# Inclusivity in Communication: Exploring Social Robots for Cultural Integration and Stuttering Therapy

by

Priyank Avijeet

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Applied Science
in
Electrical and Computer Engineering

Waterloo, Ontario, Canada, 2025

© Priyank Avijeet 2025

## Author's Declaration

This thesis consists of material all of which I authored or co-authored: see Statement of Contributions included in the thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

## Statement of Contributions

This thesis comprises first-authored, peer-reviewed material that has been accepted and published in the 33rd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN) 2024 and accepted for publication in the International Conference on Social Robotics (ICSR) 2024.

**Research presented in Chapter 3 (Study 1):**
Priyank Avijeet, Pourya Aliasghari, and Prof. Kerstin Dautenhahn designed the study. Priyank Avijeet implemented the study and developed the questionnaires on Qualtrics to collect the data. Pourya Aliasghari and Priyank Avijeet performed the data analysis. Priyank Avijeet drafted the manuscript, which was then reviewed and edited by Pourya Aliasghari and Prof. Kerstin Dautenhahn. Pourya Aliasghari mentored the study, and Prof. Kerstin Dautenhahn supervised it.

> *Citation:*
> Avijeet, Priyank, Pourya Aliasghari, and Kerstin Dautenhahn, "Greeting Preferences in a Hospitality Context: A Cross-Cultural Study with a Social Robot," *2024 33rd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, 2024, pp. 1786–1793.
> DOI: doi: 10.1109/RO-MAN60168.2024.10731433.

**Research presented in Chapter 4 (Study 2):**
Priyank Avijeet and Dr. Shruti Chandra initiated this study under the guidance of Prof. Kerstin Dautenhahn. They were primarily responsible for data collection. Prof. Kerstin Dautenhahn also participated in one of the interviews with the Speech and Language Pathologist. Prof. Smita Misra provided introductions and guidance on the Grounded Theory Methodology. Priyank Avijeet analyzed the data and collaborated with Dr. Shruti Chandra, Prof. Smita Misra, and Prof. Kerstin Dautenhahn to interpret the results. Priyank Avijeet drafted the manuscript, which was then reviewed and edited by Dr. Shruti Chandra, Prof. Smita Misra, and Prof. Kerstin Dautenhahn.

> *Citation:*
> Avijeet, Priyank, Shruti Chandra, and Smita Misra, "Design Considerations for Applications of Social Robots in the Stuttering Clinic," *Social Robotics: 16th International Conference, ICSR+ AI 2024, Odense, Denmark, October 23–26, 2024, Proceedings, Part II*, Springer Nature, 2025, p. 54.

**Research presented in Chapter 5 (Study 3):**
Priyank Avijeet designed and implemented the study under the guidance of Prof. Kerstin Dautenhahn. Priyank Avijeet implemented the study and developed the questionnaires on Qualtrics to collect the data. Priyank Avijeet also ran the study and conducted data cleaning and data analysis. Prof. Ewen MacDonald guided Priyank Avijeet in effective data collection and analysis of speech samples from participants. Prof. Ewen MacDonald lent Priyank Avijeet the equipment to collect clear audio recordings and taught him how to use them. Priyank Avijeet will draft this work for submission to a journal publication with Priyank Avijeet, Prof. Ewen MacDonald, and Prof. Kerstin Dautenhahn as co-authors.

## Abstract

Communication is fundamental for humans to exchange ideas, interact, and collaborate. These factors make it important to explore human-robot interactions through the lens of effective communication. This work explores inclusivity from two key dimensions: cultural integration and speech impairments.

Human-robot interaction (HRI) research has increasingly recognised the influence of culture on human-robot interactions, highlighting both the opportunities it brings and the need for careful consideration due to the evolving nature of cultural dynamics. In an online study with 103 participants, we explored how preferences for cross-cultural greetings performed by a humanoid robot change based on a restaurant theme. We examined factors influencing these preferences by analyzing two groups who experienced different ethnic greetings. We studied how ethnicity, percentage of life lived in Western countries, personality characteristics, and implementation of cultural aspects influenced the likability of a robot's greeting gestures.

In the context of speech impairments, the use of social robots in clinical settings for stuttering therapy of preschoolers by Speech-Language Pathologists (SLPs) remains a relatively unexplored area. To help HRI researchers explore this field, we identified potential applications of social robots for stuttering therapy. We discussed key considerations for these interactions by conducting interviews and shadowing sessions with two SLPs. The results suggest practical applications and interaction and communication design considerations for integrating social robots into speech therapy sessions for preschoolers who stutter. Key applications for social robots include engaging children through various games and providing a consistent speech model, among others. Essential properties of interaction design for sessions include being interactive, structured, and supervised by SLPs. Additionally, the robot's communication could involve slow, relaxed, personalised speech and other techniques to support children's speech goals.

Following recommendations from a handbook on stuttering and SLPs, we explored an application area focused on slow and relaxed speech that is to be used by parents and caregivers of children who stutter. As a first step and proof-of-concept, I carried out an HRI study with university students. I designed two sessions with four and five tasks, collecting natural speech samples from 63 participants across different verbal tasks. Throughout the interactions, the voice system (two conditions: the robot speaking or the same voice coming from a laptop) maintained a slow speech. I report our results on which kind of tasks had more impact on participants' speech and the difference in participants' speech metrics in the two sessions. People with different English language-related self-reported factors also

showed varied patterns in adapting to the Voice System's slow speech, showcasing the need for customised implementations in the future.

This work contributes to HRI by highlighting the challenges of developing culturally adaptive robots and providing design considerations for using social robots in stuttering therapy for preschoolers. By studying speech entrainment through a Voice System exhibiting slow speech, we explored how different verbal tasks influence entrainment levels, with potential future applications in training parents and caregivers of children who stutter.

