6-2017

Attracting Human Attention Using Robotic Facial Expressions and Gestures

Venus Yu
Union College - Schenectady, NY

Attracting Human Attention Using Robotic Facial Expressions and Gestures

By

Venus Yu

* * * * * * * * *

Submitted in partial fulfillment

of the requirements for

Honors in the Department of Computer Science

UNION COLLEGE

March 16, 2017

Abstract

YU, VENUS   Attracting Human Attention Using Robotic Facial Expressions and Gestures. Department of Computer Science, March 16, 2017.

ADVISOR: Nick Webb

Robots will soon interact with humans in settings outside of a lab. Since it will be likely that their bodies will not be as developed as their programming, they will not have the complex limbs needed to perform simple tasks. Thus they will need to seek human assistance by asking them for help appropriately. But how will these robots know how to act? This research will focus on the specific nonverbal behaviors a robot could use to attract someone's attention and convince them to interact with the robot. In particular, it will need the correct facial expressions and gestures to convince people to help them.

Contents

List of Figures

1 Introduction

As more advances are made to robots, we will increasingly see them out in public spaces. They will ideally be carrying out tasks and assisting humans with their everyday needs. In order for these social robots, which should interact with humans in a socially acceptable way, to communicate with their human conversation partner, they will have to speak and act as a human would in the same conversation. Otherwise their human partner may not feel comfortable speaking with the robot or helping it with any tasks it might need to fulfill. It may also be likely that the robot will not have a body capable of performing simple tasks, such as pushing an elevator button or picking up an item from the ground, so it will need to rely on human assistance to complete them. For this research, our robot will not have the luxury of arms or complex body movements, so she will have to rely on her speech and gestures to effectively communicate that she needs help to humans in the vicinity.

As humans, we ask for assistance by using a combination of verbal (”can you please help me?”) and nonverbal (a hand wave, a nod, etc.) communications. Nonverbal cues, as the name suggests, do not rely on the spoken dialogue to convey a certain emotion. Though our robot will be able to ”speak” through the text on her screen and synthesized voice, we know that gestures and other such nonverbal behaviors contribute to a conversation [4]. The gestures for this particular experiment will be very simple, involving mimicking eye contact, but it is still important to mimic them properly and eventually implement them into the robot to get a richer interaction.

The robot will need to be able to act as a human to get help, so as to not scare the humans it asks. It will have to use its facial expressions and gestures to express a desire to be helped, or else it will not get the help it needs. But how can a robot, with limited power and parts, possibly convey the same emotions as a human if it cannot fully mimic them? The beauty of robots is that they are not required to be the same as humans when they communicate. This is partly due to the difficulty of making a robot as life-like as possible, but it is also a product of people expecting a robot to not be able to mimic human mannerisms perfectly because they know that it is difficult. If the robots are close to acting like a human or use mannerisms specific to their robotic nature, the same emotional message can be conveyed. Walters et al. [12] showed that the more a robot acts like its appearance suggests (i.e. a more robot-looking robot speaking with a mechanical voice and beeps versus a human-looking robot that speaks with a human-like voice and does not have flashing lights), the more comfortable the human is when communicating with the robot. If there is dissonance between the robot appearance and its behaviors, the human will find it strange and off-putting. It would be strange to see, for example, a robot that was in the shape of a rectangle that could speak and act perfectly human, despite its lack of an actual body. Likewise, it would be off-putting to see a humanoid robot only speak in noises and have limited limb movements, despite its complex limb structures.

Our robot will be attempting to attract people's attention such that they feel inclined to approach her and help her with a task through the verbal and nonverbal behaviors described above. Thus for the purposes of this research, attracting attention extends further than just an acknowledgement and includes the interaction that follows. As mentioned before, the robot will perform a simple gesture in which she turns to face her screen at the people passing by, mimicking eye contact. Will using this gesture and a greeting convince more people to interact with her than if she stayed stationary and greeted people? Though humans naturally turn towards something that moves, we want to know if using this gesture will help our robot get more help.

