How embodiment impacts the perception of a system's weaknesses in understanding accents

Travis Clauson
Department of Mechanical Engineering, Tufts University, Medford, USA
Boston, United States
Travis.Clauson@tufts.edu

Kyu Rae Kim
Department of Computer Science, Tufts University, Medford, USA
Boston, United States
Kyu_Rae.Kim@tufts.edu

Srisharan Kolige
Department of Mechanical Engineering, Tufts University, Medford, USA
Boston, United States
Srisharan.Kolige@tufts.edu

Pierrick Lorang
Department of Mechanical Engineering, Tufts University, Medford, USA
Boston, United States
Pierrick.Lorang@tufts.edu

ABSTRACT

Proper human-robot interaction requires a good understanding of shared information on both sides, and natural language is currently one of the most common and convenient means of sharing information. English is the most commonly used language in the world, and it has a very diverse set of accents that robots must understand to communicate verbally with humans. Current natural language understanding techniques do not sufficiently consider various pronunciations, and robots generally recognize only American English accents. In this research, the effects of natural language processing - specific to American English - on the experiences of people from different linguistic backgrounds during human-robot interaction are studied. In order to test the effects, two games were made that require verbal interactions with an embodied agent and a disembodied agent. This paper focuses on two main aspects. First, the impact of accents on human-robot cooperative task completion is assessed. Then, the influence of the embodiment on the experiences of human agents, such as how frustrating the experience is or to what extent they blame the system during verbal communication, is evaluated.

KEYWORDS

Natural speech processing, Human-robot interaction, Social study, Embodied systems, Accent understanding, Satisfactory collaboration

1 INTRODUCTION

With the increase of robots in daily life, it is imperative that AI must have robust verbal communication skills. The first step a robot must take to understand natural language is translating the words to text, known as natural language processing or NLP. This paper analyzes this aspect of verbal communication because it is the most tangible and consequential means of sharing information between a robot and a human. When this step fails, the agent is left either with an improper or no understanding of the user's commands, which can have negative effects on their experiences. This can be easily relatable due to the increasingly regular use of voice assistants. However, this emotional distress may vary when the misunderstanding comes from an embodied agent such as Misty or a disembodied agent such as Siri or Alexa. Understanding whether embodiment affects users' perceptions towards the agent can help designers create better user experiences, especially for those who naturally struggle to be understood by NLP.

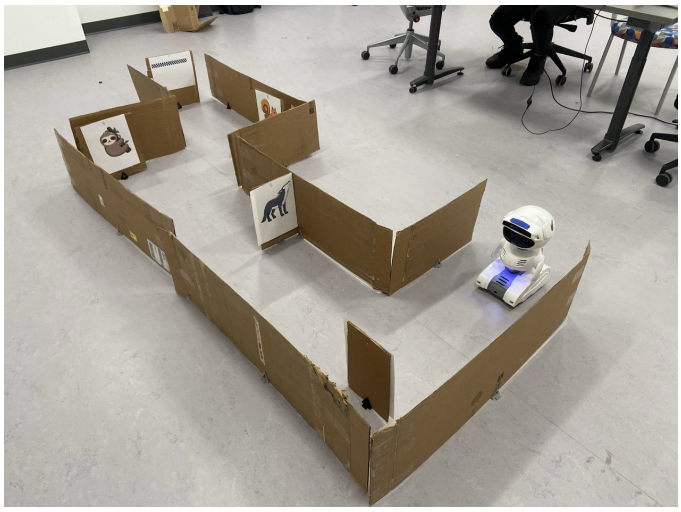

Figure 1: Misty is sitting at the start line of the maze, awaiting verbal instructions from a participant.

This study looks at the emotions and level of blame users place on embodied NLP versus disembodied NLP when it is not performing proficiently. Specifically, the study focuses on comparing the experience of native English speakers and non-native English speakers. We will give the participants questionnaires beforehand asking about their experience with English, their confidence in communicating, and their pronunciation in relation to their appreciation of native English. The actual experiment is divided into two stages: one where a user repeats given phrases into a computer interface and the other where a user employs a list of commands to direct the robot through a maze. Identical phrases are used in the two stages so that the performance of participants is not misconstrued by certain words that may be difficult to pronounce. We will measure

, , , . 2022.

the participant's speaking time, the average number of repetitions of questions, the time taken to complete each task, and success in collaboration and compare these data to each other. We will then ask them to rate their trust in the robot during the collaboration, their comfort with the task, and their frustration or satisfaction, as well as overall feedback.