**Acknowledgements**

I want to express my sincere gratitude to my supervisor, Professor Kerstin Dautenhahn. She has provided me with guidance and suggestions that were rooted in her decades of experience, helping me shape my raw ideas into practical study designs. Her support has played an instrumental role in enabling me to conduct my research and complete this thesis. Her belief in her students goes a long way in helping them develop problem-solving skills throughout their academic journeys.

I would also like to thank Pourya Aliasghari, Dr. Shruti Chandra, Professor Smita Misra, and Professor Ewen MacDonald, who have been my mentors and collaborators in the three studies documented in this thesis. Their input has been pivotal in implementing the studies and analyzing the results. I would also like to thank Speech and Language Pathologists Jean Macleod and Jill Clermont for allowing us to shadow speech therapy sessions and for permitting us to interview them.

My learning during my master's journey would have been incomplete without my supportive SIRRL labmates and Professor Chrystopher L. Nehaniv. Their feedback during research group meetings has helped me avoid mistakes in my study design or analysis that I might have otherwise overlooked. Thank you to our current and former lab members—Pourya Aliasghari, Ali Yamini, Keith Robello, Soomin Shin, Ali Noormohammadi and Kevin Fan—who were there when I needed to discuss the practical challenges I faced while implementing my studies.

In the end, I cannot forget to mention my friends who became family in Canada. Jagarti Meena, Vinamra Singh, Vinayak Gajendra Panchal, Yashvardhan Kukreja, Ayan Dogra, and Ishrat Patel have supported me whenever I faced challenges that life threw at me over the past two years. They made me feel that I was not alone in a land far from my home country. Krishanu Takhtani, Shubham Goel, Satyam Sinha, Nishita Hegde, and Simran Taneja, my friends from my undergraduate program in India, regularly motivated me through their phone calls.

Finally, I would like to acknowledge the funding from the Canada 150 Research Chairs Program, under which all my projects were supported. Thank you to all the participants who took part in my studies.

## Dedication

I dedicate this work to my parents, Mr. Ajay Kumar Srivastava and Mrs. Vinita Srivastava, and my brother, Saumya Avijeet, who have always believed that I am capable of delivering results and have supported me every step of the way. They are the reason behind my goals and the people I want to make proud.

# Table of Contents

**Author's Declaration** ii

**Statement of Contributions** iii

**Abstract** v

**Acknowledgements** vii

**Dedication** viii

**List of Figures** xiii

**List of Tables** xvi

**1 Introduction** 1

1.1 Motivation . . . . . 2

1.1.1 Cultural Integration Considerations for Social Robot in Hospitality 2

1.1.2 The Role of Social Robots in Speech Therapy . . . . . 3

1.2 Research Goals . . . . . 4

1.3 Summary of Contributions . . . . . 6

1.4 Thesis Overview . . . . . 8

1.5 Author's Contributions . . . . . 8

**2 Background Literature** **10**

2.1 Cultural adaptation and its challenges in HRI . . . 11

2.1.1 Cultural integration in HRI . . . 11

2.1.2 Culturally-adaptive social robots . . . 11

2.1.3 Challenges with cultural integration in HRI . . . 12

2.2 Applications of Social Robots in Speech Therapy and Grounded Theory in HRI . . . 12

2.2.1 Speech Therapy and SARs: . . . 13

2.2.2 Understanding GT in Technical Context: . . . 13

2.2.3 GT analysis and SARs: . . . 14

2.3 Speech Entrainment in HCI and Fluency Shaping Techniques in Speech Therapy . . . 14

2.3.1 Speech Entrainment studies in HCI . . . 14

2.3.2 Speech Therapy and Fluency Shaping Techniques . . . 16

**3 Study 1: Greeting Preferences in a Hospitality Context: A Cross-Cultural Study with a Social Robot** **18**

3.1 Research Questions . . . 19

3.2 Method and Implementation . . . 19

3.2.1 Artifacts . . . 20

3.2.2 Procedures and Measures . . . 23

3.2.3 Participants . . . 25

3.2.4 Statistical Analysis of Questionnaires . . . 25

3.3 Results . . . 30

3.3.1 Likability of Robot Greeting in Scenario 1 . . . 30

3.3.2 Likability of robot greeting in Scenario 2 . . . 32

3.3.3 Thematic Analysis of Open-ended Questions . . . 33

3.4 Discussion . . . 38

3.5 Limitations and Future Work . . . 40

**4 Study 2: Design Considerations for Applications of Social Robots in the Stuttering Clinic** **41**

4.1 Research Objective . . . 42

4.2 Methodology . . . 43

4.2.1 Grounded Theory Methodology . . . 43

4.3 Results . . . 50

4.3.1 Core Category 1: Practical Applications of Social Robots: . . . 50

4.3.2 Core Category 2: Interaction Design Considerations: . . . 51

4.3.3 Core Category 3: Communication Design Considerations: . . . 52

4.4 Discussion . . . 52

4.4.1 Practical Applications of a Social Robot . . . 54

4.4.2 Interaction Design Considerations . . . 55

4.4.3 Communication Design Considerations . . . 56

4.5 Limitations and Future Work . . . 57

**5 Study 3: Investigating Human Speech Rate Alignment in Response to Slow Robot Speech** **59**

5.1 Research Questions . . . 60

5.2 Method and Implementation . . . 61

5.2.1 Artifacts . . . 62

5.2.2 Procedures and Measures . . . 63

5.2.3 Participants . . . 76

5.2.4 Data Processing and Analysis . . . 76

5.3 Results . . . 81

5.3.1 Articulation Rate Models . . . 87

5.3.2 Pause Rate Models . . . 89

5.3.3 Pause Ratio Models . . . 91

5.3.4 Effect of Session on Speech Metrics . . . 93

5.3.5 Self-Assessment Manikins (SAM) . . . 96
5.3.6 Experience with the Voice System . . . 97
5.3.7 Thematic Analysis of Interview Questions . . . 100
5.4 Discussion . . . 105
5.4.1 Effect on Participants' Speech Metrics . . . 106
5.4.2 Effect on Participants' Emotional States . . . 108
5.4.3 Perceptions of the Voice System . . . 109
5.5 Limitations and Future Work . . . 109

