The Effect of Embodiment in Sign Language Tutoring using Robots

Joan Ressing
s4384555
15-06-2017

Supervisors

Louis Vuurpijl
Beata Grzyb

Acknowledgments

I would like to thank my supervisors and external advisors for sharing their experience and constructive criticism with me. Your availability has greatly helped me during my time working on my thesis.

Louis Vuurpijl †	For the passion that he embodied and his drive to push students to their maximum potential.
Beata Grzyb	For her constructive criticism and to-the-point attitude as well as always being available for help.
Onno Crasborn	For his uplifting spirit and the great feedback he gave on how to improve my gestures.
Henk van den Heuvel	For his enjoyable presence during our meetings and his interesting ideas.

Table of contents

Abstract

This thesis presents a 16-participant study on the effect of embodiment of a robot. The experiment was conducted in a sign language tutoring setting using two tutoring conditions: (1) physically embodied robot tutor and (2) video's of the robot tutor. All participants were tested in both conditions. The effect of the embodiment was measured in rate of learning, enjoyment, motivation, engagement and human likeness. The results suggest that the rate of learning increases when learning from a physically embodied robot as do levels of engagement, motivation and enjoyment. The results of this study strengthen the case for developing embodied agents for tutoring purposes, specifically for sign language tutoring.

Introduction

Approximately 2% of Americans suffer from a hearing disability [4], many of those people speak sign language. Instead of verbal communication, sign languages use mostly manual communication, such as hand movements and body posture, to transmit a message. Most deaf people have sign language as their mother tongue and therefore are generally fluent in sign language. However, there are also people who are less proficient in speaking sign language. Such as people who have suffered head trauma and therefore need to learn sign language, or friends who would like to learn sign language to communicate more easily. These are groups of people that could potentially benefit a lot from having robot tutoring available to them.

With the advances in robotics and the ever growing acceptance of technological means in education, it is becoming increasingly popular to employ robots as tutors. As robots gain more degrees of freedom, their motions increase in complexity. This gain in mobility lends itself well for many forms of motor behavior. On the one hand you have development of robots with robust movement and heavy work capabilities, like Atlas and BigDog by Boston Dynamics [10] and SCHAFT by Google. On the other hand there are also many developments in companion robots. This is where Aldebaran is doing a lot work with their Nao, Pepper, and Romeo robots. Companion robots are usually not designed to do heavy work, but rather to assist people in their day-to-day lives in various ways.

One of the ways in which companion robots can be used is for tutoring. There are settings in which a companion robot can already be of added value in a classroom. One could for example think of a beginners English class where children can ask for translations of specific words. However, another area in which robots are becoming increasingly relevant is sign language tutoring. Honda has been working on their ASIMO robot for the last two decades. Asimo is a companion robot that traverse all kinds of terrain including stairs [11]. Since 2014 however, Asimo can now also communicate in sign language. Toshiba has produced Aiko Chihira, a humanoid robot that can also communicate using sign language. However, not only commercial brands are interested in robots using sign language. The national university of Taiwan came out with Nino in 2013 and the technical university of Istanbul (ITU) may not have its own type of robot, but does extensive research into sign language tutoring with robots.

- ## Related work

Kose et al. have implemented an interactive game for children to learn sign language from the Nao. The Nao would tell a story and sometimes perform gestures, after each gesture, children had to guess the meaning. The guessing was done by using flashcards. Using this framework they found that young children very much enjoyed learning from the Nao [6]. This finding was backed by a study by Akalin et al. where they tested the recognition rate of gestures performed by the Nao. The second study however was performed with graduate students [1]. In one more study by the ITU Özkul et al. compared the Nao to a modified Robovie R3. They found that users often preferred the Robovie R3 to the Nao [9]. It should be noted though that the R3 had modified and was able to perform gestures better.

Leyzberg et al. performed an experiment where participants had to solve a series of puzzles where they would occasionally get advice from a tutor. There were five different tutor conditions. They found that participants who received advice from a physical robot significantly outperformed participants who received similar advice from a video representation of the same robot [7]. Interesting about this results is that participants did not perceive this increase in performance themselves. Leyzberg et al. did another experiment with a similar setup in which they found that personalizing the advice of the robot yields significantly better results compared to non-personalized advice [8].

Wainer et al. performed an experiment where participants had to do the towers of Hanoi puzzle while interacting with a robot and virtual robots. Interesting about this experiment was that they also included a condition with a remote real robot shown via a video-conference call. Their data suggests that participants preferred the present robot over a remote and virtual robots [12]. The previous results which suggest that learning with an embodied robot is very beneficial compared to a virtual robot, seem promising. Nonetheless, the effect of embodiment is not yet well established in the literature. For example, Kennedy et al. found no significant learning improvements with an embodied agent compared to a virtual agent [6].

- ## Purpose of research

There exists quite a rich body of research done on the effect of embodiment of robots. However there is no well established conclusion yet on whether the embodiment of a robot helps with learning. Even less so, how it helps with learning. With my research I want to add to the body of existing research to come closer to understanding the influence of embodiment in a tutoring setting. Knowing how embodiment affects pupils in a tutoring setting will give us insights into how important or necessary embodiment is for a humanoid robot tutor.

Many of the experiments performed so far give the subject a puzzle he or she has to solve and place the robot in an advisory role. I would like to take an approach, closer to the one used by the ITU, where the participant learns directly from the robot. I believe this approach increases the effect of the embodiment robot on the performance, compared to approaches in which the robot only fulfills an advisory role, since the robot fulfills a more central role. To study the effect of embodiment the effect on rate of learning, enjoyment, engagement and motivation was measured.

Pepper and the gestures used

- Sign language

Sign language is a natural language that primarily uses manual communication to transfer a message. There are many people for whom sign language is the main form of communication. Examples are the deaf community, people who can hear but are unable to physically speak, and their family and friends.

Signs in sign language can consist out of many components. A sign may use one or both hands, body posture, mouthings and many more features. The components of a sign can be divided into manual components (using the hands) and non-manual components. A brief overview of most of the important components can be found in table 1 below.

Component	**Description**	**Manual**
Handshape	Shape made by the hands	Yes
Location	Location of the hands	Yes
Orientation	Orientation of the hands	Yes
Posture	Body posture of the signer	No
Expression	Facial expression of the signer	No
Mouthing	Production of visual syllables with the mouth	No
Eye gaze	Gazing direction of the eyes	No

Table 1: Overview of most common components of gestures in sign language.

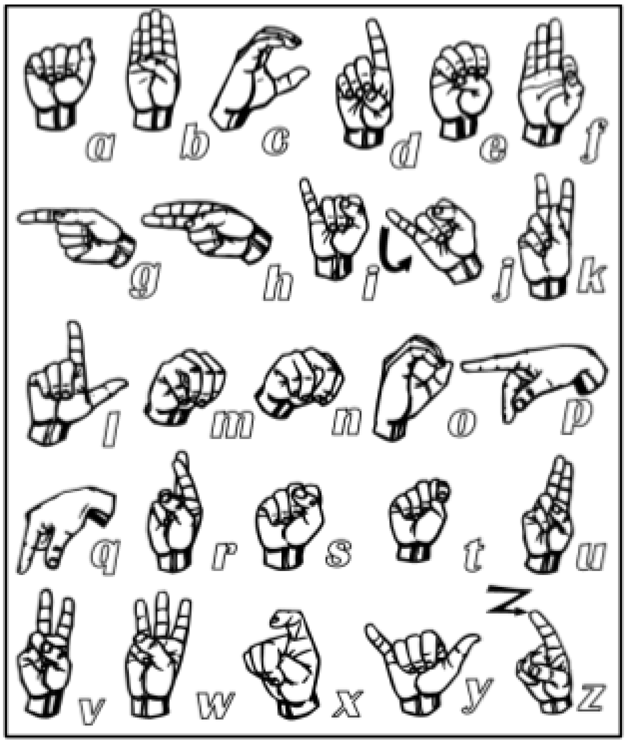

Figure 1: The SLN alphabet.