This paper offers a unique perspective by analyzing two problems side by side that have previously been studied individually: NLP systems' weakness in understanding users with accents and how anthropomorphic features influence users' experiences and expectations. A paper by Julia Fink [7] stated that users have higher expectations of intelligence for agents with more anthropomorphic or lifelike designs. This claim would predict that users may be more frustrated by an embodied robot since their expectations for its intelligence are subconsciously higher. The data collected by this paper comes to a similar conclusion.

2 BACKGROUND

Verbal communication is defined by many features: from pronunciation, speed, tone, and expressions to the epistemic assumptions that frame our understanding of our environment. This leaves room for many sources of misunderstanding in communication, which are nevertheless irreplaceable features of our communication. This is emphasized in the current interactive systems between man and machine [8] and allows simplified access to robotics even for non-experts [3]. However, these techniques are not yet optimal and force humans to adapt their way of speaking to be understood by robots [12, 19]. Moreover, current speech understanding techniques are more focused on a native English-speaking population, and lack performance when interacting with people from different linguistic backgrounds [14]. Some researchers have sought to understand the mechanisms that lead to these intra-system uncertainties for years [20], and more recent work has also attempted to improve their robustness for a wide range of language proficiencies [28]. Understanding speakers with accents is notably an acknowledged issue for NLPs and many efforts are being put into solving this challenge [4, 10, 11, 16, 17, 22, 23, 27]. Other work focus on analogical issues of speech recognition and propose new methods to solve them like dialect recognition [9] or language adaptation [5]. Finally, some research studies focus on misunderstandings due to a lack of ontological, semantic, or epistemic knowledge where it is necessary to ask for contextual clarifications for an accurate understanding [6]. Overall a strong emphasis has been put in the last decade on improving human-system interactions. This research tries to improve the collaboration by studying participants' feelings during their experience with a robot [2, 18, 26], bringing solutions to recognize and measure them [1, 15, 25] or implement users differences [24] and experiences [21] flexibility into the systems. These efforts are necessary because a poor comprehension system could bring more harm than good [13].

These prior works evaluate how speech understanding systems can be impacted by speech variation in direct performance on the system itself. In this study, we are interested in assessing the impact of speech modulation factors on a complex collaboration between a human and a robot to complete a task. We will evaluate both the performance on the collaborative tasks and the participant's

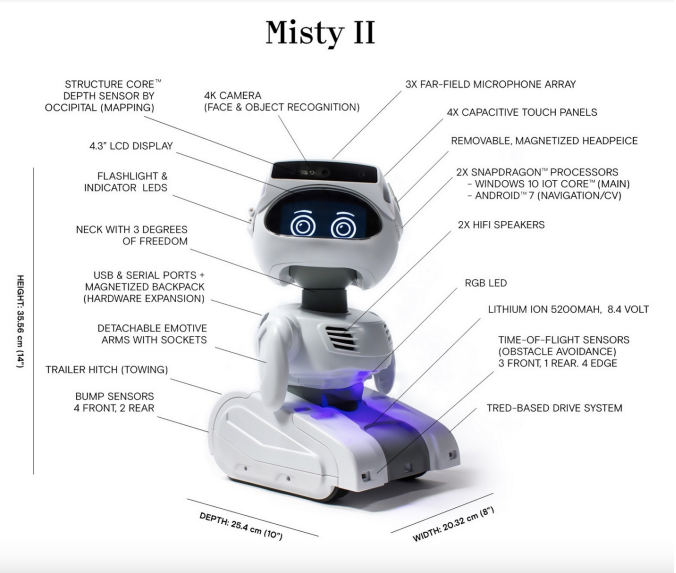

Figure 2: Misty is an autonomous roaming robot designed by Misty Robotics to make robot programming less intimidating and more accessible through its advanced perception and actuation.

experiences of the interaction. For each task, we will have pooled results with native English speakers and non-native speakers. Then we compare their performance and satisfaction over the two tests.

3 METHODOLOGY

This experiment was designed to test how participants with non-native English accents fare with verbally directing a robot through a maze compared to how native English speakers do. It was hypothesized that the natural language understanding system used throughout the study would find it difficult to understand non-native speakers' commands, making it harder for the participants to complete the task. We specifically designed the tasks so that the participants had to interact verbally with the robot. We tested our hypothesis by measuring the ability to communicate effectively and the level of satisfaction involved in completing the task. The second part of the experiment concentrated on measuring the emotional outcomes of interacting with an anthropomorphic robot compared to a computer screen when the NLP system was not able to understand what the user was trying to convey. We hypothesized that embodiment would increase the expectations, and therefore, the participants would be more frustrated when interacting with a robot compared to interacting with a computer.