**6 Conclusion** **112**
6.1 Summary of Findings . . . 112
6.1.1 Findings of Study 1 . . . 112
6.1.2 Findings of Study 2 . . . 113
6.1.3 Findings of Study 3 . . . 114
6.2 Contribution to Knowledge . . . 115
6.3 Limitations and Future Work . . . 117

**References** **119**

**APPENDICES** **132**

**A Ethics Clearance Certificates** **133**

**B Study 1 - Study Materials (Flyer, Web Interface and Questionnaires)** **140**

**C Study 1 - Extra Plots and Figures** **177**

**D Study 3 - Study Materials (Flyer, Scripts, Web Interface and Questionnaires)** **180**

**E Study 3 - Extra Plots and Figures** **262**

# List of Figures


1.1 Relationship between the studies and the central theme of Inclusivity in Communication. . . . . 2

3.1 Scenarios and different greeting categories studied within the scenarios. The order of the scenarios and greeting categories was randomized. . . . . 20

3.2 Pepper robot being recorded in front of a green screen . . . . . 21

3.3 Snapshots of the Pepper robot performing various welcoming greeting gestures: (a) Huān yíng guāng lín in a Chinese restaurant (b) As-salāmu alaykum in an Arabic restaurant, and (c) Hello in a Chinese restaurant. . . 22

3.4 User interface with embedded video and GodSpeed Likeability questionnaire 26

3.5 Histogram and density plots for Overall Ratings for HY and ASA in both scenarios . . . . . 27

3.6 Q-Q plots for Overall Ratings for HY and ASA in both scenarios . . . . . 28

3.7 Participants' overall ratings by groups for both scenarios. *: $p < .05$. . . . 31

4.1 Data Collection and Analysis Stages. The image highlights the first phase ("first diamond") of the Double Diamond Process Model [46], as applied in this study. This work is informed by prior research ( [28, 29, 64]) and the Handbook on stuttering ( [62]). . . . . 43

4.2 Core categories identified in our GT Analysis . . . . . 51

4.3 Categories, subcategories, and properties identified through GT Analysis. . 53

5.1 Images of some of the equipment used in the study . . . . . 64

5.2 The setup for Session 2 in Robot Condition. The setup for Session 1 was the same, except there were no objects on the table on the right. . . . . 65

5.3 The Self-Assessment Manikin - valence (top panel), arousal (middle panel), and control (bottom panel). . . . . 67
5.4 Web User Interface for Session 1 Task 1 . . . . . 69
5.5 Sample pictures shown for Session 1 and Session 2 on themes food and music. 71
5.6 Age distribution of Participants by Condition . . . . . 77
5.7 Histogram and Density Plot for Articulation Rates, Pause Rates and Pause Ratios of Participants . . . . . 82
5.8 Histogram and Density Plot for Overall Likability, Involvement with the Voice System, Perceived Intelligence of the Voice System and Attention Towards the Voice System . . . . . 83
5.9 Bar Plots for Control, Arousal, and Valence Scores for both Pre- and Post-Session . . . . . 84
5.10 Q-Q Plots for Articulation Rate, Pause Rate, Pause Ratio, and Perceived Characteristics of the Voice System . . . . . 85
5.11 Participant count by English Accent Category . . . . . 86
5.12 Participant count by English native language status . . . . . 86
5.13 Participants' articulation rates across tasks for both sessions. Significance levels: *: $p < .05$, **: $p < .01$, ***: $p < .001$, .: $p < .1$. . . . . 88
5.14 Participants' pause rates across tasks for both sessions. Significance levels: *: $p < .05$, **: $p < .01$, ***: $p < .001$, .: $p < .1$. . . . . 90
5.15 Participants' pause ratios across tasks for both sessions. Significance levels: *: $p < .05$, **: $p < .01$, ***: $p < .001$, .: $p < .1$. . . . . 92
5.16 Participants experience with the Voice System in terms of Overall Likability, Involvement, Intelligence, and Attention. Significance levels: *: $p < .05$ . . 99

C.1 Residuals Assumption Check for HY Scenario 1 . . . . . 178
C.2 Residuals Assumption Check for HY Scenario 2 . . . . . 178
C.3 Residuals Assumption Check for ASA Scenario 1 . . . . . 179
C.4 Residuals Assumption Check for ASA Scenario 2 . . . . . 179

E.1 Residuals Assumption Check for Articulation Rate in Session 1 . . . . . 263

E.2 Residuals Assumption Check for Pause Rate in Session 1 . . . . . 263
E.3 Residuals Assumption Check for Pause Ratio in Session 1 . . . . . 264
E.4 Residuals Assumption Check for Articulation Rate in Session 2 . . . . . 264
E.5 Residuals Assumption Check for Pause Rate in Session 2 . . . . . 265
E.6 Residuals Assumption Check for Pause Ratio in Session 2 . . . . . 265
E.7 Residuals Assumption Check for Attention Paid to the Voice System . . . 266
E.8 Residuals Assumption Check for Perceived Intelligence of the Voice System 266
E.9 Residuals Assumption Check for Involvement with the Voice System . . . . 267
E.10 Residuals Assumption Check for Overall Likability . . . . . 267

