gesture C and D ("Point to the ceiling"). Gesture C was perceived as the highest effort to engage and was significantly different not only to Gesture E, but also to Gesture F ("Put your hands upon your knee").

3. Overall, how much effort did the teacher make to show the movement? (QV3)

In terms of effort in showing the movement, Gesture C was seen averagely as the highest effort, and in statistical comparison was significantly different to Gesture E which had the least average on the effort.

4. Overall, how much information/knowledge did the teacher give? (QV4)

The participants perceived Gesture C as highly informative. It was significantly different to Gesture D and E which was perceived as where the teachers were least informative.

5. Overall, how precise (not vague) was the teaching? (QV5)

Gesture C received the highest average score with regards to how precise (not vague) was the teaching. The gesture was significantly different in statistical paired comparison to Gesture F and E. Gesture E received the lowest score and had a significant difference to Gesture F.

6. For each of the gestures, how much effort do you think it should take to do the teachings? (QF)

The evaluators felt that Gesture D required the lowest effort to do the teaching. But statistically, there was no significant difference between Gesture D and the others. The second lowest average was Gesture E, and it had a significant difference to Gesture A ("Wind the bobbin up") and B which was seen to require the highest effort.

## 6.8. Comparing the Data from the Teacher and the Evaluator

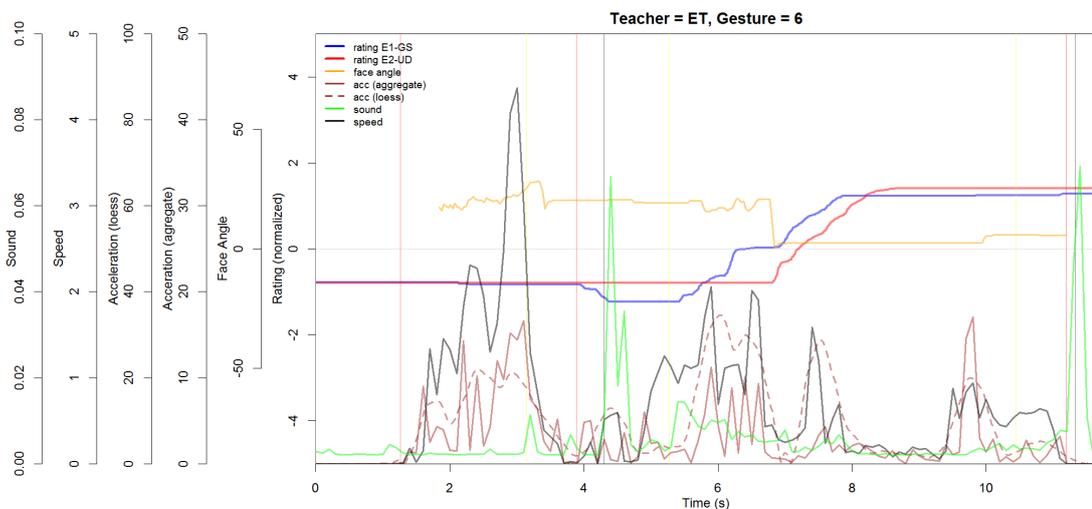
This research eventually aimed to find "measurable" events that are potentially useful for detecting the level of engagement of a human when interacting with a robot, especially in teaching a robot. The research evaluated these events by comparing what were the activities of human teachers that are perceived by other humans as engaging. Furthermore, these events were compared to the data that was captured/sensed by the robot. In this

section, the following discusses the comparison of rating data from the evaluators to several data recorded from the teachers.

### 6.8.1. Data Preparation

Two set of rating data were available for each teacher on each gesture. To evaluate the rating visually the ratings were plotted in graphs such as those shown in Figure 6.13. In this case, both sets of rating were normalized.

From the teachers, three sets of raw data were available for evaluation. They were: (i) video+audio, (ii) Kinect skeletal tracking, and (iii) Kinect face tracing. For the evaluation, the audio data from the video was extracted and used to show indications of voice activities. The sample rate of the audio was 48000 samples per second (sps). For the evaluation, the sample rate was reduced to 100 sps for automatic calculation through a program. The process to decrease the rate was using an aggregate function where every n samples were reduced to one average value. The signals were furthermore reduced to 10 sps for visual display purposes.



**Figure 6.13 Example plot of evaluation signals**

Two datasets were produced from the Kinect skeletal tracking. They were the speed and acceleration of the wrists' movements. The speed values plotted in Figure 6.13 are average values of both the left and right wrists. The accelerations also calculated as an average from both wrists but the calculation was using absolute value which made negative values as positive values.

The Kinect face tracking data was used in this evaluation to detect the face direction of the teachers when interacting with the robot. The plotted "face angle" in Figure 6.13 will be positive if the participant looked at the computer (0 meant looking straight to the robot).

### 6.8.2. Video Annotation Software

To evaluate the results of the main experiment (part 1 and part 2) the author developed a video annotation program. The appearance of the program is shown in Figure 6.14.

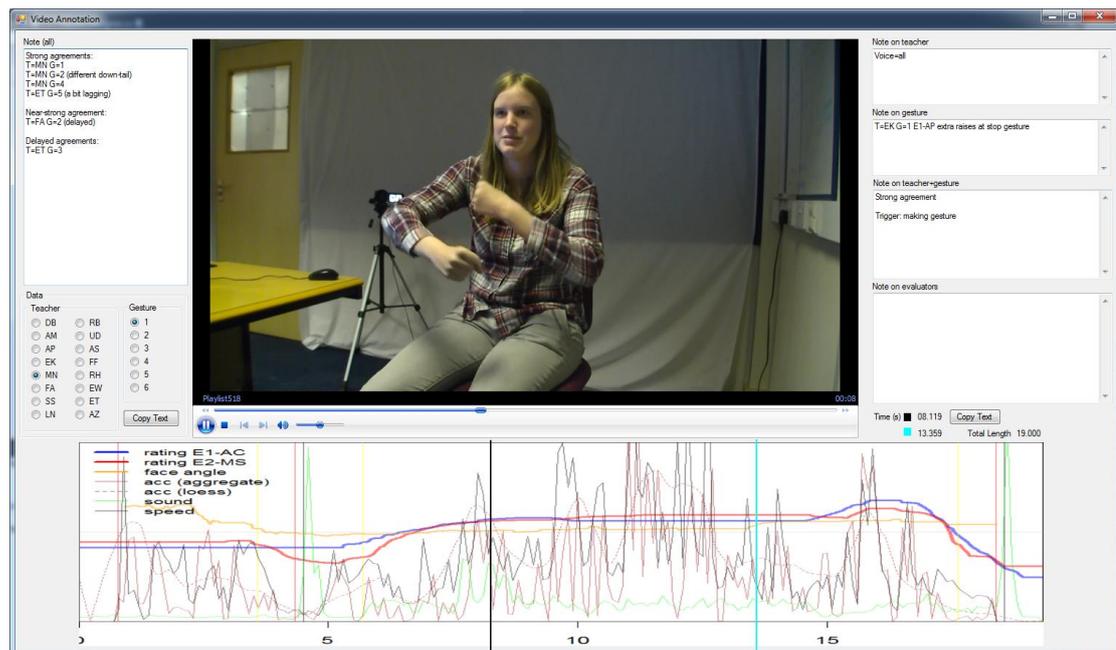


Figure 6.14 Video annotation program

The program was hard-coded with a list of datasets that can be selected with a single click of a mouse. Five text boxes were provided for storing notes and individually useful for evaluating (i) certain gestures, (ii) certain evaluators, (iii) certain teachers, (iv) a combination of teacher+gesture, and (v) general/all.

The program was developed to visually track events on a graph while playing the video. It was also able to jump to a certain time location by selecting the graph. The graph was produced by the data processing discussed in the previous section.

### 6.8.3. Behaviour Analysis

This research had tried to analyse the data by comparing the rating data to the pre-processed data produced from the teacher activity. Unfortunately, no general conclusion can be made from this as it is difficult to distinguish which data contributed to the dynamics

of the engagement rating. The research then tried to analyse how the human (evaluator) evaluated the level of engagement by observing what activities were shown in the video. The following discusses the gathered data regarding each set of evaluators (the sets of teachers and evaluators are shown earlier in Figure 6.10). For simplification, in each set, the evaluators are called  $E_A$  and  $E_B$  and the teachers are called  $T_A$  and  $T_B$ .

#### Evaluation Set:

1. In this set, both  $T_A$  and  $T_B$  said the name of gesture while teaching the robot. The difference was  $T_A$  said it normally, and  $T_B$  sang it. The face expression of  $T_A$ , in general, was consistently neutral throughout all the videos.  $T_B$  looked at the computer while demonstrating the gestures.

$E_B$  in this set used a regular pattern to raise the rating after both teachers commenced the gesture.  $E_B$  also exhibited a regular pattern to lower the rating at the end of the movement.

$E_A$  consistently always lagged behind  $E_B$  in raising the level of engagement. Sometimes  $E_A$  only moved the mouse up (raising the rating) near the end of the video.

2.  $T_A$  in this set said the gesture name when teaching except for the first gesture. The expression was neutral. On some occasions,  $T_A$  looked down to the robot and triggered the rating lower down.

$T_B$  in this set seemed anxious. This was confirmed with the answer in Godspeed questionnaire regarding "Perceived Safety". The subject ticked value number 1 (the leftmost) in anxious vs relaxed question.

$E_B$ , in general, moved the mouse down when the teachers lose eye contact. With regard to teacher  $E_A$ , the author did not see a general pattern that triggers the rating dynamics.

3.  $T_A$  said the gesture name with regular voice when making all gestures.  $T_B$  said the gesture name and made facial expressions as if the subject was teaching a toddler. In the first gesture,  $T_B$  did not produce any sounds.

Many of the ratings from  $E_A$  and  $E_B$  almost overlapped which may indicate that they might be triggered by the same factors. In general, both rate the engagement up when the teachers showed the movements.

- Both  $T_A$  and  $T_B$  said the gesture name while teaching.  $T_B$  added more words such as "winding it up for the third time". Compared to other teachers in different sets,  $T_B$  was considered very animating (more physical movements) when teaching the robot.

With  $T_A$ , both  $E_A$  and  $E_B$  seemed to assess in the opposite direction when making the rating. When  $E_A$  went up,  $E_B$  went down. With  $T_B$ , both  $E_A$  and  $E_B$  seemed to have similar agreements in making the rating.

- $T_A$  mostly gave a slight smile when doing the gestures.  $T_B$ 's expression was nearly always neutral for all the gestures. Both  $T_A$  and  $T_B$  said the name of gesture.

$E_A$  and  $E_B$  had strong agreements at least in three gestures of  $T_A$  and  $T_A$ . Like most evaluators, both raised the level up when the teacher started making the movement.

- $T_A$  said the gesture name slowly and also demonstrates at a slow pace.  $T_B$  always made his/her own "start gesture" before demonstrating the movement. The pose was to cross the arms over the chest.  $T_B$  always repeated the movement three times.  $T_B$ 's speed was relatively fast, but with clear pauses between intervals.