2 Motivation and Related Works

Nonverbal behaviors such as facial expressions and gestures work with verbal dialogue to enrich the interaction between humans. Facial expressions will be dependent on how the face is rendered. For this we will be using the robot_face ROS (robot operating system) modules[1] created by the team at University of Koblenz and Landau. It features a simple face with two eyes and a mouth (figure 1). According to research done by Bennett et al. [1], a minimalist face such as the one described above was enough to emote such that a human was able to recognize and identify the emotion that was displayed on the robot's screen. This is especially useful because the more components the robot face has, the harder it will be to control its facial expressions; this type of facial design is simple yet effective. Prajapati et al. [10] showed similarly that having a simple robot face mimic human expression was accurate enough that emotion reading software they used in their experiments were able to detect the emotion on the robot's face. For this experiment I only need our robot to express happiness and sadness. Though expressing these two emotions is not the focus of the experiment, because we want our robot to act human, she needs to be "happy" (expressed using a smile) when she is thanking the participant for helping her and "sad" (expressed with a frown) if they decide to not help her.

Gaze, or rather eye contact, also plays an important role in establishing initial contact because it signals a desire to engage with the other person. Heenan et al. [8] found that eye contact was highly effective in simulating social behavior since people felt that the robot was responding to their presence. To prevent discomfort due to the robots prolonged gaze, they would have the robot avert its gaze from time to time, simulating what most people do when gazing at the other party for too long. Cary [2] studied the role of gaze in the initiation of a conversation between two unacquainted individuals and found that not only was

[1] `https://gitlab.uni-koblenz.de/robbie/homer_robot_face`

Figure 1: SARAH's face.

gaze an important indicator of a willingness to engage in a conversation, but also a lack of gaze correlated to the individuals not engaging with each other. That is why one of the greetings will involve the robot turning and following a person as they greet them; it should mimic eye contact and show that robot is interested in the person's help. If she were to not follow them, they might not understand that she is addressing the greeting at them.

Unfortunately, the need for clear eye contact may be a problem due to the nature of our robot. Delaunay et al. [6] conducted a study where they projected a face, either human or animated, onto different surfaces, both 2D and 3D. They found that the human gaze was easily discernible by the participants when projected on any surface, but the animated gaze had varied results. Surprisingly, when the animated face was projected on a 3D surface, its gaze was identified better. They hypothesized that since the animated face was missing many subtle human face features, it was harder to identify on a 2D surface, but the 3D surface makes up for some of these. The turning gesture was motivated by the above studies; since our robot's face is displayed on a 2D surface, the only way to mimic eye gaze is to turn the entirety of her body. By doing so it will be clear that our robot wants a specific person's attention.

Humans can also be influenced by the robot's emotional behavior conveyed in the form of gestures and subtle body movements. Leyzberg et al. [9] performed a study in which a robot was scored based on how well its human participant demonstrates dance moves, they found that when the robot appropriately responded to a score (i.e. we did great! to a higher score), the participant was more likely to form a bond with the robot. On the other hand, the groups where the robot responded inappropriately or apathetically showed that this same bond was not generated. Similarly, several other studies yielded similar results in which the robot's expressiveness influenced the human's behavior and interactions with the robot [5, 7]. Takayama et al. [11] also found that having the robots show forethought (i.e. turning in the direction they want to go to before moving in that direction) made people more confident in what they thought the robot was going to be doing and that having the robots appropriately respond to their success or failure to complete a task made them seem more intelligent and competent to the humans observing them. Though the forethought nor the reactions are the focus of this research, they were added to make the robot more human-like and more approachable.

3 Approach

3.1 SARAH

We worked with our robot SARAH (**S**ocially **A**ppropriate **R**obot that **A**pproaches for **H**elp) to perform experiments (figure 2). Her face displays on a monitor held up by two metal rods, which in turn are attached to her Pioneer P3-DX robot base. The face and voice are provided by homer_robot_face as mentioned above, and Festival Text-to-Speech[2], respectively. Her voice is projected using a set of USB speakers, and video and audio were streamed using a webcam with a built-in microphone. A keypad was used for subject input.