3.1 Experimental Design

The first stage of the experiment tested the participant's interaction with a disembodied NLP system by requiring them to repeat a given set of phrases into a computer. If the computer was able to understand the users, it allowed them to continue to the next phrase. If it did not understand the user, it requested the user to repeat. In the second stage, the user had to interact with an embodied NLP

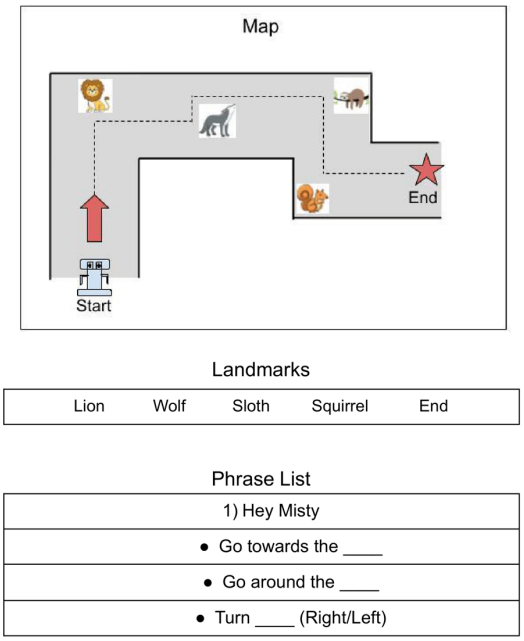

Figure 3: A diagram of overhead map of the maze, and a list of landmarks and commands given to participants.

system, represented by a small robot, Misty. The user's goal was to verbally navigate the robot through a maze with landmarks and obstacles. These features are detailed on the handout, shown in Figures 3, given to the participants to help them understand the playing field and which phrases they could use to communicate with the robot. Then, the process of communicating with Misty was explained to the participants, including how they must prime it by saying, "Hey, Misty", before directing it. Finally, they were given a microphone connected to the NLP system and told to begin.

3.2 Robot Design

In this project, human participants had to communicate with a robot to navigate through a maze while avoiding obstacles. Mounted with numerous features, a Misty robot was chosen as the robotic agent for this study.

Hardware. Because of the specific settings of this experiment, robotic arms and grippers were not necessary, and the compact size of Misty - 8×10×14 inches - and its stability in movements due to its continuous-track-based driving system made Misty robot perfectly apt for the project. In order for Misty to navigate through the maze, the IMU and the wheel encoders enabled the robot to be aware of where it was in the environment, and the various sensors on the robot, such as depth, distance, and collision sensors, ensured Misty to navigate through the maze without being immobile.

Behavior Design. For this study, the only NLP system used within Misty was key phrase detection that listens to the phrase "Hey Misty", since its onboard NLP is still in the beta phase. This triggered Google's speech recognition system to translate the user's

Figure 4: A graphical user interface for testing the interaction with disembodied system.

speech to text, which was then compared with the expected phrases to send commands to Misty based on the correctness of the interpretation. If Misty is given an unexpected commands, such as move towards a landmark not in its sight, it gives a verbal feedback to the user about its current perception of the maze. A set of commands and vocabularies that the robot can understand were explicitly programmed, and the possible answers or actions according to the command and the sensor data were carried on with its internal state machine. Once the robot starts at its initial location within the maze, it navigates through the environment in response to the user's command until it detects a landmark or an obstacle. When the robot is unable to understand the command, it asks the human repetitively until it does. It repeats this sequence of actions until it reaches the goal location.

3.3 Participants

Selecting the right set of participants was one of the most crucial aspects of our study. We selected around 20 participants from various linguistic backgrounds:

- Native American English Speakers - 8 participants
- Native Non-American English Speakers - 2 participants
- Non-native English speakers (English as a second language) - 10 participants

The main objective of the study is to identify the Robot's responsiveness to participants who speak English as their second language compared to those from English speaking countries especially from America, as most of the Natural Language Processing models are built based on American English accents. Due to the linguistic diversity, the participants were selected from the students in Tufts University. The average age of the participants was around 25, and there were a total of 14 male participants and 6 female participants.

3.4 Data collection

The participants were given a set of tasks to be completed in 10 minutes. They were advised that they were being recorded for research purposes, but they were not told what the exact research was