Both  $E_A$  and  $E_B$  has different onsets of timing across gestures when rating the engagement. There was not a clear pattern on which one triggered which.

- Both  $T_A$  and  $T_B$  never produced sounds for all of the gestures. The length of  $T_A$ 's videos was the longest among all the teaching participants. The subject always added additional movements to let the robot fully imitate the movement.  $T_B$  always focused on the computer monitor before doing the demonstration, but faced toward to the robot and made eye contact when starting the gesture. Both  $T_A$  and  $T_B$  moved relatively in slowly.

$E_A$  and  $E_B$  were both triggered by the animation of  $T_A$ , but at different onsets.

$E_A$  and  $E_B$  triggered to raise the level of engagement rating when  $T_B$  started to make eye contact with the robot.

- $T_A$  was the most voice active teacher among all of the teachers. For example, on the 4th gesture, the subject said "We're going to point to the ceiling" when making the gesture. Then said "That is right" at the end of the gesture.

$T_B$  never produced any sound except a slight coughing at the beginning of gesture 6.  $T_B$  moved at a slower pace.

For  $T_A$ , who used voice actively,  $E_B$  had a tendency to lag behind  $E_A$ . To  $T_B$ , who was silent, both ratings were raised when  $T_B$  started to make the gesture.

Based on the observations above the author saw some general patterns of what triggered the evaluator to rate the level of engagement higher. They were:

1. Animation of gesture
2. Explicit turning of head toward the robot ("eye contact")
3. Start smiling

The author also observed that losing eye contact (such as looking back to the computer while doing the gesture, or looking down to the robot leg) might trigger the evaluator to rate the level of engagement lower.

## 6.9. Conclusion

The study presented here was the second part of two sub-experiments that aimed to evaluate what elements are used by humans to evaluate the level of engagement of a human teacher when teaching a robot. Two software programs were developed to support the study. The first one was for capturing the engagement rating, and the second one for video annotation that worked effectively for the datasets in the experiment.

The behavioural analysis suggested a confirmation that immediacy affected the perceived level of engagement. While the literature research earlier in Chapter 2 and Chapter 3 discussed the relation in human-human interaction, this result from evaluating teachers that taught a robot might suggest that this immediacy effect also applied to the human-robot interaction.

The robot program developed for this study was able to record some physical human activity such as the joint movements, face tracking and video recording. Unfortunately, it needed more intensive study and further detailed analysis to use as data in this study. It may also be that more sensors and further sensor filtering may be necessary to be able to use it to measure the level of engagement. Not only that, exploration of algorithms might also be needed to successfully measure the level autonomously.

## Chapter 7. Conclusion and Future Directions

### 7.1. Conclusion

The study presented in this thesis had investigated what criteria could or should be used by a robot to measure whether a person is (or could potentially be) a good teacher. For this, the study evaluated the literature by looking at teacher-student relationships from a human-human interaction perspective. From the gathered literature two factors are considered to play an important role in defining a good teacher. They are *engagement* and *immediacy* elements, which are interconnected, for example, in a conversation, a person might use immediacy to increase the engagement of a conversation partner. These two factors are considered to play an important role for a teacher in a classroom setting to give a quality teaching.

The study has gathered a literature review to list sub-elements of engagement and immediacy. An experiment was then conducted to study these sub-elements that can be used by the robot in identifying a “good” teacher.

In the study, it was first decided which modality should be used in conducting an experiment in measuring the level of engagement. For this, an investigatory modality preference experiment was conducted to evaluate which modality the users prefer to teach a robot if the robot can be taught using voice, gesture demonstration, or physical manipulation. To support this modality preference experiment, a robotics software was developed in order to provide autonomous behaviour for a human participant to teach the KASPAR robot 5 arm gestures. Initially, the experiment was planned to include child participants to be conducted at local schools. However, due to equipment constraint, the experiment was only conducted in the robotics lab in the university with adult participants.

The main study aimed to measure the criteria that can be used by the robot to detect a good teacher. The main study was separated into two sub-experiments. In sub-experiment 1, each participant acted as a teacher to teach the robot a number of arm gestures. In sub-experiment 2 each participant acted as an evaluator of engagement of the participants in sub-experiment 1.

For sub-experiment 1, the robotics software used in the investigatory experiment was extended to allow the robot to imitate human’s arm movement.

For sub-experiment 2, two programs were developed. The first one was used mainly as an interface for the evaluator to rate the engagement level of the teacher in the video in real-time (while watching the video). The second program was a video annotation program. This program was hard-coded to handle the available datasets to allow easy evaluation when comparing the result from sub-experiment 1 and 2.

In sub-experiment 1 the experiment was equipped with a directional microphone to separate the sound from the human and mechanical noise from the robot. The recording was meant to help the evaluation of the teacher activity based on the voice. Unfortunately, the microphone had a non-adjustable auto muting feature and rendered the microphone unsuitable for recording participants who spoke with low intensity.

A robot program was developed to visually track the human and record the skeletal and face tracking data. The skeletal data was used to measure the speed and acceleration of arm wrists. The face tracking data was used to track the direction of the face toward the robot. The author had tried to compare this value to the rating data but general conclusions were hard to form as it was difficult to distinguish which data contributed to the dynamics of the engagement rating. Nevertheless, the plotted face direction data in a graph for video annotation was useful in locating the onset that triggers the raising of the engagement rating regarding face direction.

## 7.2. Findings and Review of the Research Questions

The following reviews the research questions proposed in Chapter 1.

### 1. *What input modalities do humans prefer in teaching a robot? (RQ1)*

The investigatory experiment presented in Chapter 3 was conducted in relation to select which input modality to be used in the main experiment which addressed RQ3. The task in this investigatory experiment was designed to be similar to the task in the main experiment.

The findings from the modality preference experiment suggested that the users appeared to have no preference in terms of human effort for completing the task. However, there was a significant difference in human enjoyment preferences of input modality and a marginal difference in the robot's perceived ability to imitate.

### 2. *How do humans evaluate the perception of engagement when evaluating a human teaching a robot? (RQ2)*

The main experiment presented in Chapter 5 and 6 was to address this research question. The results suggested that in human teaching of a robot (human-robot interaction), humans (the evaluators) also look for some of the immediacy cues, such as eye contact, that happen in human-human interaction for evaluating the engagement.

3. *Can physical activity measured by the robot be used to measure the level of engagement? (RQ3)*

Based mainly on the literature review in Chapter 3, the main experiment part 1 recorded physical activity data from the teachers to be compared to the engagement evaluation result from the main experiment part 2.

Unfortunately, no general conclusion can be made from this comparison as it is difficult to distinguish which data contributed to the dynamics of the engagement rating. The directional microphone that was expected to separate the human voice from the mechanical noise of the robot did not work as planned, so the comparison was done with the noise from the robot included.

This does not necessarily mean that physical data measured by the robot cannot be used to measure the level of engagement. A further in-depth study to analyse the data is needed. It may also be that more sensors and further sensor filtering may be necessary to be able to use it to measure the level of engagement. Not only that, exploration of algorithms might also be needed to successfully measure the level autonomously.

### 7.3. Future Directions

The modality preference experiment presented in this chapter was mainly to select what modality to be used for the interaction in the main study to evaluate the level of engagement of the human teacher. The results of the experiment indicated that a different study to fully focus on investigating the modality preferences was open to be explored. The software system could be developed further to accommodate more complex input interfaces. It would also be useful to conduct the same experiment with different user groups, e.g. children or individuals with special needs.

The main study tried to measure the level of engagement by using an experiment scenario that was open to multiple type immediacy cues to occur in the recording. This leads to data complexity which made it hard to distinguish which event affects the dynamics of the rating.

Further study with an experiment that is targeted to a specific type of immediacy might help data isolation thus make it easier to analyse the effect of that particular type of immediacy.

Further study might also consider adding more sensors and further sensor filtering to be able to use it to measure the level of engagement. Additionally, further focused study in the exploration of algorithms might also be needed to successfully measure the level autonomously.

## Bibliography

- Alibali, M. W. et al. (2001) 'Effects of Visibility between Speaker and Listener on Gesture Production: Some Gestures Are Meant To Be Seen', *Journal of Memory and Language*, 44(2), pp. 169–188.
- Andry, P. et al. (2001) 'Learning and Communication via Imitation: An Autonomous Robot Perspective', *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, 31(5), pp. 431–442.
- Antoniou, A.-S. et al. (2006) 'Gender and Age Differences in Occupational Stress and Professional Burnout between Primary and High-School Teachers in Greece', *Journal of Managerial Psychology*. Emerald Group Publishing Limited, 21(7), pp. 682–690.
- Argall, B. D. et al. (2009) 'A Survey of Robot Learning from Demonstration', *Robotics and Autonomous Systems*. Elsevier, 57(5), pp. 469–483.
- Argall, B. D. and Billard, A. G. (2010) 'A Survey of Tactile Human-Robot Interactions', *Robotics and Autonomous Systems*. Elsevier, 58(10), pp. 1159–1176.
- Aron, A. et al. (1992) 'Inclusion of Other in the Self Scale and the Structure of Interpersonal Closeness', *Journal of Personality and Social Psychology*. American Psychological Association, 63(4), pp. 596–612.
- Bakker, A. B. et al. (2008) 'Work Engagement: An Emerging Concept in Occupational Health Psychology', *Work & Stress*. Taylor & Francis, 22(3), pp. 187–200.
- Bartneck, C. et al. (2009) 'Measurement Instruments for the Anthropomorphism, Animacy, Likeability, Perceived Intelligence, and Perceived Safety of Robots', *International Journal of Social Robotics*, 1(1), pp. 71–81.
- Beattie, G. and Shovelton, H. (1999) 'Mapping the Range of Information Contained in the Iconic Hand Gestures that Accompany Spontaneous Speech', *Journal of Language and Social Psychology*. Sage Publications, 18(4), pp. 438–462.
- Bickmore, T. et al. (2010) 'Maintaining Engagement in Long-Term Interventions with Relational Agents', *Applied Artificial Intelligence*, 24(6), pp. 648–666.
- Billard, A. et al. (2008) 'Robot Programming by Demonstration', in *Springer Handbook of Robotics*. Springer, pp. 1371–1394.