Figure 2: SARAH's full form.

3.2 Observational Wizard-Of-Oz Studies

SARAH was placed by the elevator in Wold Atrium (figure 3), a building with a high volume of people traffic, to gather as big of a sample size of the campus population as possible (figure 4). This study was conducted "in-the-wild," where there was no controlled environment. This type of study was used to ensure that the subjects who passed by gave their genuine reactions to interacting with SARAH, and these

[2] http://www.cstr.ed.ac.uk/projects/festival/

studies yield a more realistic perspective on how the robot will function in real life with people who are unaware of her purpose [3]. We used a Wizard-of-Oz (WoZ) protocol, where the experimenter controls the robot without the subject's knowledge while the subject thinks the robot is functioning autonomously. They do not see the wizard controlling the robot's facial expressions and gestures as they interact with it. Using this protocol gave us more control over the environment without being intrusive and interfering with the subjects' behaviors. I controlled SARAH from a nearby computer lab using a simple graphical user interface (figure 5), which was able to control her facial expressions and dialogue, and an Xbox 360 controller for movement controls.

Subject reactions were recorded and analyzed to see how behaviors led to good interactions and the person helping the robot with her task. Eventually the data will be used to implement similar behaviors into the robot so that she can autonomously change her expressions and gestures.

Figure 3: SARAH standing in Wold Atrium by the elevator.

3.3 Experiment

The experimental controls were the kinds of greetings SARAH gave to people as they walked by. As people passed, SARAH would greet them with ”Hello! Can you please help me?” She executed this greeting in two ways:

- While remaining stationary
- While turning to follow the person, mimicking eye gaze

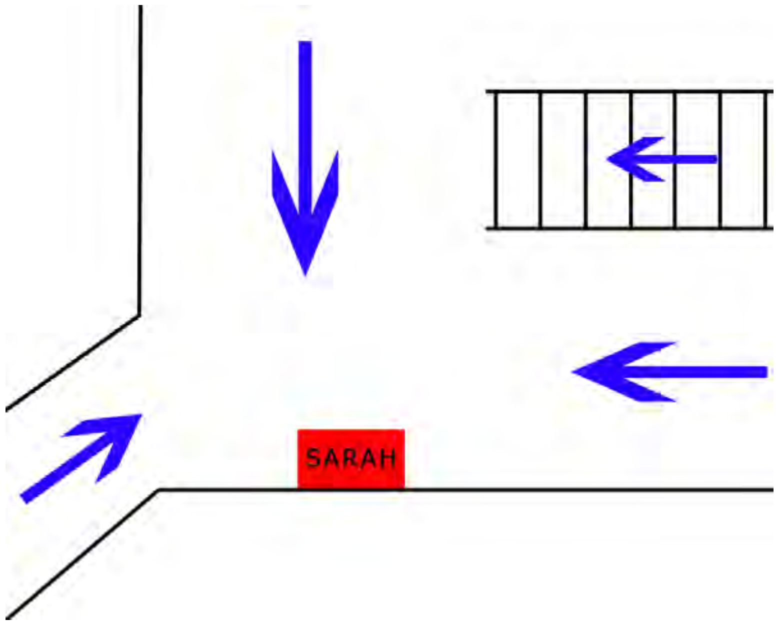

Figure 4: Bird's eye view of SARAH's location. The arrows indicate where people can come and go from.

When someone approached SARAH, she would ask them to complete a simple task. In this case the subject had to press a specific number on her keypad. If they successfully carried out the task, she would thank them with a smile.