# List of Tables


3.1 Results of the Shapiro-Wilk normality test for overall rating distributions. . 29
3.2 Results of the Shapiro-Wilk normality test for residuals of final models. . . 30
3.3 LMMs predicting participants' overall ratings in Scenario 1 . . . . . 32
3.4 LMMs predicting participants' overall ratings in Scenario 2 . . . . . 33
3.5 Greeting Preferences in Scenario 1 . . . . . 35
3.6 Greeting Preferences in Scenario 2 . . . . . 37

4.1 Description of a Subset of Games Observed During Shadowing . . . . . 47

5.1 Ten-Item Personality Inventory (TIPI) . . . . . 66
5.2 LMMs predicting articulation rate in Session 1 . . . . . 87
5.3 LMMs predicting articulation rate in Session 2 . . . . . 89
5.4 LMMs predicting pause rate in Session 1 . . . . . 91
5.5 LMMs predicting pause rate in Session 2 . . . . . 91
5.6 LMMs predicting pause ratio in Session 1 . . . . . 93
5.7 LMMs predicting pause ratio in Session 2 . . . . . 94
5.8 LMMs predicting articulation rate, pause rate, and pause ratio across all sessions and tasks . . . . . 95
5.9 CLMM predicting post-session SAM . . . . . 98
5.10 LMs predicting overall likability, involvement, intelligence, and attention . 100
5.11 Themes for the Voice System speech speed Preferences . . . . . 101

5.12 Themes for perceived usefulness of slow speech . . . . . 103
5.13 Themes for impact of the Voice System on user speech . . . . . 104

# Chapter 1

# Introduction

Effective communication plays a fundamental role in building relationships within society. To ensure meaningful interactions, it is essential to promote inclusivity in communication from both cultural and speech-impairment perspectives. However, individuals may sometimes experience social isolation due to cultural differences [44, 49] or speech impairments [61, 64], which can create challenges in communication and self-expression. By embracing diverse communication needs, we can foster a more connected and understanding society. In such cases, social robots can constructively bridge communication gaps and support interaction in both contexts - Cultural Integration and Speech Therapy. Cultural factors significantly influence the acceptance of robots and the level of discomfort experienced by individuals [118]. A small yet important aspect of communication, greetings, can enhance the positive impact of a humanoid robot when it is culturally aligned with the participants [110]. In the context of therapy, recent studies have increasingly explored the application of Socially Assistive Robots (SARs) for therapeutic interventions (e.g. [4, 47, 98]). One notable use case for further research is in speech therapy sessions. By providing interactive and supportive environments in pediatric care and speech therapy, social robots can help facilitate social interactions, cognitive engagement, and emotional well-being, making therapy more effective and engaging for children [5, 51, 80].

This thesis presents three studies: one exploring cultural greeting preferences of participants in different location contexts and the other two exploring how social robots can assist in the stuttering therapy of preschoolers. An visualisation of how the three studies are related to the central theme of Inclusivity in Communication can be seen in Figure 1.1. Both areas of focus in this thesis, cultural integration and assistance in stuttering therapy, were chosen to add insights to the field of human-robot interaction (HRI) to support

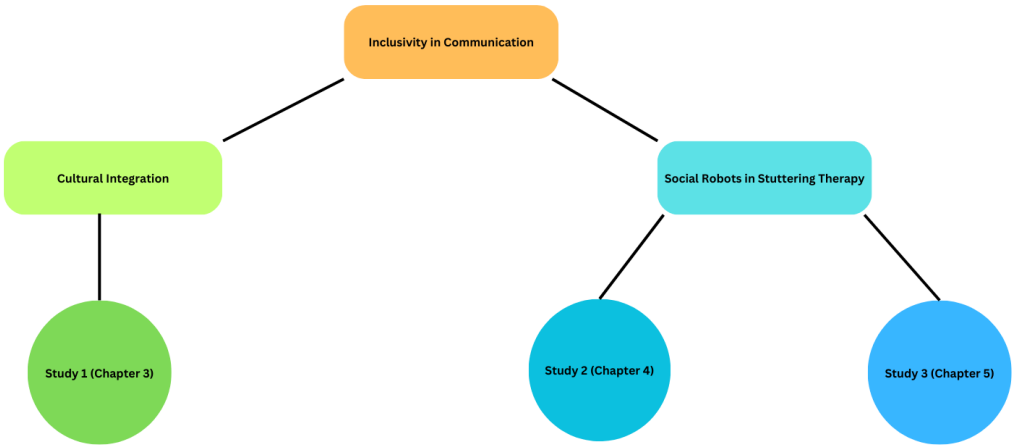



Figure 1.1: Relationship between the studies and the central theme of Inclusivity in Communication.

the overarching goal of enhancing inclusivity in communication through the use of social robots.

## 1.1 Motivation

In the following subsection, I will outline my motivations for exploring expectations and challenges regarding cultural integration through the use of social robots in restaurant settings. In the next subsection, I will discuss how social robots can be viable in stuttering therapy and modelling human speech to adopt fluency-shaping strategies.

### 1.1.1 Cultural Integration Considerations for Social Robot in Hospitality

Given the fluid nature of culture, it is important for robots to adaptively participate in social interactions. Adaptability could make room for cultural patterns to emerge directly from robot's interactions with humans. Research in cultural robotics has shifted from attempting to encode robots with static cultural norms to enhancing their fundamental skills in perception, learning, and adaptation of patterns of interaction within a group [90]. Such an approach empowers robots to independently adapt to the varied and continuously

evolving cultural contexts they encounter [90]. This raises concerns about the autonomy of individuals in shaping their interaction preferences, particularly within the sensitive context of cultural norms and expectations. Hospitality industry such as restaurants and hotels provide us with the opportunity to design our study around the cultural greetings.