- Bolt, R. A. (1980) “‘Put-That-There’’: Voice and Gesture at the Graphics Interface’, in *Proceedings of the 7th Annual Conference on Computer Graphics and Interactive Techniques*. ACM, pp. 262–270.
- Brand, R. J. et al. (2002) ‘Evidence for “Motionese”: Modifications in Mothers’ Infant-Directed Action’, *Developmental Science*, 5(1), pp. 72–83.
- Breazeal, C. (2004) ‘Social Interactions in HRI: The Robot View’, *IEEE Transactions on Systems, Man, and Cybernetics - Part C: Applications and Reviews*. IEEE, 34(2), pp. 181–186.
- Breazeal, C. and Aryananda, L. (2002) ‘Recognition of Affective Communicative Intent in Robot-Directed Speech’, *Autonomous Robots*, 12(1), pp. 83–104.
- Breazeal, C. and Scassellati, B. (2002) ‘Robots That Imitate Humans’, *Trends in Cognitive Sciences*, 6(11), pp. 481–487.
- Brown, L. and Howard, A. M. (2013) ‘Engaging Children in Math Education Using a Socially Interactive Humanoid Robot’, in *2013 13th IEEE-RAS International Conference on Humanoid Robots (Humanoids)*. IEEE, pp. 183–188.
- Cakmak, M. et al. (2010) ‘Designing Interactions for Robot Active Learners’, *IEEE Transactions on Autonomous Mental Development*, 2(2), pp. 108–118.
- Carbini, S. et al. (2006) ‘From a Wizard of Oz Experiment to a Real Time Speech and Gesture Multimodal Interface’, *Signal Processing*. Elsevier, 86(12), pp. 3559–3577.
- Cassell, J. et al. (1994) ‘Modeling the Interaction between Speech and Gesture’, in *Proceedings of the Sixteenth Annual Conference of the Cognitive Science Society*.
- Chao, C. et al. (2010) ‘Transparent Active Learning for Robots’, in *2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, pp. 317–324.
- Chartrand, T. L. and Bargh, J. A. (1999) ‘The Chameleon Effect: The Perception-Behaviour Link and Social Interaction’, *Journal of Personality and Social Psychology*. American Psychological Association, 76(6), pp. 893–910.
- Chesebro, J. L. and McCroskey, J. C. (2001) ‘The Relationship of Teacher Clarity and Immediacy with Student State Receiver Apprehension, Affect, and Cognitive Learning’, *Communication Education*, 50(1), pp. 59–68.
- Christophel, D. M. (1990) ‘The Relationships among Teacher Immediacy Behaviors, Student Motivation, and Learning’, *Communication Education*, 39(4), pp. 323–340.

- Clodict, A. et al. (2007) 'A Study of Interaction between Dialog and Decision for Human-Robot Collaborative Task Achievement', in *16th IEEE International Conference on Robot & Human Interactive Communication*, pp. 913–918.
- Collins, S. H. et al. (2009) 'Dynamic Arm Swinging in Human Walking', *Proceedings of the Royal Society of London B: Biological Sciences*. The Royal Society, 276(1673), pp. 3679–3688.
- Dautenhahn, K. (1994) 'Trying to Imitate — Step towards Releasing Robots from Social Isolation', in *From Perception to Action Conference*. IEEE, pp. 290–301.
- Dautenhahn, K. (1995) 'Getting to Know Each Other - Artificial Social Intelligence for Autonomous Robots', *Robotics and Autonomous Systems*, 16(2), pp. 333–356.
- Dautenhahn, K. et al. (2003) 'Learning by Experience from Others—Social Learning and Imitation in Animals and Robots', in Kühn, R., Menzel, R., Menzel, W., Ratsch, U., Richter, M. M., and Stamatescu, I.-O. (eds) *Adaptivity and Learning*. Springer, pp. 217–242.
- Dautenhahn, K. et al. (2009) 'KASPAR – A Minimally Expressive Humanoid Robot for Human-Robot Interaction Research', *Applied Bionics and Biomechanics*. Taylor & Francis, 6(3–4), pp. 369–397.
- Dautenhahn, K. and Nehaniv, C. L. (2002) 'The Agent-Based Perspective on Imitation', in *Imitation in Animals and Artefacts*. MIT Press, pp. 1–40.
- Decety, J. and Jackson, P. L. (2004) 'The Functional Architecture of Human Empathy', *Behavioral and Cognitive Neuroscience Reviews*, 3(2), pp. 71–100.
- Diaz, K. M. et al. (2015) 'Fitbit®: An Accurate and Reliable Device for Wireless Physical Activity Tracking', *International Journal of Cardiology*, 185, pp. 138–140.
- Donker, S. et al. (2001) 'Coordination between Arm and Leg Movements during Locomotion', *Journal of Motor Behavior*, 33(1), pp. 86–103.
- Fadiga, L. et al. (2002) 'Speech Listening Specifically Modulates the Excitability of Tongue Muscles: A TMS Study', *European Journal of Neuroscience*, 15(2), pp. 399–402.
- Feldenkrais, M. (1972) *Awareness Through Movement*. Harper & Row.
- Fitch, W. T. (2000) 'The Evolution of Speech: A Comparative Review', *Trends in Cognitive Sciences*. Elsevier, 4(7), pp. 258–267.
- Fitzpatrick, P. et al. (2014) 'A Middle Way for Robotics Middleware', *Journal of Software*

*Engineering for Robotics*, 5(2), pp. 42–49.

Friedrich, H., Münch, S., et al. (1996) 'Robot Programming by Demonstration (RPD): Supporting the Induction by Human Interaction', *Machine Learning*. Springer, 23(2), pp. 163–189.

Friedrich, H., Kaiser, M., et al. (1996) 'What Can Robots Learn from Humans?', *Annual Reviews in Control*. Elsevier, 20, pp. 167–172.

Fritsch, J. et al. (2005) 'Detecting "When to Imitate" in a Social Context with a Human Caregiver', in *Proceedings of the IEEE ICRA Workshop on The Social Mechanisms of Robot Programming by Demonstration*.

Gabel, M. et al. (2012) 'Full Body Gait Analysis with Kinect', in *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, pp. 1964–1967.

Ghahramani, Z. (2004) 'Unsupervised Learning', in *Advanced Lectures on Machine Learning*. Springer, pp. 72–112.

Glas, N. and Pelachaud, C. (2015) 'Definitions of Engagement in Human-Agent Interaction', in *The Sixth International Conference on Affective Computing and Intelligent Interaction*, pp. 944–949.

Goffman, E. (2008) *Behavior in Public Places*. Simon and Schuster.

Goldin-Meadow, S. (1999) 'The Role of Gesture in Communication and Thinking', *Trends in Cognitive Sciences*. Elsevier, 3(11), pp. 419–429.

Goldin-Meadow, S. and Morford, M. (1985) 'Gesture in Early Child Language: Studies of Deaf and Hearing Children', *Merrill-Palmer Quarterly*. Wayne State University Press, 31(2), pp. 145–176.

Goodrich, M. A. and Schultz, A. C. (2007) 'Human–Robot Interaction: A Survey', *Foundations and Trends in Human-Computer Interaction*. Now Publishers Inc., 1(3), pp. 203–275.

Graf, H. P. et al. (2002) 'Visual Prosody: Facial Movements Accompanying Speech', in *Proceedings of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition*. IEEE, pp. 396–401.

Di Gropello, E. and Marshall, J. H. (2005) 'Teacher Effort and Schooling Outcomes in Rural Honduras', in *Incentives to Improve Teaching*, pp. 307–357.

- Grunwald, G. et al. (2003) 'Programming by Touch: The Different Way of Human–Robot Interaction', *IEEE Transactions on Industrial Electronics*. IEEE, 50(4), pp. 659–666.
- Hale, J. L. and Burgoon, J. K. (1984) 'Models of Reactions to Changes in Nonverbal Immediacy', *Journal of Nonverbal Behavior*, 8(4), pp. 287–314.
- Hills, A. P. et al. (2014) 'Assessment of Physical Activity and Energy Expenditure: An Overview of Objective Measures', *Frontiers in Nutrition*, 1(5).
- Hoffmann, A. G. (1990) 'General Limitations on Machine Learning', in *European Conference on Artificial Intelligence*, pp. 345–347.
- Huang, L. et al. (2011) 'Virtual Rapport 2.0', in *Intelligent Virtual Agents*. Springer, pp. 68–79.
- Hugot, V. (2007) *Eye Gaze Analysis in Human-Human Interactions*. Royal Institute of Technology.
- Humphrey, C. M. and Adams, J. A. (2008) 'Compass Visualizations for Human-Robotic Interaction', in *Proceedings of the 3rd ACM/IEEE International Conference on Human Robot Interaction*. ACM, pp. 49–56.
- Iverson, J. M. and Goldin-Meadow, S. (2005) 'Gesture Paves the Way for Language Development', *Psychological Science*. Association for Psychological Science, 16(5), pp. 367–371.
- Jansen, B. and Belpaeme, T. (2006) 'A Computational Model of Intention Reading in Imitation', *Robotics and Autonomous Systems*, 54(5), pp. 394–402.
- Kaelbling, L. P. et al. (1996) 'Reinforcement Learning: A Survey', *Journal of Artificial Intelligence Research*, 4, pp. 237–285.
- Kaipa, K. N. et al. (2010) 'Self Discovery Enables Robot Social Cognition: Are You My Teacher?', *Neural Networks*. Elsevier, 23(8), pp. 1113–1124.
- Khan, Z. (1998) 'Attitudes towards Intelligent Service Robots', *NADA KTH, Stockholm*, 17.
- Kiesler, S. and Hinds, P. (2004) 'Introduction to This Special Issue on Human-Robot Interaction', *Human-Computer Interaction*, 19(1–2), pp. 1–8.
- Klassen, R. M. et al. (2013) 'Measuring Teacher Engagement: Development of the Engaged Teachers Scale (ETS)', *Frontline Learning Research*, 1(2), pp. 33–52.
- Knapp, M. L. et al. (2013) 'Nonverbal Communication in Human Interaction'. Cengage

Learning.

Kober, J. et al. (2013) 'Reinforcement Learning in Robotics: A Survey', *The International Journal of Robotics Research*, 32(11), pp. 1238–1278.

Kose-Bagci, H. et al. (2009) 'Effects of Embodiment and Gestures on Social Interaction in Drumming Games with a Humanoid Robot', *Advanced Robotics*, 23(14), pp. 1951–1996.

Kotsiantis, S. B. (2007) 'Supervised Machine Learning: A Review of Classification Techniques', *Informatica*, 31, pp. 249–268.

Kuh, G. D. et al. (2008) 'Unmasking the Effects of Student Engagement on First-Year College Grades and Persistence', *The Journal of Higher Education*. The Ohio State University Press, 79(5), pp. 540–563.

Langelaan, S. et al. (2006) 'Burnout and Work Engagement: Do Individual Differences Make a Difference?', *Personality and Individual Differences*. Elsevier, 40(3), pp. 521–532.

Li, Q. et al. (2009) 'Accurate, Fast Fall Detection Using Gyroscopes and Accelerometer-Derived Posture Information', in *2009 Sixth International Workshop on Wearable and Implantable Body Sensor Networks*. IEEE, pp. 138–143.

Lockerd, A. and Breazeal, C. (2004) 'Tutelage and Socially Guided Robot Learning', in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 3475–3480.

Loehr, J. et al. (2005) *The Power of Full Engagement: Managing Energy, Not Time, is the Key to High Performance and Personal Renewal*. Simon and Schuster.

Lombard, M. et al. (2000) 'Measuring Presence: A Literature-Based Approach to the Development of a Standardized Paper-and-Pencil Instrument', in *The Third International Workshop on Presence*.