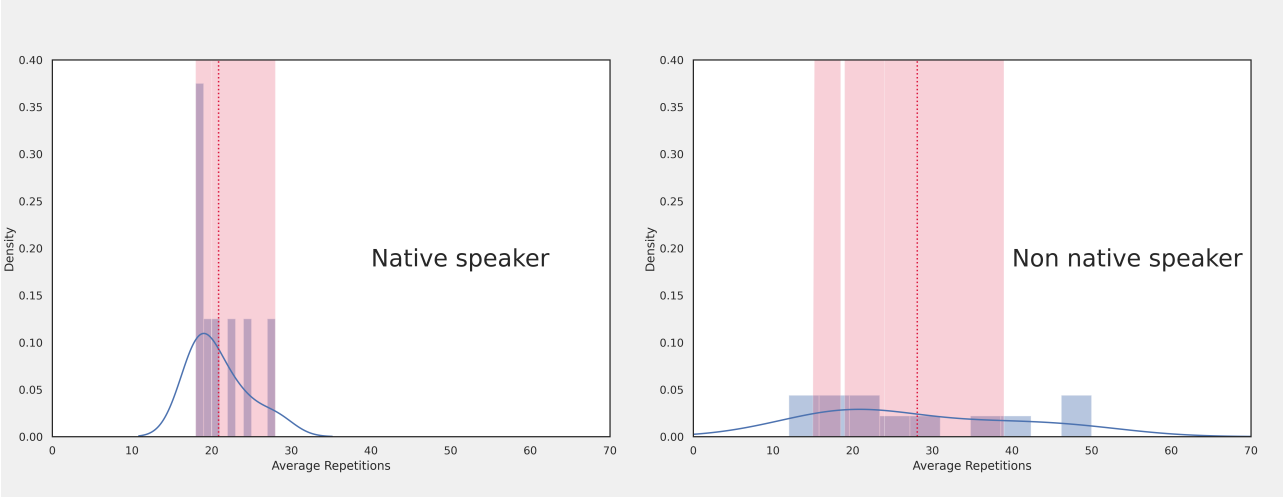

Figure 5: On the left - Average number of repetitions for American-English speakers. On the right - Average number of repetitions for non American-English speakers. The red doted line represents the average and the red area represents the standard deviation.

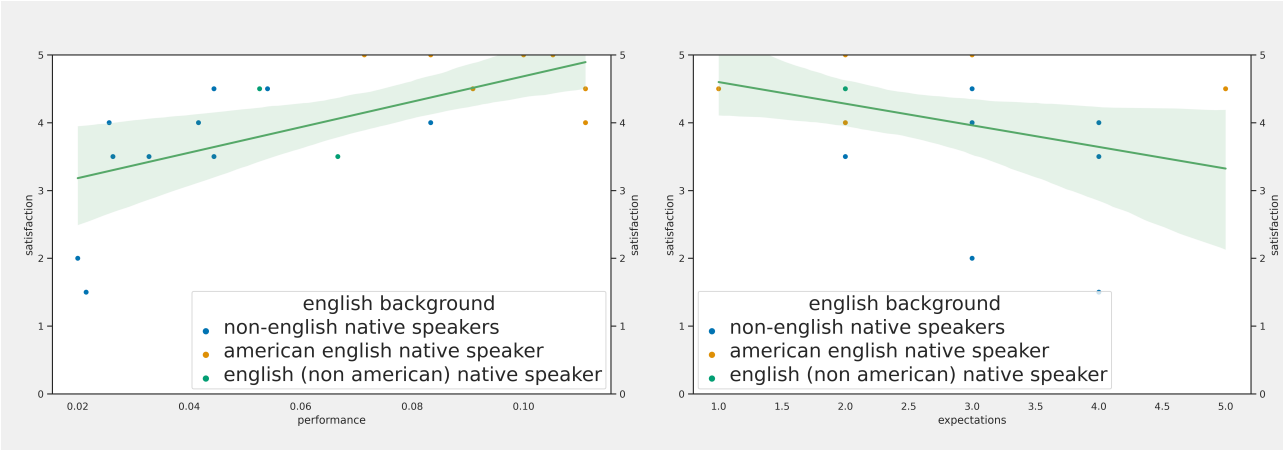

Figure 6: On the left - Satisfaction of participants over general performance. On the right - Satisfaction of participants over pre-experiments expectations. The line represents the interpolated average and the green area represents the standard deviation.

until they finished their tasks as we did not want the participants to be conscious about their accents, but rather just focus on the given tasks. The following quantitative data were collected from the recordings:

- Whether each sub-task was completed
- Time taken to complete the entire task
- Number of times a single question was repeated
- Number of times the repetitions occurred in total

After completing the task, each participant was asked to complete a survey with five questions for each of the two tasks - interaction with embodied and disembodied agent.

3.5 Data analysis

This study focused on the impact of accent misunderstandings during the collaboration and its success, as well as the emotional impact that this communication caused on the participants during their interaction with the robot and the computer. These two aspects were evaluated using quantitative and qualitative measures, respectively.