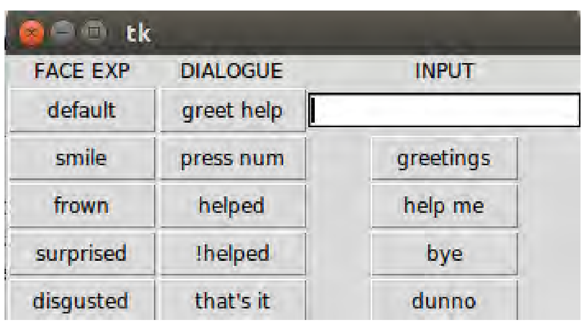

Figure 5: The face GUI.

Before sending SARAH out to her usual spot, I started her up using a roslaunch file that contained xml files to launch her base, face and GUI, webcam, and audio stream. On SARAH's computer I would run a custom rosnode that launched a 1x1px pygame window to listen for keyboard input so that I could see which number the participants were pressing. To record the webcam stream, I used gtk-recordmydesktop, which took a video and audio capture of the window in which the webcam was streaming to.

SARAH would stand outside in Wold for around 45 minutes to an hour at a time on random days at random times. During the second half of experimentation I took the time to balance out the days by doing no movement experiments on days and times where movement data had already been collected and vice versa, to ensure that there was little bias towards one day and time. Though her robot base batteries had energy for 5-6 hours, her computer battery had a much shorter life of about two hours. I found that after an hour, the lag time increases by a significant amount as the computer gets more stressed because it has less battery.

During this period of time, I would be timing my greetings based on what I could see from the webcam stream. As soon as someone got within 10 feet of her on the video stream, I would press the ”greet help” button on the GUI, corresponding to the greeting phrase ”Hello! Can you please help me?”

When someone approached her, I would press the ”press num” button on the GUI so that SARAH would say ”Can you please press # on the number pad?” to the participant who approached her. If the participant pressed the button, I would press the ”helped” button so that SARAH would say ”thank you” and change her expression to a smile. In rare cases, the ”!helped” (not helped) button was used when someone approached her but left without helping. Around one to two cases resulted in the person coming back and helping her, while in the other cases people seemed apologetic but kept walking away.

The additional buttons and text input on the GUI were for when participants wanted to pursue further conversation with SARAH. These topic of these conversations was noted for anecdotal observations, but since the purpose of this study was not related to random conversations, they were not heavily analyzed.

3.4 Data Collection and Analysis

Video and audio data from SARAH’s webcam was recorded and analyzed. The total number of people who passed by SARAH, acknowledged her, and helped her were all counted. However, people who were aware of the experiment were not part of the total count. People who passed by multiple times were only counted one. Groups were counted as one person for simplicity, but analysis was not done to see how groups behaved around SARAH due to time constraints.

In total around 14 hours of video was collected over the course of 7 weeks. There were 14 videos in total, around 45 minutes to an hour each. The video recordings were of SARAH’s camera feed, which included audio that could be picked up by her webcam microphone.

I manually watched and counted the people in the videos, simply by tallying up the people who fit the criterion above. This method is prone to human error, and as such the results may differ by 10-20 people. It was harder to determine if someone was acknowledging SARAH if they were already heading in her direction, but in most cases people would continue to look at her as they turned to go down the hallway or up the stairs, making it obvious that they were paying attention to her. I also knew a majority of the people who already knew about the experiment, so they were not counted, but it was likely that there were others who knew about it that I was not aware of.

4 Results

4.1 Empirical Results

4.1.1 Number Of People Who Helped In Total

In total, 1658 people (counting groups as one person) passed by SARAH, while 714 looked at her and 108 helped (figure 6). Of the total people who passed, 802 passed during no movement days and 856 passed during movement days.