Louis, K. S. and Smith, B. (1992) 'Cultivating Teacher Engagement: Breaking the Iron Law of Social Class', in Newmann, F. M. (ed.) *Student Engagement and Achievement in American Secondary Schools*. Teachers College Press, pp. 119–152.

Luinge, H. J. and Veltink, P. H. (2005) 'Measuring Orientation of Human Body Segments Using Miniature Gyroscopes and Accelerometers', *Medical and Biological Engineering and Computing*. Springer, 43(2), pp. 273–282.

Maccoby, E. E. (1992) 'The Role of Parents in the Socialization of Children: An Historical Overview', *Developmental Psychology*. American Psychological Association, 28(6), pp. 1006–

1017.

McNeill, D. (1985) 'So You Think Gestures are Nonverbal?', *Psychological Review*. American Psychological Association, 92(3), pp. 350–371.

Mehrabian, A. (1966) 'Immediacy: An Indicator of Attitudes in Linguistic Communication', *Journal of Personality*, 34(1), pp. 26–34.

Meltzer, L. et al. (2001) 'The Impact of Effort and Strategy Use on Academic Performance: Student and Teacher Perceptions', *Learning Disability Quarterly*, 24(2), pp. 85–98.

Meltzoff, A. N. (2007) 'The "Like Me" Framework for Recognizing and Becoming an Intentional Agent', *Acta Psychologica*, 124(1), pp. 26–43.

Mohri, M. et al. (2012) *Foundations of Machine Learning*. MIT Press.

Motoi, K. et al. (2003) 'Evaluation of a New Sensor System for Ambulatory Monitoring of Human Posture and Walking Speed Using Accelerometers and Gyroscope', in *SICE 2003 Annual Conference*. IEEE, pp. 1232–1235.

Munhall, K. G. et al. (2004) 'Visual Prosody and Speech Intelligibility: Head Movement Improves Auditory Speech Perception', *Psychological Science*. Association for Psychological Science, 15(2), pp. 133–137.

Nagai, Y. and Rohlfing, K. J. (2009) 'Computational Analysis of Motionese Toward Scaffolding Robot Action Learning', *IEEE Transactions on Autonomous Mental Development*, 1(1), pp. 44–54.

Nehaniv, C. L. and Dautenhahn, K. (2001) 'Like Me? Measures of Correspondence and Imitation', *Cybernetics and Systems*, 32(1–2), pp. 11–51.

Nehaniv, C. L. and Dautenhahn, K. (2002) 'The Correspondence Problem', in Dautenhahn, K. and Nehaniv, C. L. (eds) *Imitation in Animals and Artifacts*. MIT Press, pp. 41–62.

Nicolescu, M. N. and Mataric, M. J. (2003) 'Natural Methods for Robot Task Learning: Instructive Demonstrations, Generalization and Practice', in *Proceedings of the Second International Joint Conference on Autonomous Agents and Multiagent Systems*. ACM, pp. 241–248.

Obaid, M. et al. (2012) 'User-Defined Body Gestures for Navigational Control of a Humanoid Robot', in *Social Robotics*. Springer, pp. 367–377.

- Otteson, J. P. and Otteson, C. R. (1980) 'Effect of Teacher's Gaze on Children's Story Recall', *Perceptual and Motor Skills*. *Perceptual and Motor Skills*, 50(1), pp. 35–42.
- Oviatt, S. et al. (1997) 'Integration and Synchronization of Input Modes during Multimodal Human-Computer Interaction', in *Referring Phenomena in a Multimedia Context and Their Computational Treatment*, pp. 1–13.
- Oviatt, S. (1999) 'Ten Myths of Multimodal Interaction', *Communications of the ACM*, November, pp. 74–81.
- Oviatt, S. et al. (2004) 'When Do We Interact Multimodally?: Cognitive Load and Multimodal Communication Patterns', in *Proceedings of the 6th International Conference on Multimodal Interfaces*. ACM, pp. 129–136.
- Perzanowski, D. et al. (2001) 'Building a Multimodal Human-Robot Interface', *IEEE Intelligent Systems*. IEEE, 16(1), pp. 16–21.
- Peters, C. et al. (2005) 'Engagement Capabilities for ECAs', in *AAMAS'05 workshop Creating Bonds with ECAs*.
- Peters, C. et al. (2009) 'An Exploration of User Engagement in HCI', in *Proceedings of the International Workshop on Affective-Aware Virtual Agents and Social Robots*. ACM, p. 9:1-9:3.
- Pontzer, H. et al. (2009) 'Control and Function of Arm Swing in Human Walking and Running', *Journal of Experimental Biology*. The Company of Biologists, 212(4), pp. 523–534.
- Profanter, S. et al. (2015) 'Analysis and Semantic Modeling of Modality Preferences in Industrial Human-Robot Interaction', in *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, pp. 1812–1818.
- Quigley, M. et al. (2009) 'ROS: An Open-Source Robot Operating System', in *Proceedings of the ICRA Open-Source Software Workshop*.
- Robins, B. et al. (2010) 'Tactile Interaction with a Humanoid Robot for Children with Autism: A Case Study Analysis Involving User Requirements and Results of an Initial Implementation', in *Proceedings of the 19th IEEE International Symposium on Robot and Human Interactive Communication*. IEEE, pp. 704–711.
- Rohlfing, K. J. et al. (2006) 'How Can Multimodal Cues from Child-Directed Interaction Reduce Learning Complexity in Robots?', *Advanced Robotics*, 20(10), pp. 1183–1199.

- Roth, W.-M. (2001) 'Gestures: Their Role in Teaching and Learning', *Review of Educational Research*, 71(3), pp. 365–392.
- Russell, S. J. and Norvig, P. (2009) *Artificial Intelligence: A Modern Approach*. Prentice Hall.
- Rutter, R. A. and Jacobson, J. D. (1986) *Facilitating Teacher Engagement*.
- Salem, M. et al. (2011) 'A Friendly Gesture: Investigating the Effect of Multimodal Robot Behavior in Human-Robot Interaction', in *2011 RO-MAN: 20th IEEE International Symposium on Robot and Human Interactive Communication*,. IEEE, pp. 247–252.
- Salem, M. et al. (2012) 'Generation and Evaluation of Communicative Robot Gesture', *International Journal of Social Robotics*. Springer, 4(2), pp. 201–217.
- Sallis, J. F. (2000) 'Age-Related Decline in Physical Activity: A Synthesis of Human and Animal Studies', *Medicine & Science in Sports & Exercise*. American College of Sports Medicine, 32(9), pp. 1598–1600.
- Schaufeli, W. B. and Bakker, A. B. (2004) 'Job Demands, Job Resources, and Their Relationship with Burnout and Engagement: A Multi-Sample Study', *Journal of Organizational Behavior*. John Wiley & Sons, 25(3), pp. 293–315.
- Schüssel, F. et al. (2013) 'Influencing Factors on Multimodal Interaction during Selection Tasks', *Journal on Multimodal User Interfaces*. Springer, 7(4), pp. 299–310.
- Searle, J. (1969) 'What is a Speech Act?', in *Speech Acts: An Essay in the Philosophy of Language*. Cambridge University Press.
- Shen, Q. et al. (2008) 'Acting and Interacting Like Me? A Method for Identifying Similarity and Synchronous Behavior between a Human and a Robot', in *IEEE IROS Workshop on 'From Motor to Interaction Learning in Robots'*.
- Sheridan, T. B. (1992) 'Telerobotics, Automation, and Human Supervisory Control', *IEEE Technology and Society Magazine*.
- Sidner, C. L. . et al. (2004) 'Where to Look: A Study of Human-Robot Engagement', in *Proceedings of the 9th International Conference on Intelligent User Interfaces*. ACM, pp. 78–84.
- Sidner, C. L. and Dzikovska, M. (2002) 'Human - Robot Interaction: Engagement between Humans and Robots for Hosting Activities', in *Proceedings of the 4th IEEE International Conference on Multimodal Interfaces*. IEEE, pp. 123–128.

- Siegrist, J. (1996) 'Adverse Health Effects of High-Effort/Low-Reward Conditions', *Journal of Occupational Health Psychology*. Educational Publishing Foundation, 1(1), pp. 27–41.
- Skinner, E. A. and Belmont, M. J. (1993) 'Motivation in the Classroom: Reciprocal Effects of Teacher Behavior and Student Engagement Across the School Year', *Journal of Educational Psychology*. American Psychological Association, 85(4), pp. 571–581.
- Smith, A. (2006) 'Speech Motor Development: Integrating Muscles, Movements, and Linguistic Units', *Journal of Communication Disorders*. Elsevier, 39(5), pp. 331–349.
- Steinfeld, A. et al. (2009) 'The Oz of Wizard: Simulating the Human for Interaction Research', in *2009 4th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, pp. 101–107.
- Stiefelhagen, R. et al. (2004) 'Natural Human-Robot Interaction Using Speech, Head Pose and Gestures', in *Proceedings of 2004 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, pp. 2422–2427.
- Straube, B. et al. (2011) 'The Differentiation of Iconic and Metaphoric Gestures: Common and Unique Integration Processes', *Human Brain Mapping*, 32(4), pp. 520–533.
- Sutton, R. S. and Barto, A. G. (1998) *Reinforcement Learning: An Introduction*. MIT Press.
- Szafir, D. and Mutlu, B. (2012) 'Pay Attention!: Designing Adaptive Agents That Monitor and Improve User Engagement', in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, pp. 11–20.
- Takacs, J. et al. (2014) 'Validation of the Fitbit One Activity Monitor Device during Treadmill Walking', *Journal of Science and Medicine in Sport*. Elsevier, 17(5), pp. 496–500.
- Tan, H. Z. et al. (1994) 'Human Actors for the Design of Force-Reflecting Haptic Interfaces', *Dynamic Systems and Control*. American Society of Mechanical Engineers, 55(1), pp. 353–359.
- Tickle-Degnen, L. and Rosenthal, R. (1990) 'The Nature of Rapport and Its Nonverbal Correlates', *Psychological Inquiry*, 1(4), pp. 285–293.
- Trost, S. G. et al. (2002) 'Age and Gender Differences in Objectively Measured Physical Activity in Youth', *Medicine & Science in Sports & Exercise*. American College of Sports Medicine, 34(2), pp. 350–355.
- Uswatte, G. et al. (2005) 'Ambulatory Monitoring of Arm Movement Using Accelerometry:

An Objective Measure of Upper-Extremity Rehabilitation in Persons with Chronic Stroke', *Archives of Physical Medicine and Rehabilitation*, 86(7), pp. 1498–1501.

Wang, J.-S. et al. (2010) 'An Inertial-Measurement-Unit-Based Pen with a Trajectory Reconstruction Algorithm and Its Applications', *IEEE Transactions on Industrial Electronics*. IEEE, 57(10), pp. 3508–3521.

Wang, J.-S. and Chuang, F.-C. (2012) 'An Accelerometer-Based Digital Pen with a Trajectory Recognition Algorithm for Handwritten Digit and Gesture Recognition', *IEEE Transactions on Industrial Electronics*. IEEE, 59(7), pp. 2998–3007.