Figure 1 provides an overview of the SLN alphabet. As you can see all the the letters consist solely of manual components. Many signs in sign language use a letter from the alphabet in combination with some kind of movement or expression to create a more complex sign. This kind of behavior can for example be found in the signs for a few of the months.

- ## Gestures used

The experiment uses twenty gestures from the sign language of the Netherlands (SLN). Prior to choosing the gestures the possibilities of the Pepper were explored. To determine which gestures were appropriate I first had to learn about what the Pepper can and cannot do. Pepper has many degrees of freedom in the arms and can thus move his arms reasonably well, however, it faces plenty of limitations that will be discussed in the 'Limitations of Pepper' subsection. This subsection will discuss the gestures that were used in the experiment.

As a group we chose a set of gestures that were all in some way related to the same domain, see table 2. We chose 'time' to be our domain and used NGT Signbank [13] to search for gestures that the Pepper could perform. NGT Signbank is a database of gestures that is steadily growing and can be used to look up SLN gestures. Choosing all the gestures from a specific domain can be helpful in many scenarios, for example, if you attempt to teach gestures through storytelling. Despite me not having used such a mechanic, the set of gestures we agreed upon as a group was used in the experiment.

Set A		**Set B**	
always	july	april	monday
autumn	life	begin	september
evening	may	december	slow
fast	october	earlier	summer
februari	saturday	march	year

Table 2: The 20 gestures used in the experiment, divided into two sets of ten.

When we chose the set of gestures we expected the Pepper would be able to produce the gestures well. This was partially due to not having started on actually implementing the gestures yet. Pepper had some trouble with many of the signs. This could be attributed to two reasons.

1. Limitations of the Pepper (which will be thoroughly discussed)
2. Insufficient understanding of sign language

Upon implementing all the chosen gestures, dr. Crasborn, associate professor at the department of linguistics, was consulted for feedback. Dr. Crasborn provided important feedback on the execution of the gestures. To give an example, to an untrained eye the gesture for 'september' may appear simply like a fist making circles. Important, however, is that the fingers form the letter 's' from the alphabet while making the circles. Additionally, dr. Crasborn, gave me many pointers on the speed of the gestures, where to pause, relative speed of parts of each movement, and how to deal with the limitations of the Pepper. Insights such as these were essential for improving the overall quality of the gestures.

In table 2 the gestures are divided into two sets. The gestures were deliberately divided into two sets in order to create two equally hard sets of gestures. Having two equally hard sets meant that the effect of which gestures are learned in which tutoring condition was minimized. To reduce this effect even further, half of the participants learned set A from the embodied robot and the other half learned set B from the embodied robot. The same idea held for the learning from the video’s.

The division of the sets was carefully determined using two guidelines.

1. Find sets of gestures that look similar
2. Find gestures that are easy to remember

Sets of gestures that looked similar were split into the two different sets as fair as possible. Similarly gestures that were easy to remember, due to simple mnemonics or other reasons, were evenly distributed over the sets. The sets were also balanced using the results from a preliminary pilot study. The result section includes data that confirms the gesture sets were approximately equally difficult.

- An introduction to Pepper

For this research the Pepper robot produced by Aldebaran Robotics was used. Pepper is a humanoid robot designed to be a companion. Pepper may not be great at doing precise or heavy physical work but makes up for that with its social skills. One of Pepper’s strong points is that he is able to recognize four basic emotions; joy, sadness, anger and surprise. He is also able to interpret tone of voice and some facial expressions. Despite this being very interesting for the development of a sign language tutor, this study will not make use of this feature.

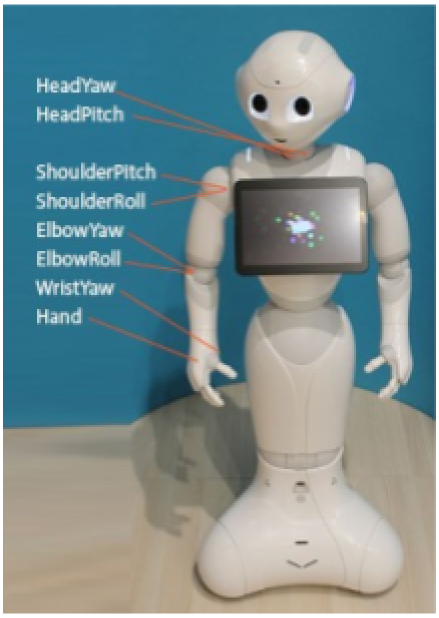

Figure 2: An overview of the relevant motors located in the Pepper.

Pepper’s body has twenty internal motors used for moving his body parts. Twelve of these motors are located in his arms which makes his arms flexible and easy manoeuvre into many positions, which is very useful in the context of sign language. A full overview of the motors located in the Pepper can be found in figure one. Despite having many motors in the arm, Pepper lacks some essential degrees of freedom for proper gesture production, especially in the hands and wrists. This is discussed extensively in the ‘Limitations of Pepper’ subsection.

The Pepper has an autonomous life mode which lets people interact with the Pepper. Enjoying a simple conversation with the Pepper is among the many options of using this mode. However, when using Pepper in an experimental setting you will likely not use this mode. To write a program or behavior that can run on the Pepper, one can make use of Choreographe. Choreographe is a software program developed by Aldebaran that can be used to write your own programs and behaviours. Choreographe allows you to immediately test your programs on the robot. Users with limited experience coding can also use Choreographe by using boxes that represent code. This means that for building simple programs, you often do not need to write any actual code. For more complicated blocks of code it is possible to make python boxes in which you can use the python code to do whatever it is you aim to do. To illustrate, most of my experiment within a python box because, among other things, wifi connections using sockets, and randomization were needed.

- ### Nao comparison

Within this research the Pepper robot was used. Another option would have been to use the Nao robot. Nao can be seen as the older brother of Pepper. Nao was released in 2006 and was the first humanoid robot produc by Aldebaran Robotics. Just like Pepper, Nao was designed to be a companion robot. Nevertheless, there are some key differences between the two. Where Pepper has three wheels used to roll around and a firm base to always stand stable, the Nao has legs and can actually walk around like humans. Since my research uses sign language the legs provide no additional functionality. On the other hand, the stability of the Pepper means you can focus on performing the gestures without having to worry about the robot falling over. This is one reason why the Pepper is more suitable than the Nao.

A more important reason as to why a Pepper is better suited for my research are the hands. The Nao has very basic hands with only three fingers on each hand. It only makes sense that for sign language a human hand is preferred over a hand that does not look like a human hand. In this aspect the Pepper robot scores a point over the Nao. The hands of both robots are depicted in figure 3.