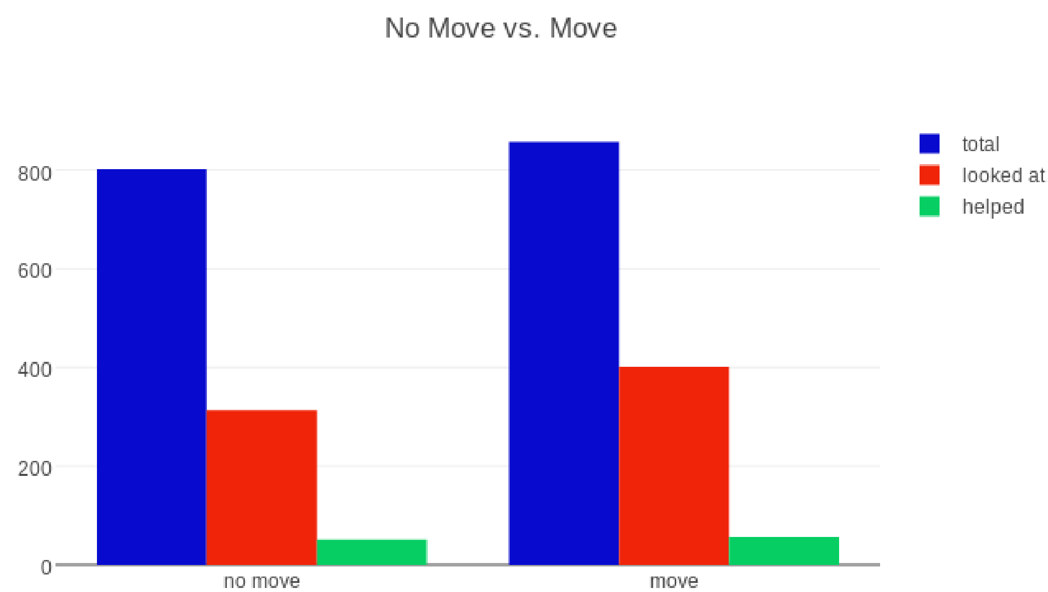

Figure 6: Total people passed (blue), looked (red), and helped (green).

4.1.2 Number Of People Who Looked And Helped On Movement And No Movement Days

Of the total people who looked, 313 (39.03%) looked during no movement days and 401 (46.85%) looked during movement days (figure 7). Using Pearson's Chi-squared test we found the above results were statistically significant. Of the people who helped, 51 (6.36%) people helped on no movement days and 57 (6.66%) people helped on movement days.

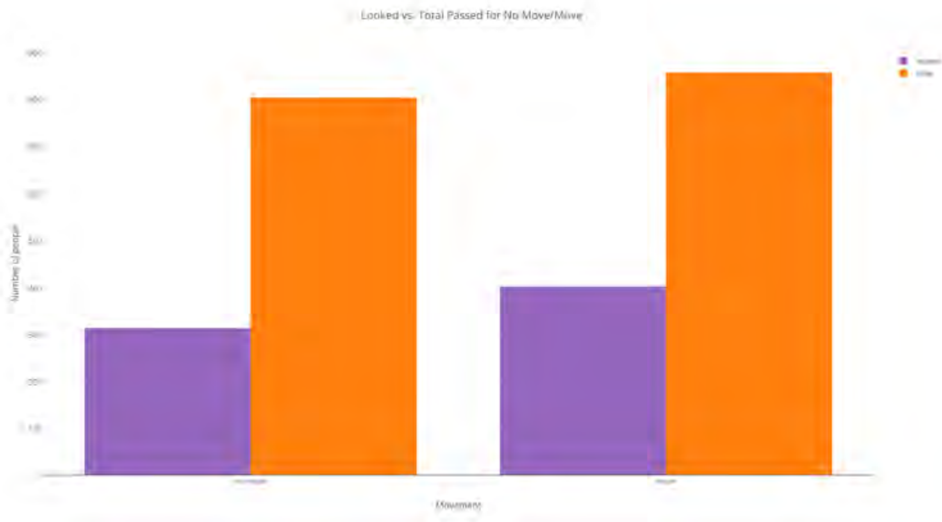

Figure 7: Looked (purple) vs. Passed (orange)

4.1.3 Number Of People Who Helped After Looking On Movement And No Movement Days

Of the total number of people who looked, 16.29% helped her on no movement days and 14.21% helped her on movement days (figure 8).