Wood, L. (2015) *Robot-Mediated Interviews: A Robotic Intermediary for Facilitating Communication with Children*. University of Hertfordshire.

Wood, L. J. et al. (2013) 'Robot-Mediated Interviews - How Effective Is a Humanoid Robot as a Tool for Interviewing Young Children?', *PLoS ONE*. Public Library of Science, 8(3), p. e59448.

Xu, Q. et al. (2013) 'Designing Engagement-Aware Agents for Multiparty Conversations', in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, pp. 2233–2242.

Zhang, Z. (2012) 'Microsoft Kinect Sensor and Its Effect', *IEEE MultiMedia*. IEEE, 19(2), pp. 4–10.

Zhu, X. (2010) 'Semi-Supervised Learning', *Encyclopedia of Machine Learning*. Springer.

Zuo, X. (2011) *Two Key Technologies for a Flexible Speech Interface: From the Perspective of Human–Robot Interaction*.

## Appendix A. Publication

This appendix lists the publication by the author at the time of the writing of the thesis that is related to this thesis together with a brief comment on the relationship to this research.

- Novanda, O., Salem, M., Saunders, J., Walters, M. L., & Dautenhahn, K. (2016). **What Communication Modalities Do Users Prefer in Real Time HRI?** 5th International Symposium on New Frontiers in Human-Robot Interaction 2016, 5-6 April 2016, Sheffield, UK arXiv:1606.03992

This paper discusses the work presented in Chapter 3 which was the investigatory experiment that evaluates what input modality humans prefer in teaching a robot.

# What Communication Modalities Do Users Prefer in Real Time HRI?

Ori Novanda<sup>1,2,3</sup>, Maha Salem<sup>3</sup>, Joe Saunders<sup>3</sup>, Michael L. Walters<sup>3</sup>, and Kerstin Dautenhahn<sup>3</sup>

**Abstract.** This paper investigates users' preferred interaction modalities when playing an imitation game with KASPAR, a small child-sized humanoid robot. The study involved 16 adult participants teaching the robot to mime a nursery rhyme via one of three interaction modalities in a real-time Human-Robot Interaction (HRI) experiment: voice, guiding touch and visual demonstration. The findings suggest that the users appeared to have no preference in terms of human effort for completing the task. However, there was a significant difference in human enjoyment preferences of input modality and a marginal difference in the robot's perceived ability to imitate.

## 1 INTRODUCTION

Humans often use multi-modal interaction in daily communication and frequently use speech, physical gesture, and eye gaze when communicating with each other. In contrast, people do not usually interact with machines in the same way they interact with other humans. For example, when we open the fridge door in the morning, we do not usually greet it as we would another person.

With the recent advances in technology, it is now quite common for people to speak to some machines. High-end consumer products such as smartphones and tablets have enough computing power to capture human speech and translate it into text commands. This allows people to use their voice to interact with the applications running on the device. This technology has given rise to digital virtual assistants such as: Siri [1] on the iOS platform, Google Now [2] on the Android platform, and Cortana [3] on the Windows platform. These systems enable people to get information simply by asking the device. For example, asking what the weather will be like, or when a flight will leave. Language learning programs, such as Duolingo [4], prompt users to say sentences and use a voice to text translation method to accept their answer.

Traditionally robots have been associated with factories for building products such as cars. However, robots are now increasingly being used in a number of application areas where people can interact with them in a more natural way, in some ways similar to how they would interact with living creatures, such as indicated in the survey by Leite et al. [5]. For example, Pleo [6] changes its behaviour depending on how the user interacts with it, and Fernaeus et al. [7] used it to learn how people play with a robotic animal. KASPAR, a child-size

humanoid robot, has primarily been developed as a mediator to interact with children with autism in order to encourage basic communication and social interaction skills [8]. The consumer and research robot NAO [9] has been programmed to fulfil many tasks, one of which is as a companion robot (see Dautenhahn [10]) such as used in the research by Baxter et al. [11].

Since Sheridan [12] first associated Human-Robot Interaction (HRI) with teleoperation of factory robotic platforms, HRI research has extended into a number of different research areas (Goodrich and Schultz [13]). One of the areas of particular interest in recent years is multi-modal interfaces for multi-modal interactions. Stiefelhagen et al. [14] suggested that multi-modal interfaces are required to facilitate natural interaction. When humans are interacting with machines that have some human-like characteristics, they have a tendency to anthropomorphise with the machine and communicate in ways similar to human-human communication [15]. One of the objectives of HRI is to make human-robot interaction easier, more intuitive and more user friendly. By providing a multi-modal interface it may help keep the users engaged and interact with them in a more familiar manner, similar in some ways to which they may interact with other humans.

Although interactive multi-modal systems have some distinct advantages, developing such systems poses many challenges. According to Turk [16], the performance of a multi-modal system depends on each unimodal technology. Currently each modality has its own ongoing progress as an active research field. For example, a survey by Argall and Billard [17] lists research that solely focuses on investigating the tactile input modality.

Developing multi-modal interactive systems requires a substantial amount of computing power and robust integration algorithms. The integration algorithm of the robot's sensing system needs to make decisions in real-time on which input to consider for giving an appropriate response or action through the robot's actuators. The system has to be powerful enough to process different inputs such as visual, audio, and gesture cues. Integrating these social queues to flow naturally throughout the interaction session will also consume additional processing power. Providing a robust input modality and fusion to integrate all input data is a technically challenging task. Many hours of work would need to be devoted just to prepare the robot for a relatively simple task. This is one of the reasons that some HRI studies use Wizard-of-Oz [18] approaches to run experiments. By using these approaches, limitations on the technology can be set aside and replaced by behind-the-scene controllers to produce behaviour for the robot which is perceived by users as autonomous.

The challenge of creating a multi-modal interactive robotic system has inspired the research in the current study which investigates users' preferences of input modality when providing

<sup>1</sup> Dept. of Electrical Engineering, Universitas Sumatera Utara, Indonesia, ori@usu.ac.id

<sup>2</sup> This author received a scholarship from the General Directorate of Higher Education of Ministry of Education and Culture of Indonesia

<sup>3</sup> Adaptive Systems Research Group, University of Hertfordshire, United Kingdom

information to a robot. The study was designed to ask users to experience three different modalities whilst delivering the same instructions to the robot.

## 2 RELATED WORK

The study took related research in Human-Computer Interaction (HCI) into consideration. As suggested by Kiesler and Hinds [19], and Breazeal [20], existing work in HCI offers rich resources and inspiration for research in HRI.

The experiment “Put That There” by Bolt [21] is widely considered a pioneering demonstration that first showed the value and opportunity of multi-modal interfaces over uni-modal interfaces in HCI. The experiment was conducted using speech and gesture as command channels to draw a map.

The multi-modal interface raised a question of when the system is capable of multi-modal interactions, will the users utilise the ability to interact multi-modally? Oviatt [22] discussed ten myths about multi-modal interaction that give useful guidance to researchers building multi-modal systems. He stated that with multi-modally capable systems, users tend to switch between uni-modal and multi-modal interaction with the multi-modal interactions being the most predictable, based on the type of action being performed. In a previous study Oviatt et al. [23] found that 86% of the time participants used multi-modal commands when navigating a map in order to move, add, modify, or calculate the distance between objects. For performing tasks that require no navigation of the map, such as printing the map, the participants interacted multi-modally less than 1% of the time.

Later, Oviatt et al. [24] conducted an experiment using a Wizard-of-Oz approach, and concluded that the cognitive load of the task will drive the users’ preference towards either uni-modal or multi-modal interaction. Tasks with higher difficulty will often cause the users to utilize the multi-modality of the system. With repetitive tasks, users would initially communicate multi-modally. Once the tasks became more familiar they then tended to prefer one particular interaction modality four times more often than interacting multi-modally.

Schüssel et al. [25] experimented using speech, gesture, and touch in multi-modal interactions to select graphical icons on a computer monitor. This experiment was also conducted using the Wizard-of-Oz approach and measured what modality was used and combined by the users to complete the task. The overall results of the modalities used were: touch (63.2%), speech (21.6%), gesture (11.2%), speech+gesture (3.6%), speech+touch (0.5%). None of the participants used speech+gesture+touch at the same time.

Carbini et al. [26] observed users’ preferences for using a story telling game. Each user was given a task to compose a coherent story from a set of objects on a computer screen. It was found that children could easily interact using speech and gesture as compared to adults. The results of the full dataset were: gesture (45%), speech (5%), gesture+speech (50%).

All of the research cited above was conducted in HCI domains, where the users interacted with computers. This current research is focused on the interaction between humans and robots. Presented below are some studies that are more closely related to research in HRI.

Research by Khan [27] surveyed 134 respondents about their preferred interaction modalities with a robot. One of the

questions asked in this survey was the preferred method of communicating with a service robot to take care of clothes on a couch, or when the robot is to inform the user that the task has been completed. The results showed that speech was the most preferred interaction modality (82%), followed by touch screen (63%), gestures (51%), and typing commands (45%). However, the results of this study are limited because the survey was conducted by asking participants to complete a questionnaire without the participants having interacted with an actual robot.

Salem et al. [28] conducted research to compare the preference of modality in HRI. In contrast to the current research, they investigated the output side of the multi-modal interface. They examined the perceptions of users regarding a robot when the robot provides information to the human uni-modally (voice only) and multi-modally (voice and gesture). It was found that the robot was evaluated more positively if it displayed non-verbal behaviours, such as hand and arm gestures along with speech, even if they do not semantically match the spoken utterances.

Humphrey and Adams [29] also conducted a study relevant to our current research, by measuring users’ preference for visualising a tele-operated robot’s compass. They compared two different compass visualisations: top-down and world-aligned. The top-down visualisation received higher preference, but there was no significant difference to the world-aligned visualisation

## 3 THE STUDY

The study presented in this paper builds on two main observations from the related work discussed above which are:

1. As described in [24], simple task interaction can be conducted sufficiently using a uni-modal system only.

2. Previous research established significant differences of modality preference one over another and the most-preferred modality also differed ([25], [26], and [27]).

Those considerations above come from the HCI research domain where humans interact with computers. This study puts them in HRI perspective, where humans interact with robots, to see whether they can be applicable to the HRI domain.

Based on the first observation (1), our research investigated further the modality comparison by conducting an experiment that asked users to do a simple-task, comparing the using of specific and different modalities in different sessions. Based on the second consideration (2), the study also evaluated which modality was most preferred.

This research aimed toward developing an autonomous humanoid robot that can perform a real-time multi-modal interaction. The developed system provides the capability to detect voice commands, and interprets gestures and touch. All processes run in parallel in real-time. In the discussion section, this paper presents the comparison of user preferences for the three input channel modalities when instructing the robot to move its arms.

The basic idea of the experiment for the research was to develop a robot that can be taught to dance following music. This idea was limited in the required capability in order to match the robot’s physical limitations in speed of movement. The dance was changed to a simple mime task, and the music was limited to a single nursery rhyme. With these changes, the experiment became teaching the robot to mime following a nursery rhyme. The robot could be instructed to move its arms

using voice commands, by the users' gestures, and by physically guiding the arms.