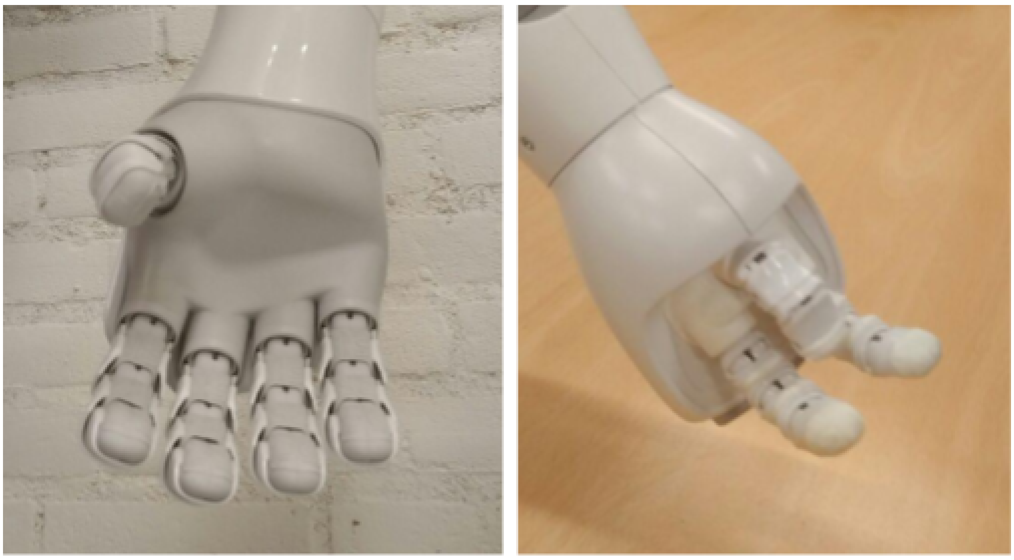

Figure 3: Side-by-side comparison of hands, Pepper (left) vs Nao (right).

Additionally, the shape of the hand deserves a mention. The back of Nao's hands is disproportionately large and the fingers come out of the palm of the hand. Consequently it is hard to see Nao's fingers under some orientations of the hand. The Pepper robot does not face this issue as his hands more closely resemble human hands. All in all the Pepper has an edge over the Nao when it comes to producing sign language. The only clear benefit that

Nao has over Pepper is that Nao does not have a tablet attached to his chest. The issues of the tablet are discussed in the next subsection.

- ## Limitations of Pepper

During my time working with the Pepper it became apparent that to make a proper sign language tutor there are many aspects of the robot that need to be improved. This subsection will be dedicated to describing the problems that were encountered while trying to implement gestures for the Pepper. In addition insights into how some of the problems could be solved are provided. The end of the section will contain conclusions on what kind of role Pepper can fulfil in further development of a robotic sign language tutor, based on my experience.

The first and likely the most significant limitation of the Pepper is that he is unable to move individual fingers. This limitation alone greatly reduces the number of gestures that can be produced by Pepper. The root of this problem is that Pepper uses a single motor per hand (see figure …) to control the position of all the fingers. As a result the Pepper can only extend or curve all fingers on a hand in unison. Many gestures, even basic ones, use individual finger movement. For example the gesture for the number "1" is produced by pointing your index finger upwards and keeping the remaining fingers curved with the palm of your hand facing towards yourself. To get a rough idea of how significantly this impacts the production of gestures we could take a look at the SLN alphabet. Out of the 26 gestures in the alphabet only 'c', 'e', and 'o' can be properly produced.

Another major problem for the execution of gestures is the tablet that is attached to the chest of Pepper. The tablet tends to form a major obstacle for types of gestures that require both hands touching each other and types of gestures where the hands need to reach the other side of the body. In the prior case the tablet forces the hands to touch at a substantially larger distance to the torso. In the latter case the tablet forces the arm to bend around the tablet making it hard to reach anything past the sternum.

A problem that amplifies the issues of the with the tablet lies in the wrist joint. Humans have a complex structure of many bones in their hands and wrist to achieve a wide range of motion. Pepper however has only one motor in the wrist as depicted in figure 4 As a result the hand can only be rotated along the yaw direction. Typical gestures where this is a problem are when the palms of the hands need to touch something, the general result is that the fingers are in the way of the palm because you cannot rotate them out of the way. Figure 4 shows Pepper signing the gesture for saturday. Here it becomes clear that the lack of wrist rotation along the yaw means that the palms of the hands cannot touch.

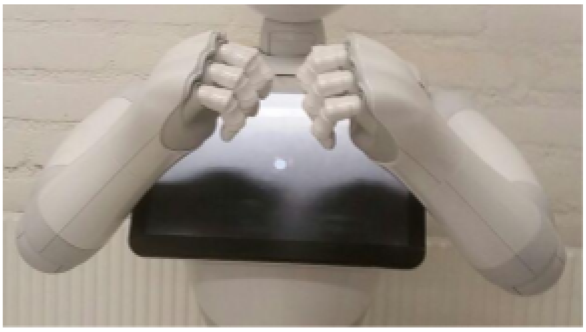

Figure 4: Pepper signing the gesture for 'saturday'.

The elbow is restricted in the angles it can make which forms an issue. The elbow joint of the Pepper can fully extend the arm, however, bending the arm is physically limited to an angle of approximately 90 degrees. When trying to bend the arm further the bending is prevented by the plastic of the under arm and upper arm colliding. Humans can achieve a way sharper angle and therefore have more range of motion. This limitation can be seen well in the gesture 'vroeger' which translates to earlier.

Then there is collision detection. Within my experiment I left it on for safety reasons as well as potential risk of losing warranty. This meant that any gesture where two parts of Pepper are supposed to touch will be different because the parts will not touch. Instead Pepper will stop the touching motion when there is about five centimeters left between both components. Of course, if you are careful you could turn off collision detection but this would generally be own risk venture.

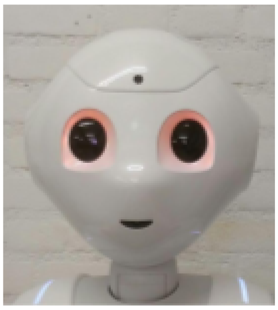

Figure 6: Pepper's head.

Last, but definitely not least, is Pepper's face. Pepper's face is very static and can only moves when Pepper moves his head. The features of the face are locked and cannot move. This is a problem because mouthings play an important role in recognizing the meaning of a gesture [2]. Pepper's face does not move nor can it make any expressions during signing which is detrimental for the quality of the gestures.

- Suitability of Pepper as a sign language tutor

This subsection reiterates the issues that Pepper faces in a sign language tutoring setting. It also discusses potential measures that could be taken to solve the problems. To be clear, the fixes discussed would not improve Pepper's general usefulness, but rather improve his sign language capabilities. The end of the section contains a conclusion on whether Pepper is suitable for sign language tutoring based on my experience working with the robot.

Two issues that were discussed that should be relatively easy to fix are the tablet and the elbows. The tablet located on Pepper's chest could probably be removed as it should not be necessary for a sign language tutor. As for the elbows, the plastic in the elbow area needs to be adjusted to allow for more range of motion. These are relatively simple fixes, that you may not be able to perform yourself, but a company like Aldebaran can.

The next two limitations of the robot are more significant, namely the degrees of freedom in the hands and in the wrists. A major issue Pepper faces is that he is unable to move individual fingers and that his wrists have only a single degree of freedom. To address this you would need at least one additional motor in each wrist and four additional motors in each hand. This would amount to a total of ten additional motors. The earlier mentioned Cognitive and Social Robotics department of the ITU has achieved something similar by replacing the hands of a Robovie R3 with custom hands which allow for additional degrees of freedom [7]. Maybe in cooperation with Aldebaran it would be possible to develop similar custom hands,

however, replacing the hand would likely also require an update on the Choreographe software to account for the additional motors.
The biggest issue that would remain are the problems with Pepper's head. Mouthings, gazing and expressions play a major role in most sign languages as they in many cases provide essential information to distinguish gestures [2]. Pepper has little to no options for any of those components. Neither is it very likely that this will change as it would require major adjustments to Pepper's face. One option would be to use Pepper's tablet to play video's of Pepper's face, however, this would not only be weird, it would also mean that the issues of the tablet remain present.