The hospitality industry over the years has experienced an increase in the deployment of social robots [69, 119]. It has been suggested that customers' inclination towards social robots stems from their perceptions of value, empathy, and information sharing [45]. The effectiveness of social robots has been studied in the hospitality and service industry, specifically in roles such as serving as information desk assistants [88, 103, 125], information broadcasting [91, 92], acting as bellboys [95] and collecting customer feedback [36] [87].

Social robots are less likely to be perceived as odd in hospitality settings since they have been studied and tested in those environments. Also, with the flexibility that restaurants or hotels offer to adapt themes representing different cultures, we designed our study to explore the relationship between the background context of a restaurant setting and cultural greeting preferences. Greetings were chosen for our exploration as they are commonly observed in the context of welcoming guests in a restaurant setting. Greetings also allow for the study of culture in short interactions, as greetings vary across most cultures.

### 1.1.2 The Role of Social Robots in Speech Therapy

Incorporating social robots into stuttering therapy has been suggested through their roles as engaging and motivational partners, consistent speech models, reinforcement for social and speech skills, homework partners, and transition aides for different phases of treatment, such as establishment, transfer and maintenance [29]. However, research on the integration of social robots in stuttering clinics, specifically considering the child-robot interaction design aspects for the treatment of preschoolers, remains underexplored. Existing studies predominantly reflect the perspectives of HRI researchers and often lack the involvement of key stakeholders in the solution design phase [27, 100, 123]. This direction may lead to a disregard for the practicalities and real-world challenges that Speech Language Pathologists (SLPs) may encounter in clinical settings. This inspired us to explore and report on the potential applications of social robots in stuttering therapy and how to implement them by involving SLPs in the solution design phase.

I then identified an application of social robots where they can help influence a human's speaking patterns in a way that can help in the speech therapy outcomes of preschoolers who stutter. Speech Entrainment (SE), a phenomenon where individuals' speech patterns

such as pitch, rhythm, intonation, and speech rate become more similar during verbal interaction with another person [17] can be leveraged to affect human speech. In social psychology, SE is often explained through Communication Accommodation Theory (CAT), which has been developed and extensively advanced over decades by H. Giles and colleagues. CAT describes conscious and unconscious adjustments individuals make to align their communication with others, focusing primarily on convergence behaviours for alignment [54,55]. For subconscious adjustments, this alignment is commonly attributed to the chameleon effect [32] and the mirror mechanisms [99], where individuals automatically and unconsciously mimic their partner's behaviours, speech patterns, or postures that enhance social bonding through smoother interactions, increased empathy, and mutual liking [32].

Research on human-machine communication has examined how Voice-Based Assistants (VBAs) and social robots manage communication accommodation to influence user perceptions. In this context, perceived accommodation brings into discussion the relationship between warmth and competence, with overaccommodative VBAs receiving more favourable evaluations than underaccommodative ones [41]. Similarly, social robots establish rapport through social dialogue and prosodic entrainment, where pitch adaptation enhances perceptions of social presence [81].

In the study discussed in Chapter 4, I noted that a slow and relaxed speech for the caregivers of preschoolers who stutter is suggested by SLPs, and it is also supported by the suggestions in a handbook on stuttering [62]. SE and CAT inspire us to explore the applications of social robots in modelling expected speech patterns in humans by leveraging SE and CAT to support speech therapy either directly or indirectly.

## 1.2 Research Goals

In this section, I will discuss the research goals of my three studies. The three studies have similar broad goals to contribute towards inclusivity in communication HRI, but they are also unique in the way they provide these solutions. Detailed research questions can be found in the respective Chapters 3, 4 and 5 of these studies.

Our first study (discussed in Chapter 3) examined how interaction context affects the perception of cultural adaptation in two culturally diverse groups. Here, we focus on the first and basic form of interaction with a robot: greeting. We assume that a robot in the hospitality sector might already know a person's cultural or ethnic background, through the guest's prior declared preferences, e.g. at the time of booking a restaurant table. Yet, there is an underlying uncertainty regarding the type of robot greeting the user may

expect. In particular, in a restaurant with a specific cultural theme, how should a robot greet guests who may be accustomed to their own ethnic greetings[1]? Should the robot use a widely recognised, 'popular' greeting that adheres to Western norms, a greeting that aligns with the customer's ethnic background, or one that complements the restaurant's ethnic theme? Our study focused on participants who use ethnic greetings distinct from Western norms and consistent within their communities. Centering on ethnic greetings as a key research aspect, we recruited people from Chinese and Muslim/Islamic groups, due to their significant presence at the University of Waterloo, cf. the University's equity survey results [3]. Our recruitment materials therefore specifically invited individuals who identified with the greetings "Huān Yíng" or "As-salāmu alaykum". We refer to these two groups of participants as **HY** and **ASA**, respectively. In the study of Shidujaman et al. [110], participants experienced more positive feelings when the robot used a greeting gesture that was common in their own ethnic community. We expanded upon this finding by examining how the preference for ethnic greetings changes by taking into account the context of the location where the robot is employed. To the best of our knowledge, the complexity of cultural integration in greeting scenarios in hospitality settings, analyzing how user preferences can differ from one cultural group to another, has not yet been studied.