The experiment was run non-intrusively so that the users did not need to use gloves or markers. The users also did not have to wear a microphone or headphone. The voice command system used a speaker-independent system so it did not have to be trained prior to the experiment.

## 4 EXPERIMENT SETUP

This section describes the experimental setup for the study. The study was approved by the University of Hertfordshire Ethics Committee under protocol number a1213/10.



Figure 1. KASPAR Robot

### 4.1 The Robot

This research uses KASPAR [30], a child-alike humanoid robot (shown in Figure 1). It has 17 Degrees of Freedom (DoFs) and has an internal PC to run the robot autonomously. The robot uses eSpeak [31] text-to-speech engine for speaking.

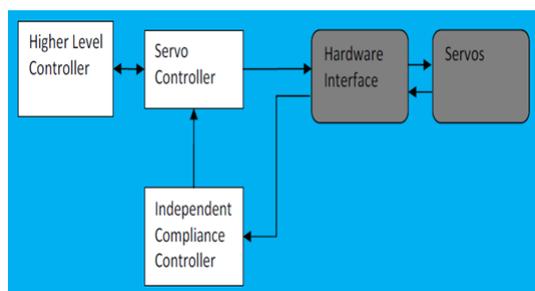


Figure 2. Compliance Mechanism

For the study, a program was developed to feature a servo compliance system. The block diagram of the compliance system is shown in Figure 2. It has a controller that measures the servos' torque values. This measurement is used to allow the software to detect whether the arms are being moved by an external force. It will then adjust the servos' positions to comply with the external

force. With this feature, users can move KASPAR's arms without breaking the servos. This controller works independently and can override any arm movement commands sent by the higher level controller.

In the current implementation, there was a time delay in the compliance controller's loop path introduced by the hardware interface. This made the control bandwidth of the servos only achieve 1 Hz, which is lower than the human force control bandwidth which is around 20 Hz [32]. This made the arms slightly stiff to move.

The system used an additional external PC beside the internal PC. The PC's communicated using TCP/IP through an Ethernet connection. The robot was built to have a WiFi connection as well but this wireless connection was never used in the experiment because of the latency in data transmission.

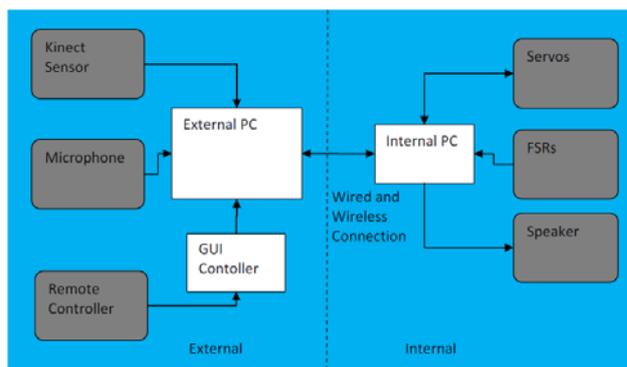


Figure 3. System Architecture

The external PC runs the high demand processes, such as the gesture detection and speech recognition. The global architecture of the system can be seen in Figure 3. The GUI controller runs on the external PC and sends commands to the internal PC to control the robot. The robot has several force sensitive resistor (FSR) sensors to detect touches. They are located on both palms and on the upper arms. This research did not restrict the participants on where they could touch the robot when moving its arms. During the experiment, the system only used the compliance system mentioned above to allow the participants to move the robot's arms physically.

### 4.2 Sensors

KASPAR was equipped with sensors to provide the following input modalities: (i) voice command, (ii) gesture, and (iii) touch. The developed system uses the Microsoft speech recognition engine. With non-intrusive interaction in mind, the system uses a directional microphone to listen to the user's voice. The microphone location was adjusted so the sound coming from the robot (voice and mechanical servo movements) was less likely to disturb the user's voice.

The speech recognition engine was programmed to detect 5 different commands that could be used to instruct the robot to move its arms. The robot has colour markers on its fingers (see Figure 1) to refer to the arms by colour instead of left and right (the former was deemed to be easier for participants to use when facing the robot). The markers are red and blue. The commands are: (i) red up, (ii) blue up, (iii) arms open, (iv) red down, and (v)

blue down. As suggested by the name, ‘up’ and ‘down’ commands will instruct the corresponding red or blue arm to go up or down. The ‘arms open’ command will make both arms open wide.

The system could only detect one particular command at a time. After saying a command, the user was expected to wait for the robot to respond before saying the next command.

A Microsoft Kinect was used by the system to detect the human partner's gestures. The Kinect SDK provided a skeleton representation of the user's position and pose. The position of the wrists were measured and interpreted as commands to move the robot arms. The system was programmed so that it only detected 5 positions, which were equivalent to the 5 voice commands.

Touch input modality was provided to the robot by using the developed compliance system. The users could move the robot's arms by moving the arm directly. They could hold any part of the arm in order to move it e.g. the users could move the arms by moving the upper arm or moving the hand. The latter requires smaller force because it is further away from the shoulder joint.

### 4.3 Layout

The physical layout of the experiment is shown in Figure 4. The robot was ‘sitting’ on the table and the Kinect sensor was located next to the robot. Video cameras were used to record the activities during the experiment sessions.



Figure 4. Experiment layout

Next to the robot was an instruction sign (see Figure 5) which reminded the user of the five instructions that could be used to control the robot. The instruction sign showed arrows to reflect the direction of the arms movement.

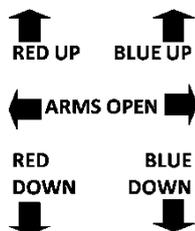


Figure 5. Instruction sign

### 4.4 Interaction Scenario

The task given to the participants in this study was teaching a humanoid robot to mime to a rhyme. The rhyme was ‘Hickory Dickory Dock’. The participants had to instruct the robot to move the arms to mime by following the lines of the rhyme. The task was repeated in several sub-sessions by only allowing one or two of these modalities in each session: voice, gesture, touch, and voice+gesture.

### 4.5 Experiment Procedure

Before starting the experiment, the participants completed a demographic questionnaire and signed a consent form.

The experiment was divided into two main sessions:

#### 1. Introduction session

In the beginning, the participant was introduced to the robot and asked to shake its hand. This was to familiarise the participants with the robot, and to let them know that it was fine to physically move its ‘red arm’ (right arm), even though it felt slightly stiff. Next, they were introduced to the nursery rhyme, and told what to do during the main trial session. The participants were also instructed on how to move the arms using each input modality.

During the introduction session, the robot was operated semi-autonomously using a wireless clicker to advance between sub-sessions. At the end of the introduction session, the participants were told that the following was the main trial, and the robot would run fully autonomously.

#### 2. Main trial session

In the main trial, the participants were left alone interacting with the robot which ran autonomously. The investigator stayed in the same room reading a book and sat back-facing the participants at a table without any computer or electronics devices. The participants were told that in case of emergency or if they wanted to stop, they could notify the investigator at any time.

The trial was run individually with a single participant for each trial session. The robot first asked the participants to instruct it on how to move in order to follow the nursery rhyme. The robot said the rhyme, and the participant should then instruct the robot to move for each line of the rhyme. The participant could instruct the robot to move the arms while the robot said the rhyme, except in the voice command mode session, where the participants were instructed (by the robot) to say the command after the robot has finished saying the rhyme. In the touch modality sessions, the participants had to move forward close to the robot to move its arms.

In total, there were 4 sub-sessions in the main trial. Each sub-session presented to the participant a different input modality. The first three were arranged so each participant had a different order of voice, gesture, and touch modalities. In total there were 9 possible different orders. In the fourth sub-session, the participant was asked to instruct the robot using a freely chosen combination of gesture and voice commands. After each sub-session, the robot performed the complete ‘dance’ with movements and timings specified by the commands that had been given by the participant.

After the main trial session, a second questionnaire recorded the users’ preferences of the methods to teach the robot. Before the whole session ended, the participants were also asked

verbally whether they had any comments they wanted to express regarding the experiment.

#### 4.6 Dependent Measurements

The post-trial questionnaire asked four questions using the Likert scale, and the participants rated their answers on a scale from 1 to 5. The first one was “Did you fully understand what instructions KASPAR said during the main session?” (1 being “not very well” and 5 being “very well”).

The second question was “In terms of effort, how did you feel about the different methods to teach KASPAR to dance?” (1 being “very hard” and 5 being “very easy”).

The third question was “In terms of enjoyment, how did you feel about the different methods to teach KASPAR to dance?” (1 = least enjoyable, 5 = most enjoyable).

The fourth question asked “When KASPAR showed what it had learned, how well did you feel KASPAR followed your instruction?” (1 = not very well, 5 = very well).

Every question from 2 to 4 had separate answers for each interaction modality.

## 5 RESULTS

The experiment was conducted with 16 participants; six females and 10 males aged 20 to 48 years old. They were recruited from the university staff and students. The invitation was advertised verbally and they were given a link of an online scheduler (Doodle [33]) to pick the available time slots that were suitable for them. In each gender category, 1 person was very familiar with robotic systems, while none had a prior knowledge of the robot setup that was used in this experiment.

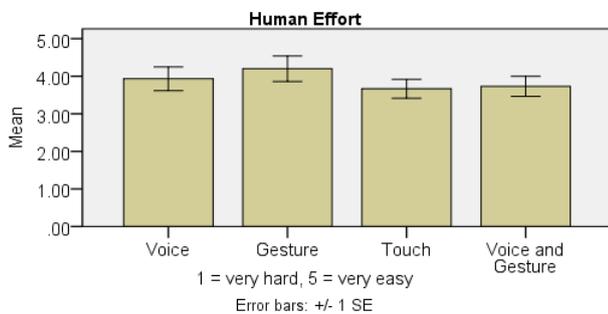


Figure 6. Questionnaire result on human effortlessness

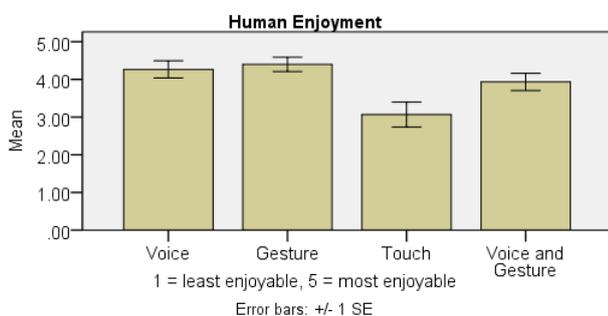


Figure 7. Questionnaire result on human enjoyment

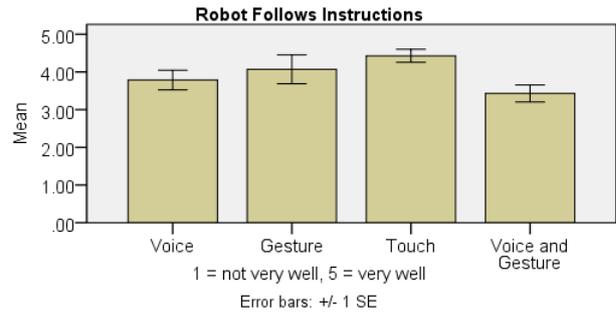


Figure 8. Questionnaire result on different instruction modalities

For the first question of the questionnaire, that asked whether the participants fully understood what the robot said during the experiment, no participant selected a value lower than 4. The mean score was 4.56 (SD = 0.51). The middle point of the answer was weighted as 3.