As you can see there are numerous issues with the internal and external design of the Pepper. In conclusion, Pepper, as he is produced by Aldebaran at the moment, would unfortunately not be suited to be a sign language tutor. However, this does not mean that Pepper has no place in future research on this topic. Many of the problems Pepper faces could be addressed in future versions of the robot and even with the current version there is plenty of useful research (examples are discussed in the 'Future Work' section) in which Pepper, despite his limitations, can be utilized. The completely working version of a sign language tutoring robot will likely not be a Pepper model as we have used it in this research, but Pepper may play a great role in getting to that point.

Methods

- Research questions

The main research question debates whether the embodiment of the Pepper positively influences the learning process of the participant in a sign language tutoring setting. In this question Pepper is the tutor, and the participant is the pupil. This question, however, is very broad and therefore was divided into multiple sub questions.

- RQ1: Does the embodiment of the Pepper affect the rate of learning of the gestures?
- RQ2: Does the embodiment of the Pepper motivate people?
- RQ3: Does the embodiment of the Pepper make the tutoring more humanlike?
- RQ4: Does the embodiment of the Pepper increase the enjoyment of the pupil?
- RQ5: Does the embodiment of the Pepper make the tutoring more engaging?

My hypotheses were that for sub questions two through five the answer would be yes. Which, when combined, could be the basis for an increase in the rate of learning as well, as questioned in RQ1. The results from the sub questions were used to formulate an answer to the main research question. The zero hypotheses are defined as that the embodiment of the robot has no effect on the respective variable. My expectations were that the zero hypotheses would be rejected.

Research question one was formulated to look at the effect of embodiment on the rate of learning. To address this question an experiment was set up with two tutoring conditions.

1. Physically embodied robot

2. Video's of the robot

Each subject learned from both conditions and was tested in both conditions. The scores were used as a measure for rate of learning and were used to answer the first research question. Research questions two through five were addressed using Likert scales. This was easy to do using a questionnaire that was filled out after the experiment.

- Research design

Within my research I have employed a quantitative approach. A little qualitative data was also gathered but this was only used for feedback. The experiment employs a within-subject design, this design was chosen for multiple reasons. The main reason advantage of using a within subject design is that every subject is in both groups this has a few nice benefits. Firstly this reduces the number of subjects needed for my experiment, and secondly this reduces the amount of variance between persons that naturally arises when each person is in only one condition.

A risk of using a within subject design are inherent effects that one needs to account for, namely fatigue and practice (also known as learning effects). In order to reduce learning effects the order in which the tutors were used was counterbalanced. Therefore half of the participants used the embodied robot first and the other half started with the video tutoring. To minimize any effects from fatigue the experiment was designed to be relatively short. The final experiment lasted about 20 to 25 minutes. Any fatigue effects that were still present were likely neglected by the aforementioned counterbalancing of tutoring order.

- Experiment

After participants were seated comfortably they were handed an explanation of the experiment. The explanation contained information about the goal of the study, the procedure and data collection. After having read the explanation they were asked to sign a consent form. The experiment consisted out of two major phases. Within each major phase the participant was subject to one of the two tutoring conditions. In both tutoring conditions the participant would first learn from the tutor, then have a short break followed by a test by the tutor. The learning consisted of the tutor showing each gesture once and saying the meaning of the gesture. During this part there was no interaction and the participant was only required to watch and listen. If the participant was unable to understand the tutor for any reason, the meaning was also displayed on the GUI (figure 7). After the tutor had performed each gesture, the participant was given a one minute break. This break was meant for the participant to relax and mentally process the learned gestures. During this time the

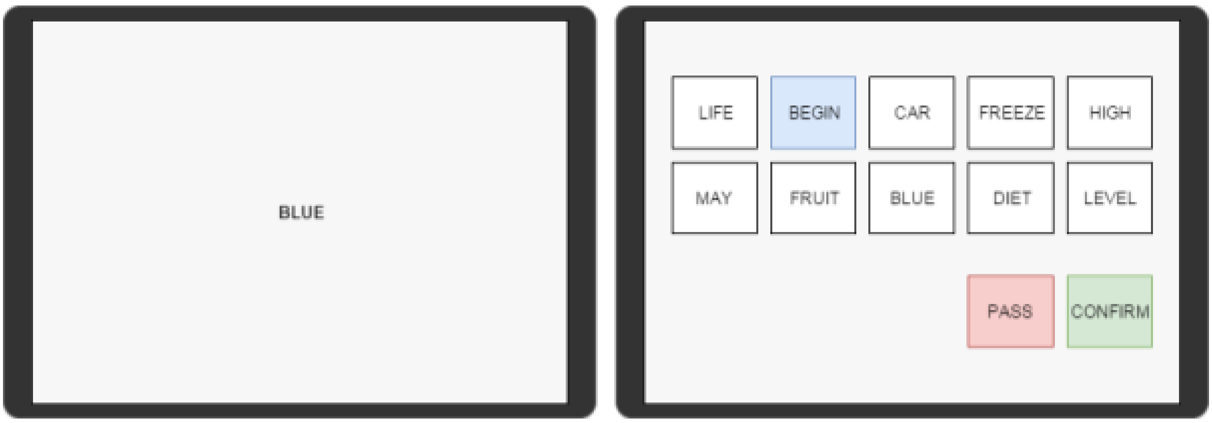

Figure 7: Shows the display of the laptop during the learn phase (left) and during the test phase (right).

participant was free to do whatever he or she wanted. Following the break the participant was quizzed by the tutor. In the phase the tutor showed each sign three times. After each sign the tutor waited until the participant had entered an answer using the laptop. The moment all signs had been asked three times the test phase was over and the tutoring condition had ended. At the end each participant filled in a questionnaire about their experience with the experiment.

- Considerations

The previous subsection described the procedure of the experiment. This subsection contains information about why the experiment was set up as such.

For the data to be most useful it was crucial that the average scores were somewhere in the range of 40-70% of questions correctly answered. In order to accomplish this a few measures were taken. To start off, the twenty gestures had to be carefully divided into two groups of gestures of similar difficulty. If the groups were too far apart in terms of difficulty this would create large variance in the data as well as increase the potential for people scoring 100% or 0% which is very undesired. If people would score 0% or 100% that would mean that the accuracy score possibly does not represent the knowledge of the participant. In that case it is possible that the participant could have scored lower/higher if it were possible. How the sets were balanced was described in the ‘Gestures used’ subsection.

Not only was balancing the gesture sets essential to keep scores close, the difficulty of the test was another major focus point. A pilot study found that showing each gesture more than once resulted in too high accuracy scores, however, showing them once resulted in average scores around 70%. Just within the target margin.

Eye-gazes play an important role in sign language and interaction with humanoid avatars [2][3]. However, for this study the decision was made not to include them for practical reasons. Eye-gazes would be hard to implement using the pepper as the only option for this is to rotate the head as a whole. Eye contact was not used either as it would be impossible to properly employ this in the video tutoring condition.

A general disadvantage of within-subject designs are learning effects and fatigue effects. Fatigue effects were prevented by keeping the experiment relatively short. Learning effects could not prevented (and were definitely present, see results section), however to counteract the effect of learning the order in which tutors were used by participants was counterbalanced.