Our research objective in our second study, presented in Chapter 4, was to explore the applications of social robots in stuttering clinics and propose design considerations for these applications intended for use by SLPs through a co-design approach. The study was an extension of previous work conducted in the Social and Intelligent Research Lab (SIRRL) [28, 29, 64], which initiated the research direction of exploring the feasibility of using social robots for stuttering interventions. Adopting a co-design approach, extensive research was conducted, focused on stuttering interventions and explored opportunities for social robots in stuttering clinics [28]. SIRRL members engaged in online interviews and focus groups with SLPs to assess the feasibility of leveraging social robots into stuttering therapy. A recent study [29], employed the Double Diamond Design Process Model [46] that comprises four phases: Discovery, Definition (see Figure 4.1), Development and Delivery. Building on this prior work, our current research strengthens our understanding of the problem space and completes the first diamond phase of the research. This research focuses on discovering and defining (first diamond) potential considerations for developing applications on social robots to be used by SLPs in their clinic in the future (second diamond). In line with the co-design approach we followed in prior work, the goal of the present work was to actively involve SLPs in the Definition phase (see Figure 4.1). By incorporating insights from SLPs we aim to develop design considerations that will facilitate the adoption of our solutions

[1]Note, if we use the term "ethnic" greeting in this paper we refer to the greeting that participants identified with at the point of recruitment.

in therapy practices.

Our third study (explained in Chapter 5) explores if and how the slow and relaxed speech of a Voice System application running either on a robot or a laptop can subconsciously influence human speech through speech entrainment, demonstrating the chameleon effect. Participants interact with a Voice System featuring a low articulation rate and frequent pauses across two in-person sessions involving four and five tasks, respectively. These tasks include repeating the sentences spoken by the Voice System, describing images displayed on the screen, and listening to the Voice System discuss its preferences in food and music. Participants' speech samples are collected at the start and end of each session to analyze articulation rates, pause rates, and pause ratios as primary speech metrics. The primary motivation behind this work was to explore whether a social robot could be employed to subtly influence human speech patterns through different kinds of verbal tasks. This approach holds potential for stuttering therapy in preschoolers, where caregivers are encouraged to adjust the child's environment by speaking in a slower and more relaxed manner [62]. In the research conducted in Chapter 4, an SLP also emphasised the challenge of altering one's speech patterns in practice.

## 1.3 Summary of Contributions

The overall contributions of the research projects in this thesis are summarised as follows:

1. For the first study, I conducted an online survey where participants watched videos in which the Pepper robot [2] greeted them using different cultural greetings in two culturally themed restaurants. Participants' likability towards the greetings was measured using the Likability subscale of the Godspeed [13] questionnaire. Note, an English language package was used consistently in all conditions to create the robot's voice. Our findings from the first study suggest that a restaurant's theme and location, alongside individuals' personality traits and ethnic background, the percentage of life lived in Western countries, and also how precisely cultural aspects are implemented in a robot might influence the likability of robot greeting gestures. For example, in a restaurant that resonated more closely with their own ethnic greetings, participants from the HY group showed a preference for a popular Western greeting, while the ASA participants favoured their own ethnic greeting. Both HY and ASA participants did not appear to have a strong preference for any particular greeting in a restaurant whose theme did not resonate with their respective ethnic greetings.

These findings point towards the ethical challenges that need to be kept in mind while incorporating cultural elements in HRI.

2. The second study involved an analysis of prior work published by members of the Social and Intelligent Robotics Research Lab (SIRRL) [2], shadowing of stuttering therapy sessions for preschoolers, followed by interviews with the SLPs, as well as using as a handbook on stuttering as an additional data source. The data collection and analysis adopted the Grounded Theory Methodology [31]. The research contributions of the second study include identifying practical applications of social robots in stuttering clinics, specifically as a toolbox containing a variety of games, storytelling, stuttering samples, and a consistent speech model. Additionally, the findings from observations in shadowing sessions of preschoolers who stutter, and through discussion with SLPs, bring to light key considerations for interaction design. The results emphasise interactions being play-based, low cognitive load, the inclusion of tangible props and turn-taking, and the achievement of speech goals. An SLP also suggested considering the integration of social robots in speech therapy, such that the absence of the social robot should not cause distractions or disappointment. We also noticed that certain communication design considerations emerged during a qualitative analysis of the findings. We learned that the robot should communicate in a personalised, enthusiastic manner, with prolonged, and syllable-timed speech. Cultural elements could be used if those elements are familiar to the child. Moreover, a natural and non-robotic speech style needs to be implemented for effective outcomes.

3. To address the research goals for the third study, I conducted an in-person study in the lab with two sessions for each participant, each session having four and five verbal tasks respectively. The tasks were designed to collect speech samples of the participants, which were analyzed later to detect any differences in speech metrics throughout the two sessions for each participant. Questionnaire data was also collected before the study, before and after each session, and after the study, along with a semi-structured interview at the end. Results from the third study inform us that more predictable and less open-ended tasks, such as sentence repetitions or describing a picture in three sentences, are capable of reducing articulation rates more than other tasks. We also learned that having more sessions can have a higher relative impact. When it comes to pause rates and pause ratios, a task where participants have to repeat phrases could be less beneficial if the goal is to increase these speech metrics. The findings of our study will not only help design specific applications that could aid in modelling speech patterns, but it also opens doors for exploring

[2] https://uwaterloo.ca/social-intelligent-robotics-research-lab/

certain fluency-shaping techniques. Techniques such as stretched syllable techniques, prolongation, or gentle initiation can be explored in future studies leveraging SE and CAT.

## 1.4 Thesis Overview

The remainder of this thesis has the following structure. Chapter 2 provides a literature review on topics relevant to the three studies, including cultural adaptation and its challenges in HRI, applications of social robots in speech therapy, applications of Grounded Theory in HRI, speech entrainment in Human-computer interaction (HCI) and fluency-shaping techniques in speech therapy. Chapter 3 presents our first study, which explores how locational contexts influence user preferences for cultural greetings. Chapter 4 describes the study on the application of social robots in stuttering therapy for preschoolers, aimed at establishing a foundation for future research in this domain. Chapter 5 presents our final study on slow speech entrainment in human participants in response to slow speech of a Voice System application running on a social robot or a laptop across different verbal tasks. The primary motivation behind this study is to find potential applications in modelling slow and relaxed speech in parents and caregivers of children who stutter. Finally, Chapter 6 summarises the findings of our studies, discusses how they address the research questions, highlights their contributions to the HRI knowledge space, and outlines the limitations of our approach, along with directions for future research.