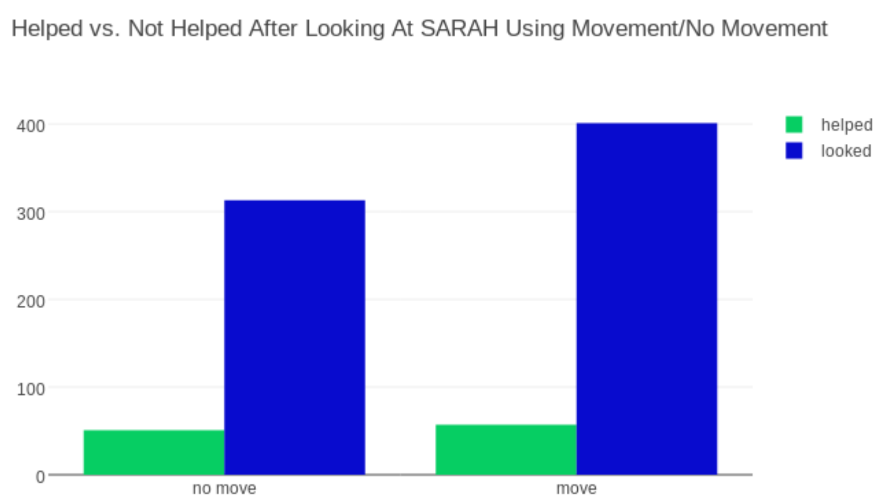

Figure 8: Helped (green) vs. Looked (blue)

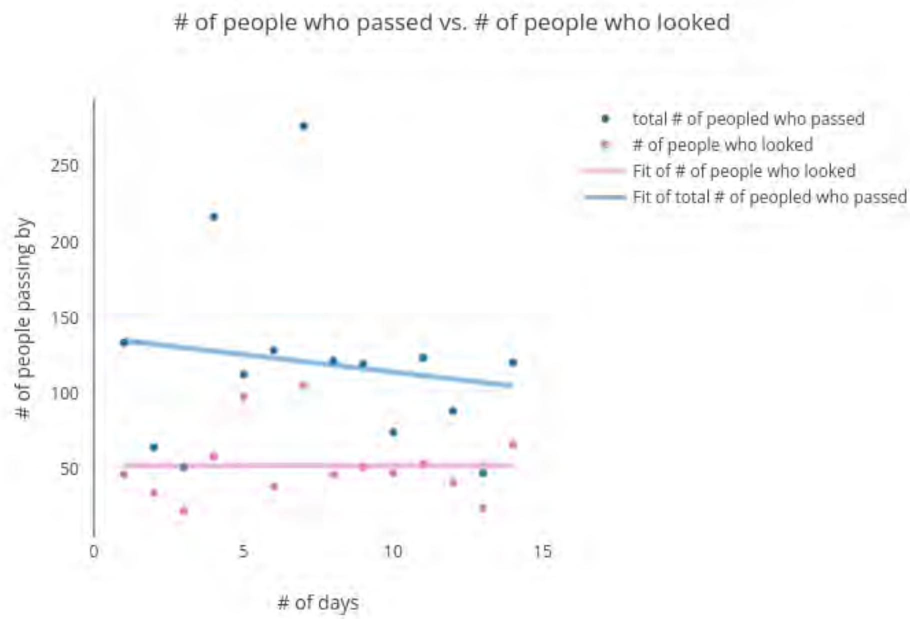

Figure 9: Total number of people who passed and looked over 14 days.

4.1.4 Trend Of People Passing By And Looking

The trend of people passing and looking over time was found to check if people were getting used to SARAH and thus were not looking at SARAH as often (figure 9). Over time the number of people who passed by decreased slightly, but the number of people who looked stayed consistent. The decrease in people passing by was likely due to the colder weather and busier schedules as the school term came to a close.

4.2 Anecdotal Results

4.2.1 People Who Helped

Of the people who helped SARAH, common reactions included:

- ”That’s it?”
- A wave, thumbs up, or other gesture to acknowledge her thanks
- Confusion
- A few subjects carried on conversations after the task, with the following reactions:
  - ”Wow! She’s so intelligent!”