The questionnaire result on the effort to teach the robot to dance is shown in Figure 6. The data were checked using one-way repeated-measures ANOVA. The result was  $F(3,42) = 0.848$ ,  $p = 0.476$ , which meant none was significant. The result suggests that no particular modality is perceived as harder than the others.

The result that is shown in Figure 7 shows participants' perceived enjoyment of conducting the task for each modality. The touch modality received the least enjoyable rating. The statistical analyses indicated a significant difference in preferences,  $F(3,42) = 6.461$ ,  $p = 0.001$ . The pairwise comparisons indicated that there was a significant difference ( $p = 0.008$ ) between participants ratings for gesture (M = 4.4, SD = 0.74) and touch (M = 3.07, SD = 1.28) interaction modalities.

Finally, Figure 8 shows the participants' perception of the robot's ability to follow instructions. The difference was marginally significant,  $F(3,39) = 2.56$ ,  $p = 0.069$ . The pairwise comparisons showed a preference ( $p = 0.011$ ) for touch (M = 4.43, SD = 0.65) over voice+gesture (M = 3.43, SD = 0.85).

## 6 DISCUSSION

This research has investigated a robotic system that can be taught movements to follow a nursery rhyme. The development of the software is only presented briefly as it would be better to be presented as a technical paper. Three modalities were provided as input channels to give information to the robot as commands to move its arms. They are voice, gesture, and touch. Two modalities were provided as output channels: voice and gesture. The robot operated autonomously during individual sessions. The robot had touch-compliance which allows humans to physically move its arms into a desired pose. The system supported integration of multiple modalities through a TCP/IP-based inter-process communication mechanism. The experiment was conducted with adult participants.

The research findings indicated that being given a task which was to teach a robot to mime actions that follow a nursery rhyme, there was no statistically significant difference in preference ratings regarding human effort.

In contrast, there were favourable preferences regarding the human enjoyment. The touch modality was the least preferred and the gesture modality was rated the highest. The authors argue that the touch modality scored lowest due to the participants worrying about breaking the arms of the robot. This was because the compliance only controlled the arms compliance at a 1 Hz cycle rate instead of 20 Hz (see [32]).

For the robot's perceived ability to follow instructions, touch modality received the highest rating. The combined voice+gesture modalities received the lowest. This could be due to the robot only performing the instructed action after the voice command had completed, while the action after the gesture mode interaction was followed immediately. However, they were not statistically significant at the 5 % level, and only indicated a trend towards higher mean preference to the touch modality.

In general, without considering the task, the results are in contrast to the result in [25], [26], and [27]. However, this contrast indicates an agreement with [22] and [24], namely that for certain tasks humans can communicate to robots effectively using a uni-modal communication channel.

## 7 FUTURE WORK

This research is eventually aiming to evaluate how best to teach a robot and what constitutes an effective teaching strategy. The work presented here is an initial attempt towards that direction, and further research is required. The software system could be further developed to accommodate more complex input interfaces. It would also be useful to conduct the same experiment with different user groups, e.g. children or people with special needs.

## 8 REFERENCES

- [1] "Siri." [Online]. Available: <https://www.apple.com/uk/ios/siri/>. [Accessed: 03-Jul-2015].
- [2] "Google Now." [Online]. Available: <https://www.google.com/landing/now/>. [Accessed: 03-Jul-2015].
- [3] "Cortana." [Online]. Available: [www.microsoft.com/en-mobile/experiences/campaign-cortana/](http://www.microsoft.com/en-mobile/experiences/campaign-cortana/). [Accessed: 03-Jul-2015].
- [4] L. von Ahn, "Duolingo: learn a language for free while helping to translate the web," in *Proceedings of the 2013 international conference on Intelligent user interfaces*, 2013, pp. 1–2.
- [5] I. Leite, C. Martinho, and A. Paiva, "Social robots for long-term interaction: a survey," *Int. J. Soc. Robot.*, vol. 5, no. 2, pp. 291–308, 2013.
- [6] "Pleoworld.com." [Online]. Available: [pleoworld.com](http://pleoworld.com). [Accessed: 03-Jun-2015].
- [7] Y. Fernaeus, M. Håkansson, M. Jacobsson, and S. Ljungblad, "How do you play with a robotic toy animal?: a long-term study of pleo," in *Proceedings of the 9th international Conference on interaction Design and Children*, 2010, pp. 39–48.
- [8] B. Robins, K. Dautenhahn, and P. Dickerson, "From isolation to communication: a case study evaluation of robot assisted play for children with autism with a minimally expressive humanoid robot," in *Advances in Computer-Human Interactions, 2009. ACHI'09. Second International Conferences on*, 2009, pp. 205–211.
- [9] D. Gouaillier, V. Hugel, P. Blazevic, C. Kilner, J. Monceaux, P. Lafourcade, B. Marnier, J. Serre, and B. Maisonnier, "The nao humanoid: a combination of performance and affordability."
- [10] K. Dautenhahn, "Socially intelligent robots: dimensions of human-robot interaction," *Philos. Trans. R. Soc. B Biol. Sci.*, vol. 362, no. 1480, pp. 679–704, 2007.
- [11] P. Baxter, T. Belpaeme, L. Canamero, P. Cosi, Y. Demiris, V. Enescu, A. Hiole, I. Kruijff-Korbayova, R. Looije, M. Nalin, and others, "Long-term human-robot interaction with young users," in *IEEE/ACM Human-Robot Interaction 2011 Conference (Robots with Children Workshop)*, 2011.
- [12] T. B. Sheridan, *Telerobotics, automation, and human supervisory control*. MIT press, 1992.
- [13] M. A. Goodrich and A. C. Schultz, "Human-robot interaction: a survey," *Found. trends human-computer Interact.*, vol. 1, no. 3, pp. 203–275, 2007.
- [14] R. Stiefelhagen, C. Fügen, P. Giesemann, H. Holzapfel, K. Nickel, and A. Waibel, "Natural human-robot interaction using speech, head pose and gestures," in *Intelligent Robots and Systems, 2004.(IROS 2004). Proceedings. 2004 IEEE/RSJ International Conference on*, 2004, vol. 3, pp. 2422–2427.
- [15] D. Perzanowski, A. C. Schultz, W. Adams, E. Marsh, and M. Bugajska, "Building a multimodal human-robot interface," *Intell. Syst. IEEE*, vol. 16, no. 1, pp. 16–21, 2001.
- [16] M. Turk, "Multimodal interaction: A review," *Pattern Recognit. Lett.*, vol. 36, pp. 189–195, 2014.
- [17] B. D. Argall and A. G. Billard, "A survey of tactile human-robot interactions," *Rob. Auton. Syst.*, vol. 58, no. 10, pp. 1159–1176, 2010.
- [18] A. Steinfeld, O. C. Jenkins, and B. Scassellati, "The oz of wizard: simulating the human for interaction research," in *Human-Robot Interaction, 2009 4th ACM/IEEE Int. Conf. on*, 2009, pp. 101–107.
- [19] S. Kiesler and P. Hinds, "Introduction to this special issue on human-robot interaction," *Human-Computer Interact.*, vol. 19, no. 1–2, pp. 1–8, 2004.
- [20] C. Breazeal, "Social interactions in HRI: the robot view," *Syst. Man, Cybern. Part C Appl. Rev. IEEE Trans.*, vol. 34, no. 2, pp. 181–186, 2004.
- [21] R. A. Bolt, "'Put-that-there': Voice and gesture at the graphics interface," *ACM SIGGRAPH Comput. Graph.*, vol. 14, no. 3, pp. 262–270, Jul. 1980.
- [22] S. Oviatt, "Ten myths of multimodal interaction," *Commun. ACM*, vol. 42, no. 11, pp. 74–81, Nov. 1999.
- [23] S. Oviatt, A. DeAngeli, and K. Kuhn, "Integration and synchronization of input modes during multimodal human-computer interaction," in *Referring Phenomena in a Multimedia Context and their Computational Treatment*, 1997, pp. 1–13.
- [24] S. Oviatt, R. Coulston, and R. Lunsford, "When do we interact multimodally?: cognitive load and multimodal communication patterns," ... *Conf. Multimodal interfaces*, pp. 129–136, 2004.
- [25] F. Schüssel, F. Honold, and M. Weber, "Influencing factors on multimodal interaction during selection tasks," *J. Multimodal User Interfaces*, vol. 7, no. 4, pp. 299–310, 2013.
- [26] S. Carbini, L. Delphin-Poulat, L. Perron, and J.-E. Viallet, "From a wizard of Oz experiment to a real time speech and gesture multimodal interface," *Signal Processing*, vol. 86, no. 12, pp. 3559–3577, 2006.
- [27] Z. Khan, "Attitudes towards intelligent service robots," *NADA KTH, Stock.*, vol. 17, 1998.
- [28] M. Salem, S. Kopp, I. Wachsmuth, K. Rohlfing, and F. Joubin, "Generation and evaluation of communicative robot gesture," *Int. J. Soc. Robot.*, vol. 4, no. 2, pp. 201–217, 2012.
- [29] C. M. Humphrey and J. A. Adams, "Compass visualizations for human-robotic interaction," in *Proceedings of the 3rd ACM/IEEE int. conf. on Human robot interaction*, 2008, pp. 49–56.
- [30] K. Dautenhahn, C. L. Nehaniv, M. L. Walters, B. Robins, H. Kose-Bagci, N. A. Mirza, and M. Blow, "KASPAR—a minimally expressive humanoid robot for human-robot interaction research," *Appl. Bionics Biomech.*, vol. 6, no. 3–4, pp. 369–397, 2009.
- [31] "eSpeak." [Online]. Available: <http://espeak.sourceforge.net/>. [Accessed: 03-Jul-2015].
- [32] H. Z. Tan, M. A. Srinivasan, B. Eberman, and B. Cheng, "Human factors for the design of force-reflecting haptic interfaces," *Dyn. Syst. Control*, vol. 55, no. 1, pp. 353–359, 1994.
- [33] "Doodle." [Online]. Available: [doodle.com](http://doodle.com). [Accessed: 03-Jun-2015].