Baseline. The basic purpose of this study was to compare the performance and emotional outcomes of non-native speakers with those of native speakers when interacting with the robot and a computer. We compared each of our metric results of the non-native speakers to those of the native speakers. We also asked participants

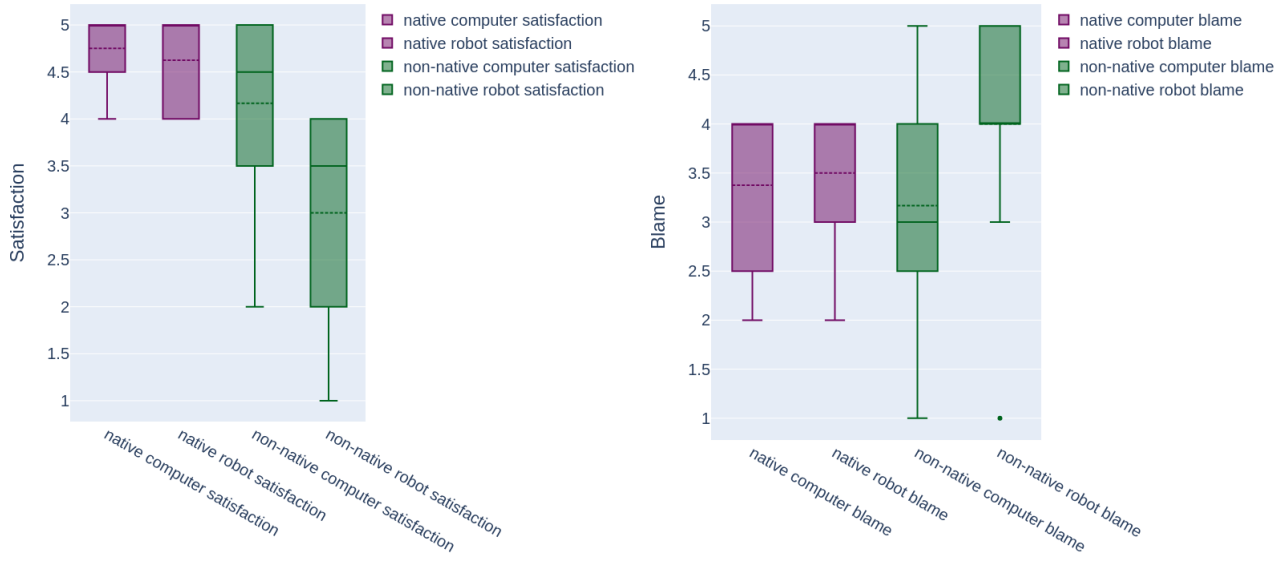

Figure 7: On the left - Box plot of the blame per group of participants for disembodied versus embodied system. On the right - Box plot of the satisfaction per group of participants for disembodied versus embodied system.

about their assessment of their own English skills compared to native speakers on a 5-point Likert-type scale in the questionnaire given before the experiment. This allows us to weigh the results obtained using their direct assessment of their confidence in their language level and subsequently to draw conclusions between the appreciation of one's language level, the performance and satisfaction in collaboration with a speech-capable robot. Additionally, we collected data from the participants to study their emotional outcomes when interacting with a robot versus with a computer. Participants were asked about their satisfaction and frustration levels during their interaction with both the computer and the robot, thereby deriving their emotional outcomes towards a robot as opposed to a computer.

Performance of the collaboration. The first performance metric was the average success rate on all tasks. In order to have a quantification of efficiency during the interaction, we have measured and computed the duration of participants' speeches, the time taken to complete the goal and the ratio of reiterations per command.

Frustration or Satisfaction of the Interaction. In order to evaluate the emotional impact of miscommunication with either a robot or a computer due to diverse linguistic backgrounds, we used the questionnaires completed by the participants upon the completion of the experiments. We were able to compare the participants' comfort in these interactions and put it into perspective with their appreciation of their language level. We were interested in the difference in satisfaction and frustration that these interactions can pose in comparison to the experience of native speakers. Additionally, we compared the participants' enthusiasm before and after the interaction, as well as their confidence in the robot during the experience. Finally, we asked about their thoughts on the experiments.

4 RESULTS

The two graphs in Fig.5 show how many times each participant had to repeat the commands before the robot or computer could understand them. The *Right* graph shows the number of repetitions of the commands for a native American English speaker, and the $Left$ graph for a non-American-born English speaker.

First, the graphs show that NLP system was better at understanding the American-born English speakers compared to the non native American-English speakers. The former had to repeat the commands on average of 22 times, while the latter had to repeat the commands on average of 28 times. This difference was also reflected in the timing; American-born English speakers completed the two tasks on average approximately 6 minutes, while the non native American-English speakers took an average of 10 minutes to complete the tasks.