Towards the end, there are also five appendices. Appendix A contains our ethics clearance certificates for Study 1 and Study 3. Appendices B, C, D and E provide study materials including web interfaces, questionnaires, scripts of conversations and some extra plots and figures that helped implement and analyze the studies.

## 1.5 Author's Contributions

This thesis consists of three research studies undertaken by me. The use of the term "we" in this thesis reflects the collaborative nature of these projects, which were carried out with the joint effort of a team—sometimes a multidisciplinary team of researchers—a common practice in human-robot interaction studies. The term "we" is also a common practice in academic writing. In regard to my contributions, I came up with the key ideas for the first and third studies. I led the design and development of the first and third studies from both conceptual and technical standpoints. The second study began with the ideas

suggested by Professor Dautenhan to visit SLPs at KidsAbility. Dr. Shruti Chandra led the initial shadowing and interviews with the SLPs, but after that, I followed up with shadowing sessions and interviews with the second SLP at KidsAbility. I came up with the idea to report our progress so far in the domain of applications of social robots in the stuttering therapy of preschoolers. Under the guidance of Professor Dautenhan, Pourya Aliasghari, Professor Smita Misra, Dr. Shruti Chandra and Professor Ewen MacDonald, I performed data cleaning and analysis for each study, followed by drafting and preparing the submissions for reporting the findings.

# Chapter 2

# Background Literature

This chapter will review relevant background literature related to our studies. The discussion focuses on two key themes of inclusivity in communication: cultural aspects and speech impairment, which are central to this thesis. The first part explores cultural integration, while the following sections broadly fall under the category of speech impairment.

Regarding our first study, it is important to understand the journey of cultural integration in HRI studies and how it has evolved over time. Advancements in technology and algorithms have enabled us to experiment with autonomous cultural adaptations during these interactions. Cultural integration and adaptations being sensitive, evolving, and context-dependent in nature raise certain challenges.

The second sub-theme — applications of social robots in speech therapy— discusses studies that have investigated the use of Social Assistive Robots (SARs) in speech therapy. As we adopted a variant of Grounded Theory (GT) for our second study, it was essential to refer to GT and its role in data collection and analysis in studies involving SARs. Since GT evolved in the discipline of sociology, we also summarise efforts undertaken to adapt GT in technical contexts.

Building on our application ideas in my third study, we will explore how speech entrainment has been studied in Human-Computer Interaction (HCI), the approaches researchers have taken, and the results they have achieved. We will connect these insights with existing literature to support the effectiveness of fluency-shaping strategies, such as slow and relaxed speech. We will also examine the impact of caregivers practicing these techniques with young children. By identifying gaps and evaluating the potential of social robots in leveraging speech entrainment to facilitate fluency-shaping strategies, we recognise the need to study an application in speech therapy that requires practice by adults (caregivers).

## 2.1 Cultural adaptation and its challenges in HRI

In this section, we first discuss cultural integration in HRI and how previous research considered different cultures from around the world in HRI. Next, we review studies on developing culturally adaptive social robots and then look into the challenges of cultural integration in HRI.

### 2.1.1 Cultural integration in HRI

The HRI community over time has progressively explored the impact of culture on participants' perception of robots by conducting studies across diverse communities. Initial studies predominantly focused on societies in the Western world and East Asia, reflecting the historical centers of developments in the HRI domain (e.g. Japan, USA, Europe), as evidenced in works such as [14, 52]. Later efforts have expanded to investigate the Middle Eastern population [8, 82], recognizing the significance of a more globally inclusive approach in HRI. In a research study conducted with participants from Egypt and Japan, it was discovered that culture played a significant role in determining the acceptance and discomfort levels associated with the greeting gestures of a humanoid robot [118]. Research on children from Pakistani and Dutch backgrounds interacting with the iCat robot revealed significant influences of age and culture, indicating Pakistani and younger children generally enjoyed the interactions more than Dutch and older children, respectively [108]. Another study compared German and Arab participants' preferences for robot design. The findings suggested that the German participants leaned towards practical, non-humanoid robots primarily for industrial applications, whereas Arab participants showed a stronger inclination towards robots that mimic human appearance and interaction, particularly for roles extending into healthcare [71]. Such findings emphasise the importance of accounting for cultural differences in social robot design for global applications.

### 2.1.2 Culturally-adaptive social robots

There has been a growing effort to explore the applications of Socially Assistive Robots (SAR) through a cultural lens, as evidenced by studies conducted under or built upon the CARESSES project [25]. CARESSES [107] is an initiative aimed at developing culturally competent robots capable of autonomously adapting their behaviours and communications to align with the personal and cultural preferences of the elderly they assist. This approach integrates principles of trans-cultural nursing to provide culturally sensitive care. A study

highlighting a 'diversity-aware' humanoid robot's acceptance among clinicians and spinal cord injury patients demonstrates the robot's ability to customise interactions according to the user's cultural background, age, gender, personal preferences, and health status, while also examining the factors influencing its adoption [59]. Research in this field has progressively become more detailed in addressing cultural nuances. Complex greeting gesture selection systems have been developed that take insights from sociology studies and incorporate contextual features such as culture, gender, location, power relationship, and social distance into an algorithm, facilitating the dynamic adaptation of gestures based on various contexts [117].