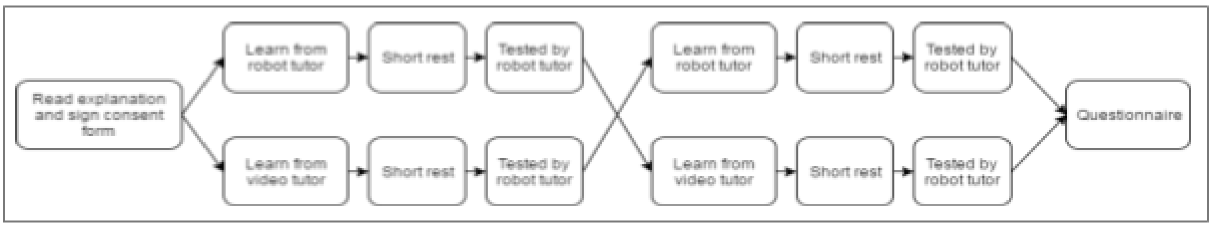

Figure 8: Schedule of the experiment.

A consideration that was made was whether to use a large flatscreen television (60") or a computer (23,8") for the testing with the video's. To purely measure the effect of embodiment you would like to have the robot occupy the same amount of the participant's visual field in both conditions. This can be achieved using a large TV and sitting a bit closer. However a case could be made for using a pc monitor as well because that is a more realistic setting for a virtual tutor. The decision to use the large tv instead of a pc monitor was made on grounds of theoretical accuracy. To ensure that the virtual robot and the embodied robot were perceived as the same size participants were seated closer in the virtual robot tutoring setting.

As mentioned earlier, the test phase of each tutoring condition consisted out of each gesture being shown three times and the participant trying to guess the meaning. The order in which gestures were shown was pseudorandom with two constraints.

1. The last gesture in a series could not be the first gesture in the next series.
2. Similar gestures could not be shown in direct succession.

Showing each gesture three times, instead of only once, was a deliberate choice to increase the precision of the scores. Additionally, asking each gesture multiple times generated additional data that could be used to look into which gestures were harder and or confusing. 'Hardness' of the gestures could be used as a measure in future work.

- Setting and materials

The experimental setup can be seen in figure 9. All of the experiments were conducted in the robotics lab of the AI department. In both conditions the participant was seated in a chair behind a desk. On the desk was a laptop with a standard usb mouse. The desk was placed facing a wall and allowed for unobstructed vision of the two tutors, Pepper and the television. During the tutoring with the robot, the person was seated at approximately three meters from the wall. During the tutoring with the video's (on the television) the person was seated at approximately two meters from the wall. The difference in distance was to account for display size. Despite using a large 60" display, the image of the Pepper was still significantly smaller than the actual robot. To prevent any bias in the data, due to one condition being perhaps easier to see, the participant was seated at a larger distance when using the robot. As a result the observed size of the robot was roughly the same.

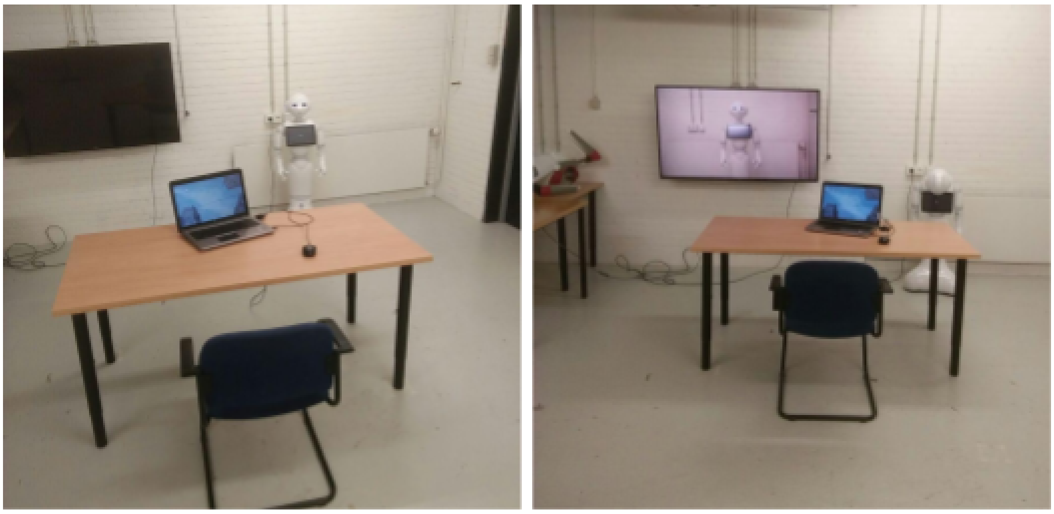

Figure 9: Setting of both test conditions. Note that in the video condition (right), the table stands further forward such that the image of Pepper closely approximates the size of the real robot.

- Hardware and software interaction

The setting of the experiment can be seen in figure 9, however, it may not be clear how all hardware components interacted during the experiment. Figure 10 contains an overview of all components involved in the experiment. In the condition in which the participant learned from the video's of the robot the television acted merely as a display. The television was connected via an hdmi cable to a laptop which was used to display all the video's on the television. The laptop was also connected to another laptop (labeled 'GUI') that was used by the participants. The laptop used by the participants had a GUI which was used for entering answers. These answers were automatically written to a local datafile. Whenever the participants entered an answer the GUI would also automatically send a message to the other laptop to notify the laptop that the next video could be played. The condition in which the participants were tutored by the physically embodied robot worked in a slightly different way. Pepper can create a wifi connection with the laptop used by the participants. This meant that the other laptop only had to send a program to the Pepper that would create a wifi connection with the laptop used by the participants and then all the communication could be handled between the robot and GUI.

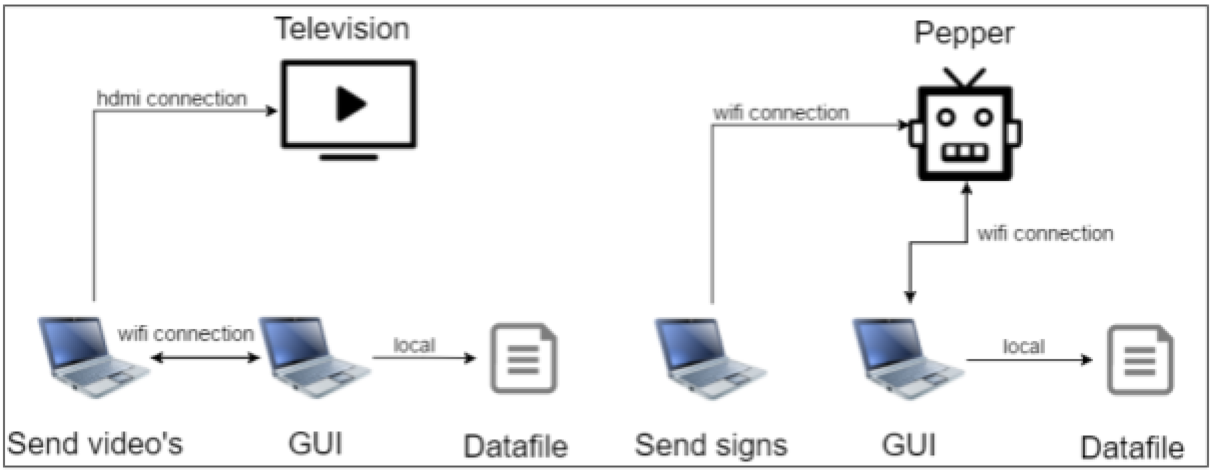

Figure 10: Overview of the interaction between hardware components. Participants only interacted with the laptop labeled 'GUI'.