- – Walking down the hallway to the computer lab to find the "wizard" after realizing that the robot was likely controlled by a person

It seemed that many people were confused about why SARAH wanted help with her task and what the importance of the task was, hence the many "that's it" reactions. They expected her to want more than help with a simple task, either in the form of wanting further conversation or perhaps some kind of unexpected action that would cause them to have a reason to fear her. In the rare cases people carried on conversation with her, they would bring up topics that humans would usually not bring up with other humans randomly. For example, a group of two women came up to SARAH, helped her with her task, and then immediately proposed to her, offering their "eternal love" to her. In another case, a man helped SARAH and attempted to buy her, asking her how much she cost and who owned her. Though they thought she was autonomous and seemed to think that she was sentient, the fact that she was a robot likely encouraged them to act strangely.

4.2.2 People Who Looked

Of the people who looked at SARAH but did not help, common reactions included:

- Waving at her, thinking she only wanted to greet them
- Taking a photo/video with their phones, but not helping
- Surprise that she could talk/move
- Strong refusals to help
- Looking back after being far up the nearby staircase or the hallway; these were missed opportunities due to limitations

Most of the people who looked but did not help were not afraid of SARAH in the slightest, but the few that were made it very clear they were uncomfortable. Otherwise, the people were mostly busy or engaged in another task. For the people who took photos and videos of her, they seemed to not care that she needed help and instead was entertained by her robotic nature.

The missed opportunities were related to hardware limitations. Because there were no environmental cameras, I was timing my responses based on how well I could see people through SARAH's camera feed. This led to many timing problems because I would not be able to plan my responses ahead of time, causing people to walk out of what they considered an acceptable range to turn back and help her.

4.2.3 People Who Passed

Of the people who passed by without looking at SARAH, common factors included:

- Wearing headphones
- Not being able to hear her in general
- Focus on getting to their destination
- Engaged in a conversation

Similar to the people who did not look, the people who passed without any acknowledgement were busy getting to a set location or not focusing on their environment at all. As discussed later, some people could not hear her and thus had no reason to turn their attention towards her, something that can be easily fixed with louder speakers for SARAH but was not an option for these experiments.

5 Limitations

The biggest limitation we faced was the lack of environmental cameras. People who looked at SARAH while they were out of view of the webcam were unable to be counted towards the total. I also could not properly time my greetings, leading to several missed opportunities where I was not able to greet people in time because they came and left the webcam's range too fast.

This exacerbated another problem involving lag time; there was around 2 to 3 seconds of lag from when I saw the person and when SARAH greeted them. This was the biggest reason why I missed several people passing by. Without knowing when someone was approaching SARAH, I was not able to properly time my greetings and intercept them. On movement days, this was slightly less of a problems; since SARAH was supposed to move, I could turn her and observe the hallways and stairwell without breaking the experimental parameters. But on no movement days, it became much more of a problem.

SARAH's speakers were also not very loud. A simple decibel test showed that her voice was around 63 decibels while the environment was also around 63 decibels. This further aggravated the lag problem. At the very least, even if I could not have timed my greeting, she could have been loud enough to attract their attention from a larger range. Some subjects admitted that they could not really hear her even from a few feet in front of her.

6 Conclusions

Though more people looked at her while she was moving, there was a negligible different between the number of people who helped her while she was not moving and while she was moving. Likewise, though more people looked at her when she moved, they were not necessarily more inclined to help her. Movement increases looks, but does not correlate to an increase in interaction.

On a more anecdotal note, being a robot is both an asset and detriment for SARAH. As an asset, SARAH easily attracts attention because being a robot is a novelty that not many will see on a daily basis. Many people came up to her, exclaiming ”oh, it's the robot!” to themselves or to their friends, regardless of whether she was moving or not. Likewise, as a detriment, people were less likely to come up to her and help because they were afraid of her. Many others refused to help her to her face or walked away faster in fear, sometimes exclaiming ”ah, it's the robot!” as they passed by.