## Appendix B. Ethics Approval Documents

**UNIVERSITY OF HERTFORDSHIRE  
FACULTY OF SCIENCE, TECHNOLOGY AND CREATIVE ARTS**

## **M E M O R A N D U M**

**TO** Ori Novanda  
**CC** Kerstin Dautenhahn  
**FROM** Dr Simon Trainis – Chair, Faculty Ethics Committee  
**DATE** 5 November 2012

---

Your Ethics application for your project entitled:

**Study to evaluate which modality of interaction is important and under which conditions in imitation learning for a humanoid robot**

has been granted approval and assigned the following Protocol Number:

**1213/10**

This approval is valid:

**From 5 November 2012**

**Until 31 October 2013**

If it is possible that the project may continue after the end of this period, you will need to resubmit an application in time to allow the case to be considered.

UNIVERSITY OF HERTFORDSHIRE  
SCIENCE AND TECHNOLOGY

## MEMORANDUM

**TO** Ori Novanda

**CC** Kerstin Dautenhahn

**FROM** Dr Simon Trainis Science and Technology ECDA Chairman

**DATE** 24 July 2013

---

Protocol number: a1213/10

Title of study: Study to evaluate which modality of interaction is important and under which conditions in imitation learning for a humanoid robot.

Your application to extend the existing protocol detailed above has been accepted and approved by the ECDA for your school.

This approval is valid:

From: 24 July 2013

To: 31 October 2014

**Please note:**

**Approval applies specifically to the research study/methodology and timings as detailed in your Form EC1. Should you amend any aspect of your research, or wish to apply for an extension to your study, you will need your supervisor's approval and must complete and submit form EC2. In cases where the amendments to the original study are deemed to be substantial, a new Form EC1 may need to be completed prior to the study being undertaken.**

UNIVERSITY OF HERTFORDSHIRE  
SCIENCE & TECHNOLOGY

## ETHICS APPROVAL NOTIFICATION

**TO** Ori Novanda  
**CC** Professor Kerstin Dautenhahn  
**FROM** Dr Simon Trainis, Science and Technology ECDA Chairman  
**DATE** 02/02/16

---

Protocol number: COM/PGR/UH/02024

Title of study: Evaluation of human teaching activities in robot learning by demonstration.

Your application for ethics approval has been accepted and approved by the ECDA for your School.

This approval is valid:

From: 15/02/16

To: 14/02/17

**Please note:**

**Approval applies specifically to the research study/methodology and timings as detailed in your Form EC1. Should you amend any aspect of your research, or wish to apply for an extension to your study, you will need your supervisor's approval and must complete and submit form EC2. In cases where the amendments to the original study are deemed to be substantial, a new Form EC1 may need to be completed prior to the study being undertaken.**

**Should adverse circumstances arise during this study such as physical reaction/harm, mental/emotional harm, intrusion of privacy or breach of confidentiality this must be reported to the approving Committee immediately. Failure to report adverse circumstance/s would be considered misconduct.**

**Ensure you quote the UH protocol number and the name of the approving Committee on all paperwork, including recruitment advertisements/online requests, for this study.**

**Students must include this Approval Notification with their submission.**

**UNIVERSITY OF HERTFORDSHIRE  
SCIENCE AND TECHNOLOGY  
ETHICS APPROVAL NOTIFICATION**

**TO** Ori Novanda  
**CC** Prof Dr Kerstin Dautenhahn  
**FROM** Dr Simon Trainis, Science and Technology ECDA Chairman  
**DATE** 24/03/2016

---

Protocol number: **aCOM/PGR/UH/02024(1)**

Title of study: Evaluation of human teaching activities in robot learning by demonstration

Your application to modify the existing protocol COM/PGR/UH/02024 as detailed below has been accepted and approved by the ECDA for your School.

Modification: As detailed in the EC2 received on 22 March 2016:  
Addition to the consent form for using the video for scientific presentations;  
Minor rewording to one of the questions in the questionnaire;  
Addition of more questions in the questionnaire.

This approval is valid:

From: 24/03/2016

To: 14/02/2017

**Please note:**

**Any conditions relating to the original protocol approval remain and must be complied with.**

**Approval applies specifically to the research study/methodology and timings as detailed in your Form EC1 or as detailed in the EC2 request. Should you amend any further aspect of your research, or wish to apply for an extension to your study, you will need your supervisor's approval and must complete and submit a further EC2 request. In cases where the amendments to the original study are deemed to be substantial, a new Form EC1 may need to be completed prior to the study being undertaken.**

**Should adverse circumstances arise during this study such as physical reaction/harm, mental/emotional harm, intrusion of privacy or breach of confidentiality this must be reported to the approving Committee immediately. Failure to report adverse circumstance/s would be considered misconduct.**

**Ensure you quote the UH protocol number and the name of the approving Committee on all paperwork, including recruitment advertisements/online requests, for this study.**

**Students must include this Approval Notification with their submission.**

## Appendix C. Paper-based Questionnaire Forms

## Demographics

Participant Number:

1. Age:
2. Gender:
3. Do you mainly speak in English?

If you don't, what is your main language?

4. Have you had any experience playing with other robots? (please tick all that apply)  
 Another KASPAR       Sony AIBO       Lego Robot       Toy Robot  
 Other (please specify).....

# Questionnaire

Participant Number:

1. Did you fully understand what instructions KASPAR said during the main session? (1 = not very well, 5 = very well)

1	2	3	4	5
---	---	---	---	---

2. In terms of effort, how did you feel about the different methods to teach KASPAR to dance? (1 = very hard, 5 = very easy)

A. By speaking to it:

1	2	3	4	5
---	---	---	---	---

B. By demonstrating using my arms:

1	2	3	4	5
---	---	---	---	---

C. By moving the robot's arms:

1	2	3	4	5
---	---	---	---	---

D. By a combination of speaking and demonstrating:

1	2	3	4	5
---	---	---	---	---

3. In terms of enjoyment, how did you feel about the different methods to teach KASPAR to dance? (1 = least enjoyable, 5 = most enjoyable)

A. By speaking to it:

1	2	3	4	5
---	---	---	---	---

B. By demonstrating using my arms:

1	2	3	4	5
---	---	---	---	---

C. By moving the robot's arms:

1	2	3	4	5
---	---	---	---	---

D. By a combination of speaking and demonstrating:

1	2	3	4	5
---	---	---	---	---

4. When KASPAR showed what it had learned, how well did you feel KASPAR followed your instructions? (1 = not very well, 5 = very well)

A. By speaking to it:

1	2	3	4	5
---	---	---	---	---

B. By demonstrating using my arms:

1	2	3	4	5
---	---	---	---	---

C. By moving the robot's arms:

1	2	3	4	5
---	---	---	---	---

D. By a combination of speaking and demonstrating:

1	2	3	4	5
---	---	---	---	---

# Demographics

Participant Number:
---------------------

1. Age:
2. Gender:
3. Have you had any experience using other robots? (please tick all that apply)  
 Another KASPAR     NAO     Pleo     Lego Robot     Toy Robot  
 Other (please specify).....
4. Have you had any experience building a robot (including robot toy kit)? (please name the robot)
5. Have you had any experience programming a robot? (please name the robot)
6. How familiar are you with the nursery rhyme "Wind the Bobbin Up"?  
(1 = very unfamiliar, 5 = very familiar)

1	2	3	4	5
---	---	---	---	---

# Questionnaire

1. In term of effort, how did you find teaching KASPAR the gestures? (1 = very hard, 5 = very easy)

A. "Wind the bobbin up"	1	2	3	4	5
B. "Pull, pull"	1	2	3	4	5
C. "Clap, clap, clap"	1	2	3	4	5
D. "Point to the ceiling"	1	2	3	4	5
E. "Point to the floor"	1	2	3	4	5
F. "Put your hands upon your knee"	1	2	3	4	5

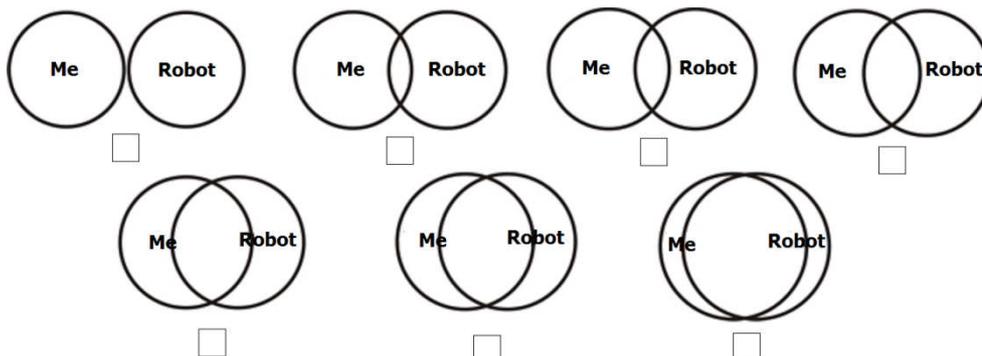
2. In term of enjoyment, how much did you enjoy teaching KASPAR the gestures? (1 = least enjoyable, 5 = most enjoyable)

A. "Wind the bobbin up"	1	2	3	4	5
B. "Pull, pull"	1	2	3	4	5
C. "Clap, clap, clap"	1	2	3	4	5
D. "Point to the ceiling"	1	2	3	4	5
E. "Point to the floor"	1	2	3	4	5
F. "Put your hands upon your knee"	1	2	3	4	5

3. When KASPAR showed you what it had learned, how well did you feel KASPAR followed your demonstration? (1 = not very well, 5 = very well)

A. "Wind the bobbin up"	1	2	3	4	5
B. "Pull, pull"	1	2	3	4	5
C. "Clap, clap, clap"	1	2	3	4	5
D. "Point to the ceiling"	1	2	3	4	5
E. "Point to the floor"	1	2	3	4	5
F. "Put your hands upon your knee"	1	2	3	4	5

4. Please tick the picture that best describes your relationship with the robot during the experiment session.



Participant number (Group A, Gesture Demonstrators):

## Questionnaire

5. Please rate your impression of the robot on these scales:

Fake	1	2	3	4	5	Natural
Machinelike	1	2	3	4	5	Humanlike
Unconscious	1	2	3	4	5	Conscious
Artificial	1	2	3	4	5	Lifelike
Moving rigidly	1	2	3	4	5	Moving elegantly
Dead	1	2	3	4	5	Alive
Stagnant	1	2	3	4	5	Lively
Mechanical	1	2	3	4	5	Organic
Artificial	1	2	3	4	5	Lifelike
Inert	1	2	3	4	5	Interactive
Apathetic	1	2	3	4	5	Responsive
Dislike	1	2	3	4	5	Like
Unfriendly	1	2	3	4	5	Friendly
Unkind	1	2	3	4	5	Kind
Unpleasant	1	2	3	4	5	Pleasant
Awful	1	2	3	4	5	Nice
Incompetent	1	2	3	4	5	Competent
Ignorant	1	2	3	4	5	Knowledgeable
Irresponsible	1	2	3	4	5	Responsible
Unintelligent	1	2	3	4	5	Intelligent
Foolish	1	2	3	4	5	Sensible

6. Please rate your emotional state during the experiment session on these scales:

Anxious	1	2	3	4	5	Relaxed
Agitated	1	2	3	4	5	Calm
Quiescent	1	2	3	4	5	Surprised