All gestures were implemented using the animation mode in Choreographe. The animation mode allows the programmer to physically set the robot into a position and then store the settings of all motors in a keyframe. Next the timeline feature was used to chain multiple keyframes together. The timeline feature will automatically interpolate the movement in between gestures to create a fluent movement. Most gestures needed between five and fifteen keyframes to reasonably approximate the real gesture as performed by a human. Where using more keyframes can sometimes improve the resemblance with the real gesture this sometimes has a negative effect resulting more shaky or stuttering motion.

The video's of the gestures were made using the back camera of a Oneplus One mobile phone. All videos were shot in 1920x1080 resolution from a frontal perspective. Figure 11: shows depicts a screenshot from one of the video's.

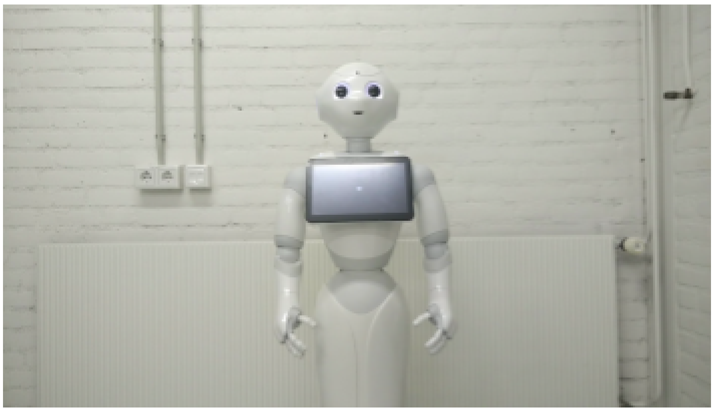

Figure 11: Screenshot from video's used in video condition.

The program running on the participant's laptop was written in Java and used the Swing library to implement the GUI. A ServerSocket was used to set up the wifi connection. A BufferedWriter was used for writing data to a comma-separated-values file.

In the video tutoring setting JavaFX was used to write an application able of playing video's and wifi communication. The embodied robot condition used Choreographe to send a program to the robot that would set up a connection and perform all the gestures.

The source code of the programs written using Python, Java and JavaFX can be accessed on my github [14].

- Subject constraints

There were three constraints on which subjects were filtered. Subject were required to:

1. read and understand english on a high school level.
2. have no prior knowledge of sign language.
3. be over 18 years old and under 30 years young.

A few participants afterwards noted that they knew the gesture for ‘slow’, this seems logical as it is a broadly used gesture even for normally speaking people. The analysis of the data was not adjusted as a result of this.

- Data collection and analysis

Data collection is accomplished in two ways. Once during the main experiment and once with the questionnaire. At the start of each tutoring phase a data file is created. Both the robot and the user write automatically to this file. This is an interactive process where the Pepper and the user take turns writing date to the file. During the test phase Pepper will show many gestures. Each time, just before performing the gesture, Pepper writes the meaning of the gesture to the data file and then performs the gesture. As the user watches Pepper perform the gesture he or she will decide on what the meaning of the gesture is. When the user has decided, he or she selects the meaning on the graphical user interface (GUI) displayed on the laptop. After pressing confirm the selected meaning is automatically written to the data file. When all gestures are performed the GUI closes and no more data are written to the file. The rest of the data are collected through a questionnaire. The questionnaire contains ten 5-point Likert scales and two open questions. The questionnaire is filled out at the end of the experiment.

There were two forms of data; the data from the experiment and the data from the questionnaire. This data was analyzed separately. First off, the data from the experiment. The data from the experiment contains two files per person, each file represents a different condition, either learning from the robot, or learning from the videos. Each files contain thirty pairs of data, where each pair contains:

1. Name of particular gesture performed during test phase.
2. Respective answer of the participant.

The data was analyzed using a repeated measures ANOVA test using Tutor (embodied robot vs virtual robot) as within-subject variable. Two between-subject variables were taken into account as well. Namely ‘SetR’ indicating which set of gestures was used in the embodied robot condition (remember that the gestures were divided into two sets), and ‘First’ indicating which tutoring condition was used first. Below is an overview of the variables analyzed and the effects that they measured.

The second form of data comes from the questionnaires. The questionnaire contains mostly likert scales and in addition also features two open questions. The open questions were used primarily to check whether there were any influences or effects in the experiment that I had not accounted for. This required no statistical analysis. The results of the Likert scales were analyzed by comparing them pairwise wherever relevant. An example of a pair of likert scales that were compared would be:

1. *The robot tutoring felt humanlike.* (strongly disagree - strongly agree)
2. *The video tutoring felt humanlike.* (strongly disagree - strongly agree)

To analyse the Likert scales the answers were translated from “strongly disagree - strongly agree”, to numbers one through five. Fischer’s exact test was used to test for significant differences. P-values below .05 were considered significant. The analysis of standalone Likert scales such as *“I would consider using a robot tutor in the future.”* was done by computing the means and standard deviations.

Results

- Description of subjects

The experiment was conducted with sixteen participants within the range of 20 to 25 years. Out of the participant eleven were male and five were female. The participants were mainly students with a few exceptions. Some of the students were AI students and thus had previous experience with robots. Prior to the experiment none of the participants had any significant knowledge of sign language. After the explanation of the experiment all of the participants signed a consent form, agreeing that they were properly informed about the experiment.

- Experiment results

The main experiment consisted out of two parts. In one part the participant was tutored and subsequently questioned by the embodied Pepper. In the other part the participant was tutored and subsequently questioned by video’s of the video’s of Pepper. Before doing the analysis the data was confirmed to be approximately normally distributed using a Shapiro-Wilk’s test ($p > .05$) and visual inspection of their histograms and Q-Q plots.

Group	**Mean**	**Standard deviation**	**Sample size**
Robot	72.71	18.35	16
Video	66.88	22.66	16
Set A	71.67	21.98	16
Set B	67.92	19.43	16
First condition	63.75	21.94	16
Second condition	75.83	17.58	16

Table 3: Means and standard deviations of experiment groups.

Table 3 describes the means and standard deviations of the accumulated data. It is clear that there is a strong learning effect and a smaller effect of tutor. The two sets of gestures were decently balanced in terms of difficulty. A repeated measures ANOVA was computed using SPSS 20 to analyze the accuracy score data. Tables 4 and 5 feature the results of the analysis.

	First	SetR	Mean	Std. Deviation	N
Robot	Robot	A	55.8350	16.18836	4
		B	90.0000	8.16497	4
		Total	72.9175	21.78034	8
	Video	A	77.5000	9.17888	4
		B	67.5025	20.61647	4
		Total	72.5013	15.71065	8
	Total	A	66.6675	16.80854	8
		B	78.7513	18.85051	8
		Total	72.7094	18.34693	16
Video	Robot	A	63.3325	14.14292	4
		B	95.0000	6.38169	4
		Total	79.1663	19.74084	8
	Video	A	50.8350	11.34591	4
		B	58.3325	26.17470	4
		Total	54.5838	19.10108	8
	Total	A	57.0837	13.62052	8
		B	76.6662	26.36701	8
		Total	66.8750	22.65548	16

Table 4: Descriptive statistics of the experiment data.