7 Future Works

We will run more experiments with SARAH where the mentioned limitations are no longer existent, so that more accurate results can be gathered. Using what we now know about how people react to SARAH, we can also modify our experiments to find ways to distract people from the SARAH's novelty and get them to pay attention to her because they are interested in helping her.

In those experiments, we will have the necessary equipment so that we can avoid the limitations we faced for this experiment.

References

[1] Casey C. Bennett and Selma Šabanović. Perceptions of affective expression in a minimalist robotic face. In *Proceedings of the 8th ACM/IEEE International Conference on Human-robot Interaction*, HRI '13, pages 81–82, Piscataway, NJ, USA, 2013. IEEE Press.

[2] Mark S. Cary. The Role of Gaze in the Initiation of Conversation. *Social Psychology*, 41(3):269–271, 1978.

[3] Aaron G Cass, Eric Rose, Kristina Striegnitz, and Nick Webb. Determining appropriate first contact distance: trade-offs in human-robot interaction experiment design. In *Workshop on Designing and Evaluating Social Robots for Public Settings at the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2015)*, 2015.

[4] Justine Cassell. Embodied conversational agents. chapter Nudge Nudge Wink Wink: Elements of Face-to-face Conversation for Embodied Conversational Agents, pages 1–27. MIT Press, Cambridge, MA, USA, 2000.

[5] Vijay Chidambaram, Yueh-Hsuan Chiang, and Bilge Mutlu. Designing persuasive robots: How robots might persuade people using vocal and nonverbal cues. In *Proceedings of the Seventh Annual ACM/IEEE International Conference on Human-Robot Interaction*, HRI '12, pages 293–300, New York, NY, USA, 2012. ACM.

[6] Frédéric Delaunay, Joachim de Greeff, and Tony Belpaeme. A study of a retro-projected robotic face and its effectiveness for gaze reading by humans. In *Proceedings of the 5th ACM/IEEE International Conference on Human-robot Interaction*, HRI '10, pages 39–44, Piscataway, NJ, USA, 2010. IEEE Press.

[7] Bradley Hayes, Daniel Ullman, Emma Alexander, Caroline Bank, and Brian Scassellati. People help robots who help others, not robots who help themselves. In *The 23rd IEEE International Symposium on Robot and Human Interactive Communication, IEEE RO-MAN 2014, Edinburgh, UK, August 25-29, 2014*, pages 255–260, 2014.

[8] Brandon Heenan, Saul Greenberg, Setareh Aghel-Manesh, and Ehud Sharlin. Designing social greetings in human robot interaction. In *Proceedings of the 2014 Conference on Designing Interactive Systems*, DIS '14, pages 855–864, New York, NY, USA, 2014. ACM.

[9] Dan Leyzberg, Eleanor Avrunin, Jenny Liu, and Brian Scassellati. Robots that express emotion elicit better human teaching. In *Proceedings of the 6th International Conference on Human-robot Interaction*, HRI '11, pages 347–354, New York, NY, USA, 2011. ACM.

[10] Sourabh Prajapati, C. L. Shrinivasa Naika, Shashi Shekhar Jha, and Shivashankar B. Nair. On rendering emotions on a robotic face. In *Proceedings of Conference on Advances In Robotics*, AIR '13, pages 19:1–19:7, New York, NY, USA, 2013. ACM.

[11] Leila Takayama, Doug Dooley, and Wendy Ju. Expressing thought: Improving robot readability with animation principles. In *Proceedings of the 6th International Conference on Human-robot Interaction*, HRI '11, pages 69–76, New York, NY, USA, 2011. ACM.

[12] M. L. Walters, K. Dautenhahn, R. t. Boekhorst, K. L. Koay, and S. N. Woods. Exploring the design space of robot appearance and behavior in an attention-seeking 'living room' scenario for a robot companion. In *2007 IEEE Symposium on Artificial Life*, pages 341–347, April 2007.