Additionally, participants with higher success rates are more satisfied with their interaction with both embodied and disembodied systems, as shown in Fig.6 $Left$. Moreover, satisfaction seems to be correlated with early perception of the agent's abilities to understand human speech, cf. Fig. 6 *Right*. In fact, higher expectation, generally correlated to lack of real life robotics experience, tends to negatively impact the perceptions towards the agents. These results are more apparent as the performance of the system to understand the participant decreases. The latter graph shows a higher overall

standard deviation, reflecting lower certainty in this outcome, especially in regards to participants who had higher pre-experiment expectations.

Finally, there were more emotional attachments to the embodied system when it failed to understand the command. Fig.7 *Left* shows that non-native speakers are more frustrated with their interaction with the robot than when they interact with a computer. Fig.7 *Right* shows a similar result except the participants blamed the robot more for not understanding them correctly, even though the NLP system used in both cases was exactly the same. We can infer from this observation that humans have higher expectations of embodied systems and thus are more likely to accept the failures of the disembodied system.

5 DISCUSSIONS

Many studies have analyzed NLP system with non-native accents along with expectations and reactions for embodied versus disembodied agents. However, our study was unique as it evaluated the satisfaction of natural language processing with a diverse speaking group of participants while interacting with embodied and disembodied systems. This study confirmed that NLP systems are very specific to one type of accent, mostly specific to American English accent, and do not perform with people with accents. The results showed a clear negative effect on the performance and user experience of non-native speakers. Furthermore, users are more likely to blame embodied agents, which demonstrates that embodied agents must have a more robust NLP system to match the users' expectations. In the future, NLP must be trained with more diverse accents to have better performance and better user experience.

6 CONCLUSION

In this work, we studied how people from different linguistic backgrounds verbally interact with robots with natural language processing systems, commonly specific to American English accents. Our hypothesis followed previous work that participants would have higher expectations and place greater blame on embodied robots. The data we collected - performance metrics on the two stages of the test and the survey data - proved this hypothesis to be correct.

7 ACKNOWLEDGEMENTS

Travis Clauson. While we planned out our experiment, I took on the responsibility of designing the physical and social aspects of the experiment. This included defining the acceptable commands and a word bank, sketching a maze design, and choosing experimental parameters such as the maximum number of fails. Later down the road, when we conducted our experiment, my main commitment was to build our physical maze with the appropriate landmarks. In regard to this paper, my contributions are focused around the Experimental Design section Methodology and the majority of the introduction.

Kyu Rae Kim. The main responsibility I had throughout the study was to program the Misty robot according to the experiment design made by the group and test it in the set environment so that we do not experience any unexpected behaviors. Although it was initially planned to solely use Misty's onboard natural language processing system for the entire experiment, I noticed that the system was not fully developed and thus decided to use only the key phrase recognition on Misty and use Google's speech-to-text system instead. I programmed so that Misty recognizes the key phrase, which then triggers Google NLP system to compare the speech to the desired sentences and to send out action commands to Misty. I have also developed and designed the computer GUI for testing the interaction between human and disembodies system.

Srisharan Kolige. My major contribution to the project was in the data collection part. I was responsible for recruiting the right set of participants for the study. My main motivation for the recruitment was to get a linguistically diverse set of participants for the study. I also consulted with Pierrick before designing the survey for the participants as to what kind of data he would prefer for the analysis part. I was also responsible for collecting other information during the study which could not be answered by the participants in the survey and was required for analysis and study. I further assisted Pierrick in the initial part of the data cleaning by taking the raw survey data and converting it into a .csv file for the analysis.

Pierrick Lorang. Pre-experimentation, I was in charge of finding appropriate metrics to qualify and quantify the experiments results as much as structure the data. Post-experimentation I focused on filtering and cleaning the data, as well as representing straightforward and easy to understand data graphs. Overall I was more on the data-analytic part of the project thus we discussed mainly with Srisharan about ways to approach HRI data collection and displaying. To clean, sort and restructure the data I mainly used *python* based on *.csv* files that I could then process under *jupyter notebook* using *panda dataframes* structures. To understand how the data should be approached I used a *profile* report. For displaying I used the *seaborn* format which is a convenient tool to build nice and refine graphs. My goal was to convey as much relevant information as possible in as few graphs as possible, while providing insights into the disparity of the data.

REFERENCES

[1] Moaed Abd, Iker González, Mehrdad Nojoumian, and Erik Engeberg. [n. d.]. Trust, Satisfaction and Frustration Measurements During Human-Robot Interaction.