Effect	**Significance**	**Effect size**	**Observed power**
Tutor	.023	.362	.668
Tutor x First	.000	.708	.999
Tutor x SetR	.120	.190	.338
Tutor x First x SetR	.045	.294	.537

Table 5: Significance levels and effect size of effects. Computed using alpha = .05.

Figure 5.2 illustrates an interesting interaction effect between the variables tutor, first, and the set used in the embodied robot setting. When the embodied robot was used in the first condition participants scored slightly higher in the video condition (due to learning effects) and scores were significantly higher for set B. However, when the video’s were used in the first condition the participants scored far better with the embodied robot (partially due to learning effects). Interesting is the notable interaction between tutor and set.

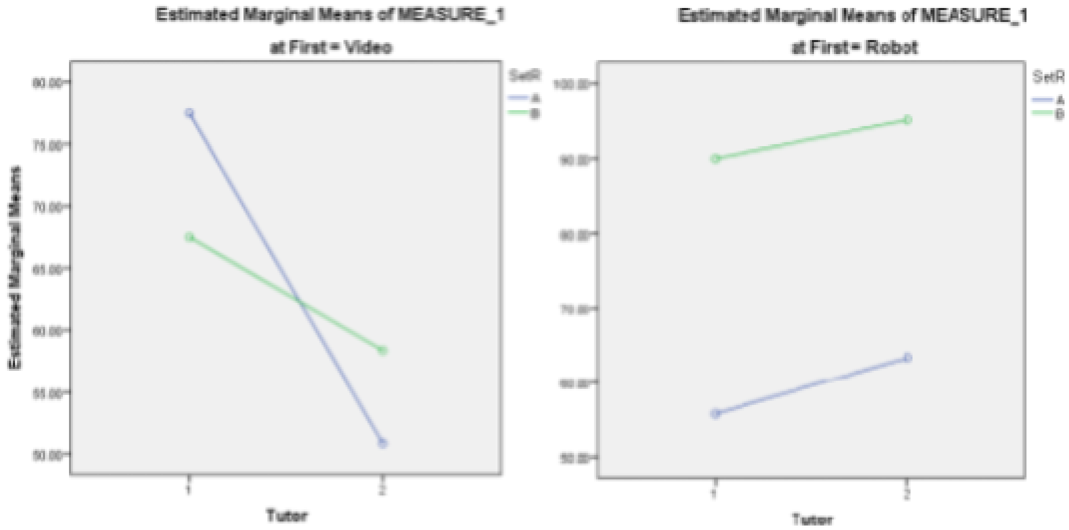

Figure 5.2: Displays the interaction effect noticed between the variables (Tutor x SetR x First). Tutor 1=robot, tutor 2=video.

The experiment data was also used to generate the confusion matrices displayed in table 6.

	always	autumn	evening	fast	februari	july	life	may	october	saturday
always	37	0	0	0	0	3	1	2	4	1
autumn	0	47	0	0	0	0	1	0	0	0
evening	1	1	38	0	0	0	5	0	2	0
fast	1	0	0	40	0	0	0	4	2	0
februari	0	1	2	1	32	1	2	1	3	3
july	6	0	2	3	2	24	0	2	4	3
life	4	1	0	0	1	0	37	0	3	2
may	0	0	3	4	1	1	0	28	5	3
october	4	2	3	2	4	1	1	2	26	2
saturday	1	0	1	1	3	0	2	2	2	35

	april	begin	december	earlier	march	monday	september	slow	summer	year
april	29	4	0	4	0	1	0	0	4	3
begin	1	31	0	0	8	0	3	2	0	3
december	0	1	45	0	0	0	0	0	2	0
earlier	0	1	0	41	1	0	0	0	3	2
march	2	4	0	0	28	1	1	1	1	10
monday	2	1	1	0	6	25	10	0	0	2
september	0	2	0	0	6	19	16	0	0	5
slow	0	0	0	0	0	0	0	48	0	0
summer	0	2	0	5	0	0	0	0	39	2
year	2	4	1	1	8	1	1	0	3	24

Table 6: Confusion matrices for both gesture sets. Rows indicate the answers that were given for a specific gesture.

- Questionnaire results

The questionnaire features ten Likert scales and two open questions. The Likert scales can be divided into four pairs and two standalone scales. All the Likert scales ranged from ‘Strongly disagree’ (1) to ‘Strongly agree’ (5). The sample size for the Likert scales varies slightly for two reasons. The first reason is that one person forgot to fill out the back side of the questionnaire. The other reason is another person had an older version where the question about how humanlike the tutoring was was not included. Let us start by taking a look at the results of the standalone Likert scales.

- *The robot felt more engaging than the video's of the robot.*

Mean	4
Standard deviation	0.845
Sample size	15

Table 7

- *I would consider using a robot as a tutor in the future.*

Mean	3.267
Standard deviation	0.961
Sample size	15

Table 8

The remaining Likert scales were analyzed pairwise using the R programming language to perform Fisher's exact tests.

Statements	p-value from Fisher's exact test
I could understand the Pepper robot well. I could understand the Pepper in the video's well.	0.8592
I enjoyed learning from the Pepper robot. I enjoyed learning from the video's of the Pepper.	0.006
I felt motivated to learn from the robot. I felt motivated to learn from the video's of the Pepper.	0.046
The physical robot tutoring felt humanlike. The video tutoring felt humanlike.	0.027

Table 9: p-values from pairwise analysis of Likert scales.

The questionnaire also contained two open questions, namely:

- *Were there any factors that affected your performance in either condition?*
- *Any additional remarks or suggestions?*

These questions however were written with the intent of finding discovering flaws or missed points of interest in the research design. As a result the feedback provided was not, and could not be, used to answer the research questions. The only interesting feedback came from a person who would have liked to have feedback on his answers during the two test phases.

Discussion

- Conclusion

Table 3 shows that the two gesture sets were balanced very well. A difference in mean scores of only 3.75% indicates that the performance of both sets was nearly the same. Given that the sets were counterbalanced they should have the same effect on the performance in either condition. The order in which tutors were presented however had a very significant effect on performance of participants. An average of 63.75% in the first condition and an average of 75.83% in the second condition shows that there was a large learning effect present, however, because of counterbalancing this effect was greatly reduced. The average scores using the embodied robot were 72.71% compared to 66.88 in the virtual robot condition. Whether this is enough to be a significant effect will be discussed below.

RQ1: Does the embodiment of the Pepper affect the rate of learning of the gestures?
The results suggest that the participants learned significantly faster from an embodied robot compared to the virtual robot ($p = .023$). The zero hypothesis, stating that the embodiment would have no effect on the rate of learning, was therefore rejected. This results could be explained by the gestures being easier to see because the gestures are observed in a three dimensional space. Another possible explanation would be that the participants were excited to work with the Pepper as for most this was a new experience. To exclude such a novelty effect a longitudinal study should be conducted where participants come back for multiple learning sessions.

Interesting is that there were also significant interaction effects between (Tutor x First, $p=.000$) and (Tutor x First x SetR, $p=0.45$). The interaction effect between the tutoring condition and which tutoring condition was used first was the result of people scoring surprisingly low in the video condition when starting in said condition (see figure 5.2). A reasonable explanation for this effect is hard to justify.

RQ2: Does the embodiment of the Pepper motivate people?
A p-value of .046 was found for the effect of embodiment on motivation using a Fisher's exact test. As such the zero hypothesis is rejected and it is concluded that there was a significant difference in motivation. Looking at the average scores of the questionnaire tells us the direction of the effect, which leads to the conclusion that participants were significantly more motivated to learn from an embodied robot than for a virtual robot.