[2] Tae Ahn and Michelle Lee. [n. d.]. User experience of a mobile speaking application with automatic speech recognition for EFL learning. *British Journal of Educational Technology* 47 ([n. d.]). https://doi.org/10.1111/bjet.12354

[3] Mustafa Can Bingol and Omur Aydogmus. [n. d.]. Performing predefined tasks using the human–robot interaction on speech recognition for an industrial robot. *Engineering Applications of Artificial Intelligence* 95 ([n. d.]), 103903. https://doi.org/10.1016/j.engappai.2020.103903

[4] Jordan J. Bird, Elizabeth Wanner, Anikó Ekárt, and Diego R. Faria. [n. d.]. Accent classification in human speech biometrics for native and non-native english speakers. In *Proceedings of the 12th ACM International Conference on PErvasive Technologies Related to Assistive Environments* (New York, NY, USA, 2019-06-05) *(PETRA '19)*. Association for Computing Machinery, 554–560. https://doi.org/10.1145/3316782.3322780

[5] Ronald Cumbal. [n. d.]. Adaptive Robot Discourse for Language Acquisition in Adulthood. In *Proceedings of the 2022 ACM/IEEE International Conference on Human-Robot Interaction* (Sapporo, Hokkaido, Japan, 2022-03-07) *(HRI '22)*. IEEE Press, 1158–1160.

[6] Fethiye Irmak Doğan, Ilaria Torre, and Iolanda Leite. [n. d.]. Asking Follow-Up Clarifications to Resolve Ambiguities in Human-Robot Conversation. In *Proceedings of the 2022 ACM/IEEE International Conference on Human-Robot Interaction* (Sapporo, Hokkaido, Japan, 2022-03-07) *(HRI '22)*. IEEE Press, 461–469.

[7] Julia Fink. 2012. Anthropomorphism and Human Likeness in the Design of Robots and Human-Robot Interaction. In *Social Robotics*, Shuzhi Sam Ge, Oussama Khatib, John-John Cabibihan, Reid Simmons, and Mary-Anne Williams (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 199–208.

[8] Patrik Gustavsson, Anna Syberfeldt, Rodney Brewster, and Lihui Wang. [n. d.]. Human-robot Collaboration Demonstrator Combining Speech Recognition and Haptic Control. *Procedia CIRP* 63 ([n. d.]), 396–401. https://doi.org/10.1016/j.procir.2017.03.126

[9] Naoki Hirayama, Koichiro Yoshino, Katsutoshi Itoyama, Shinsuke Mori, and Hiroshi G. Okuno. [n. d.]. Automatic speech recognition for mixed dialect utterances by mixing dialect language models. *IEEE/ACM Trans. Audio, Speech and Lang. Proc.* 23, 2 ([n. d.]), 373–382. https://doi.org/10.1109/TASLP.2014.2387414

[10] Irakli Kardava, Jemal Antidze, and Nana Gulua. [n. d.]. Solving the Problem of the Accents for Speech Recognition Systems. *International Journal of Signal Processing Systems* 4 ([n. d.]), 235–238. https://doi.org/10.18178/ijsps.4.3.235-238

[11] Kaleem Kashif, Yizhi Wu, and Adjeisah Michael. [n. d.]. Consonant Phoneme Based Extreme Learning Machine (ELM) Recognition Model for Foreign Accent Identification. In *Proceedings of the 2019 The World Symposium on Software Engineering* (New York, NY, USA, 2019-09-20) *(WSSE 2019)*. Association for Computing Machinery, 68–72. https://doi.org/10.1145/3362125.3362130

[12] James Kennedy, Séverin Lemaignan, Caroline Montassier, Pauline Lavalade, Bahar Irfan, Fotios Papadopoulos, Emmanuel Senft, and Tony Belpaeme. [n. d.]. Child Speech Recognition in Human-Robot Interaction: Evaluations and Recommendations. In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction* (New York, NY, USA, 2017-03-06) *(HRI '17)*. Association for Computing Machinery, 82–90. https://doi.org/10.1145/2909824.3020229

[13] Martin Labský, Jan Cuřín, Tomáš Macek, Jan Kleindienst, Ladislav Kunc, Hoi Young, Ann Thyme-Gobbel, and Holger Quast. [n. d.]. Impact of word error rate on driving performance while dictating short texts. In *Proceedings of the 4th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (New York, NY, USA, 2012-10-17) *(AutomotiveUI '12)*. Association for Computing Machinery, 179–182. https://doi.org/10.1145/2390256.2390286

[14] Audrey Mbogho and Michelle Katz. [n. d.]. The impact of accents on automatic recognition of South African English speech: a preliminary investigation. In *Proceedings of the 2010 Annual Research Conference of the South African Institute of Computer Scientists and Information Technologists* (New York, NY, USA, 2010-10-11) *(SAICSIT '10)*. Association for Computing Machinery, 187–192. https://doi.org/10.1145/1899503.1899524

[15] Youssef Mohamed, Giulia Ballardini, Maria Teresa Parreira, Séverin Lemaignan, and Iolanda Leite. [n. d.]. Automatic Frustration Detection Using Thermal Imaging. In *Proceedings of the 2022 ACM/IEEE International Conference on Human-Robot Interaction* (Sapporo, Hokkaido, Japan, 2022-03-07) *(HRI '22)*. IEEE Press, 451–460.