RQ3: Does the embodiment of the Pepper make the tutoring more humanlike?
Participants generally did not perceive the embodied robot as being very humanlike (mean: 2.867 ≈ very slightly disagree), however, the video's were perceived significantly ($p = .027$) less humanlike (mean: 2.133 ≈ disagree). It makes sense that an embodied robot is perceived as more humanlike and the data reflects this intuition. Perhaps the tutoring was not experienced as more humanlike because there was little interaction with the pupil.

Another factor could be that the robot did not make any eye contact with the human which is suggested to play an important role in interactions with humanoid avatars [5].

RQ4: Does the embodiment of the Pepper increase the enjoyment of the pupil?
There were very significant results for this question. With a p-value of .006 the zero hypothesis may easily be rejected. Looking at the questionnaire results the direction of the effect was determined. In conclusion, the embodiment of Pepper increases the level of enjoyment experienced by participants.

RQ5: Does the embodiment of the Pepper make the tutoring more engaging?
This research question was tested with a single Likert scale. The mean was 4 with a standard deviation of .845. With eight out of fifteen people agreeing and four out of fifteen strongly agreeing this question the data appears to suggest that this statements holds. This

Now that the sub questions have been discussed let's take a look at the main research question. "*Does the embodiment of the Pepper positively influence the learning process of the participant in a sign language tutoring setting?*" All in all the data from the experiment and from the questionnaire combined seems to suggest that the embodiment of the Pepper did have a positive influence on the participants. The rate of learning, enjoyment, motivation and human likeness all significantly improved. Whether this is a result of novelty effects needs to investigated in a longitudinal study.

In addition to the questions already discussed there was also a question in the questionnaire about whether people would consider using a robot as tutor in the future. Despite the significant differences in engagement and enjoyment compared to a virtual robot people did not seem very eager to use a robot as a tutor. With a mean score of 3.267 and a standard deviation of .961 the average user is relatively neutral in response to the question. Of course this could be for many reasons such as the anticipated prices could being out of budget, the robots perhaps being intimidating to some, or maybe their initial impression with the robot tutoring just were not exciting enough.

The confusion matrices give nice insight into the difficulty of the gestures. From table 6 we can see for example that people had no problem with the gestures 'december', 'slow' and 'autumn'. This is logical as these gestures have very simple mnemonics. The gesture for autumn resembles falling leaves, december's gesture looks like a christmas, and the gesture for slow is a universally used gesture. Table 6 can also be used to see which gestures were confused a lot, like (september, monday) and (year, march). Additionally, gestures that were just hard, and often guessed by participants can be found, like 'october'. This information could be valuable for potential follow-up studies using the same gestures.

- Future work

This research only scratches the surface of investigating the effect of embodiment of robots. Not only did it only test using the Pepper robot, it also only tested in a robot tutoring setting. Even in a robot tutoring setting there are many more experiments that could be done. For those reasons it is hard to generalize the found results to broader fields. This subsection will list a few of the possible directions for future research.

- Why does embodiment increase performance? Is it by the gestures being perceived better visually or do engagement, motivation and enjoyment also play a significant role? The experiment performed featured little interaction with the participant. Could the effects be amplified by additional interaction with the participant?

- A longitudinal study on the effects of embodiment would be very interesting to see whether they last or wear out after using embodied robots over a longer period of time. The code from this thesis could largely be used to set up such an experiment.

- Does the appearance of the robot affect the effect of embodiment? Would the effect of embodiment using a Aiko Chihira be larger than when using a Pepper? Or are there differences already between using a Nao and a Pepper? It is possible that some robots have larger or smaller effects of embodiments based on their appearance. Are larger or smaller effects observed when using robots associated with the uncanny valley?

- The effects of embodiment are maybe expected when tutoring a visual subject such as sign language, however, what happens when tutoring spoken language? Do the effects persist or does the embodiment lose it’s purpose?

- All subjects that participated in the experiment were in Dutch people in the age range of 20-25, of which over half studied AI, therefore the conclusions cannot be drawn for the whole population. This opens up options for follow research using a wider variety of participants such as elderly. Do older people also prefer embodied robots over virtual robots?

- Could robots in theory be better tutors than humans? When social behaviour for robots is optimized and motor behavior comes close to perfection, could robots perform equally or even outperform average human tutors?

- This study investigates the effect of embodiment of the Pepper robot. However to be precise, it investigates the effect of physical 3d embodiment. What would the effects be when compared to non-physical 3d embodiment such as holograms? Would the feeling of the robot being real have an influence on factors such as engagement, enjoyment or human likeness?

- As discussed in the discussion there seems to be an interaction effect between the tutor embodiment and the order in which tutors were presented. This makes it interesting to do a follow-up study in which the effects order of presentation are further investigated.

References

1. Akalin, N., Uluer, P., Kose, H., & Ince, G. (2013, November). Humanoid robots communication with participants using sign language: An interaction based sign language game. In *Advanced Robotics and its Social Impacts (ARSO), 2013 IEEE Workshop on* (pp. 181-186). IEEE.

2. Fontana, S. (2008). Mouth actions as gesture in sign language. *Gesture*, *8*(1), 104-123.

3. Garau, M., Slater, M., Bee, S., & Sasse, M. A. (2001, March). The impact of eye gaze on communication using humanoid avatars. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 309-316). ACM.

4. Holt, J. A., Hotto, S., & Cole, K. (1994). *Demographic aspects of hearing impairment: Questions and answers*. Center for Assessment & Demographic Studies, Gallaudet University.

5. Kennedy, J., Baxter, P., & Belpaeme, T. (2015). Comparing robot embodiments in a guided discovery learning interaction with children. *International Journal of Social Robotics*, *7*(2), 293-308.

6. Kose, H., & Yorganci, R. (2011, October). Tale of a robot: humanoid robot assisted sign language tutoring. In *Humanoid Robots (Humanoids), 2011 11th IEEE-RAS International Conference on* (pp. 105-111). IEEE.

7. Leyzberg, D., Spaulding, S., Toneva, M., & Scassellati, B. (2012). The physical presence of a robot tutor increases cognitive learning gains.

8. Leyzberg, D., Spaulding, S., & Scassellati, B. (2014, March). Personalizing robot tutors to individuals' learning differences. In *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction* (pp. 423-430). ACM.

9. Özkul, A., Köse, H., Yorganci, R., & Ince, G. (2014, December). Robostar: an interaction game with humanoid robots for learning sign language. In *Robotics and Biomimetics (ROBIO), 2014 IEEE International Conference on* (pp. 522-527). IEEE.

10. Raibert, M., Blankespoor, K., Nelson, G., & Playter, R. (2008). Bigdog, the rough-terrain quadruped robot. *IFAC Proceedings Volumes*, *41*(2), 10822-10825.

11. Sakagami, Y., Watanabe, R., Aoyama, C., Matsunaga, S., Higaki, N., & Fujimura, K. (2002). The intelligent ASIMO: System overview and integration. In *Intelligent Robots and Systems, 2002. IEEE/RSJ International Conference on* (Vol. 3, pp. 2478-2483).

IEEE.

12. Wainer, J., Feil-Seifer, D. J., Shell, D. A., & Mataric, M. J. (2006, September). The role of physical embodiment in human-robot interaction. In *Robot and Human Interactive Communication, 2006. ROMAN 2006. The 15th IEEE International Symposium on* (pp. 117-122). IEEE.

Additional resources

13. NGT Signbank, https://signbank.science.ru.nl/

14. https://github.com/JoanRessing/SignLanguageTutor