[16] Hyeong-Ju Na and Jeong-Sik Park. [n. d.]. Accented Speech Recognition Based on End-to-End Domain Adversarial Training of Neural Networks. *Applied Sciences* 11, 18 ([n. d.]), 8412. https://doi.org/10.3390/app11188412 Number: 18 Publisher: Multidisciplinary Digital Publishing Institute.

[17] Maryam Najafian and Martin Russell. [n. d.]. Modelling Accents for Automatic Speech Recognition. ([n. d.]), 1568.

[18] Anastasia K. Ostrowski, Vasiliki Zygouras, Hae Won Park, and Cynthia Breazeal. [n. d.]. Small Group Interactions with Voice-User Interfaces: Exploring Social Embodiment, Rapport, and Engagement. In *Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction* (New York, NY, USA, 2021-03-08) *(HRI '21)*. Association for Computing Machinery, 322–331. https://doi.org/10.1145/3434073.3444655

[19] Laxmi Pandey and Ahmed Sabbir Arif. [n. d.]. Effects of Speaking Rate on Speech and Silent Speech Recognition. In *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2022-04-27) *(CHI EA '22)*. Association for Computing Machinery, 1–8. https://doi.org/10.1145/3491101.3519611

[20] Archiki Prasad and Preethi Jyothi. [n. d.]. How Accents Confound: Probing for Accent Information in End-to-End Speech Recognition Systems. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (Online, 2020-07). Association for Computational Linguistics, 3739–3753. https://doi.org/10.18653/v1/2020.acl-main.345

[21] Elisa Prati, Margherita Peruzzini, Marcello Pellicciari, and Roberto Raffaeli. [n. d.]. How to include User eXperience in the design of Human-Robot Interaction. *Robotics and Computer-Integrated Manufacturing* 68 ([n. d.]), 102072. https://doi.org/10.1016/j.rcim.2020.102072

[22] Kacper Radzikowski, Mateusz Forc, Le Wang, Osamu Yoshie, and Robert Nowak. [n. d.]. Accent neutralization for speech recognition of non-native speakers. In *Proceedings of the 21st International Conference on Information Integration and Web-based Applications & Services* (New York, NY, USA, 2019-12-02) *(iiWAS2019)*. Association for Computing Machinery, 136–141. https://doi.org/10.1145/3366030.3366083

[23] Wenbi Rao, Ji Zhang, and Jianwei Wu. [n. d.]. Improved BLSTM RNN Based Accent Speech Recognition Using Multi-task Learning and Accent Embeddings. In *Proceedings of the 2020 2nd International Conference on Image, Video and Signal Processing* (New York, NY, USA, 2020-03-20) *(IVSP '20)*. Association for Computing Machinery, 1–6. https://doi.org/10.1145/3388818.3389159

[24] Sukyung Seok, Eunji Hwang, Jongsuk Choi, and Yoonseob Lim. [n. d.]. Cultural Differences in Indirect Speech Act Use and Politeness in Human-Robot Interaction. In *Proceedings of the 2022 ACM/IEEE International Conference on Human-Robot Interaction* (Sapporo, Hokkaido, Japan, 2022-03-07) *(HRI '22)*. IEEE Press, 470–477.

[25] Meishu Song, Adria Mallol-Ragolta, Emilia Parada-Cabaleiro, Zijiang Yang, Shuo Liu, Zhao Ren, Ziping Zhao, and Björn Schuller. [n. d.]. Frustration recognition from speech during game interaction using wide residual networks. *Virtual Reality & Intelligent Hardware* 3 ([n. d.]), 76–86. https://doi.org/10.1016/j.vrih.2020.10.004

[26] Alexandra Weidemann and Nele Rußwinkel. [n. d.]. The Role of Frustration in Human–Robot Interaction – What Is Needed for a Successful Collaboration? *Frontiers in Psychology* 12 ([n. d.]). https://www.frontiersin.org/articles/10.3389/fpsyg.2021.640186

[27] Xuesong Yang, Kartik Audhkhasi, Andrew Rosenberg, Samuel Thomas, Bhuvana Ramabhadran, and Mark Hasegawa-Johnson. [n. d.]. Joint Modeling of Accents and Acoustics for Multi-Accent Speech Recognition. ([n. d.]).

[28] Hui Ye and Steve Young. [n. d.]. Improving the speech recognition performance of beginners in spoken conversational interaction for language learning. 289–292. https://doi.org/10.21437/Interspeech.2005-160