### 2.1.3 Challenges with cultural integration in HRI

The design of the cultural adaptation process in HRI can be influenced by a variety of factors. A recent study reviewed the use of social robots in educational settings from a cultural standpoint with a focus on the robot's role, the objectives and goals of the interaction, and the types of interaction involved. The study highlighted the differences in interaction preferences between individualistic and collectivist cultures, how power distance affected robot roles, and the importance placed on educational content by 'masculine' and 'feminine' societies, according to the authors [24]. Note, concerns have been raised regarding questions about the suitability of robots to understand or generate descriptions of people, emphasizing the need to avoid the risks of stereotyping or misrepresentation in representing a culture [120].

## 2.2 Applications of Social Robots in Speech Therapy and Grounded Theory in HRI

In this section, we will review studies that have developed applications for SARs in speech therapy and evaluated their effectiveness. This review will help us fully understand the field and identify potential research gaps. We will then examine studies that utilise the GT approach in HRI research for practical applications. This has also guided us in refining our approach. We aim to understand subtle aspects about the application of GTM in HRI that may have been overlooked in previous studies.

### 2.2.1 Speech Therapy and SARs:

Several studies have addressed the use of social robots in speech therapy, although without any mention of involving stakeholders in the solution design phase [27, 100, 123]. Applications involved mouth pose detection to guide speech exercises for Apraxia [27]. Another study, for children with cerebral palsy, showed significant improvements in phonological, morphosyntactical, and semantic aspects of speech [100]. Yoshikawa et al. [123] explored the use of teleoperated robots to reduce stuttering-related psychological symptoms, finding improvements in speech-related attitudes. Despite these promising results, the reliance on teleoperation poses practical challenges. In [113], with tasks designed by psycholinguists and conducted by speech therapists, SARs showed the potential to improve children's language skills through higher engagement. A recent study [76] involved the co-creation of three therapeutic educational games using the Kaspar robot, in collaboration with experienced Speech, Language, and Communication (SLC) experts. Their intervention with 20 children from special educational needs schools over 9 sessions showed significant improvements in SLC skills. However, to the best of our knowledge, no research has been done on design considerations for HRI researchers, particularly those based on observations and discussions with SLPs in preschool stuttering clinics. Our second study addresses this gap by offering practical design insights, providing a foundation for researchers to build upon in using social robots for preschooler stuttering therapy.

### 2.2.2 Understanding GT in Technical Context:

GT, while less popular than Thematic Analysis in qualitative HRI research, remains a valuable and applicable methodology, but not without challenges and adaptations. The authors in [111] discuss the methodological challenges of transitioning from GT to user requirements for designing information and communication technologies. Translating qualitative data into practical design solutions is understood to be a fairly complicated task. Certain methodological adjustments were suggested by the authors, such as incorporating preliminary scenarios before interviews and integrating the exploration of user requirements within open-ended interviews. The proposed shift in epistemology and methodology bridges theory and practice, combining sociological techniques with design needs to make qualitative research both rigorous and useful for technological development. Similarly, in software engineering (SE), GT is adapted to study socio-technical phenomena, where human and technical aspects are closely connected. Traditional GT, originally developed for sociological research, often poses challenges for SE researchers due to its focus on purely social phenomena and lack of clarity for socio-technical applications. This study presented

Socio-Technical Grounded Theory (STGT) as a way to address the challenges faced by software engineering researchers when applying traditional GT. It offers flexibility in theory development modes and emphasises the importance of interdisciplinary research teams. STGT combines knowledge from social sciences, qualitative research, and technical backgrounds. STGT aims to make GT more accessible and effective for SE researchers, enabling high-quality studies that produce practical and relevant theories in SE [65].

### 2.2.3 GT analysis and SARs:

GT has been used to explore practical aspects similar to our goal but in different domains such as deploying robot teleoperation in therapy [48] and to identify desired features of assistive robots for older adults [18]. Analysis of interview data from seven older adults revealed that while they recognised the benefits of assistive robots, stereotypes about aging embedded in robot design had impeded the robot's acceptance [78]. SARs improved engagement and care quality in aged-care facilities [70] although other studies with healthcare professionals noted limitations in providing comprehensive humanistic care [112]. In [72], GT was applied to analyze observational data of interactions in dementia care, involving patients, relatives, and care providers. Note, none of these works explain the reasoning behind their choice of a particular variation of GT (Glaser/Classic or Charmaz/Constructivist or Strauss, differences among variations summarised in [39]) over other variations and its impact on their research.

## 2.3 Speech Entrainment in HCI and Fluency Shaping Techniques in Speech Therapy

HCI has extensively studied speech entrainment. For our third study, it was important for us to understand their data collection and analysis approaches for inspiration. We were particularly interested in fluency-shaping strategies that caregivers initiate and practice to promote smooth speech in young children. This was especially relevant as we planned to design a study with adult participants.

### 2.3.1 Speech Entrainment studies in HCI

In their exploratory works M. Cohn et al. [37, 38], explored the impact of physical embodiment and gender on vocal accommodation by conducting experiments to examine

how participants adjust their speech when interacting with identical TTS (text-to-speech) voices presented in three distinct physical forms: an Amazon Echo smart speaker (the least human-like, with a simple cylindrical design), a Nao robot (moderately human-like, featuring a face and body), and a Furhat robot (the most human-like, resembling a human head with animated facial expressions). In [38], the participants interacted with pre-recorded videos of three physical forms (Amazon Echo [6], Nao robot [102], and Furhat robot [101]). Each spoke with the same computer-generated voices (male: "Matthew", female: "Salli") using AWS Polly [7], a cloud-based TTS service that converts text into natural-sounding speech. These voices were used for Amazon Echo, Nao, and Furhat. The experiment consisted of a baseline production block (a structured phase in the study), where participants pronounced each of the 11 target words three times, followed by a shadowing task in which they repeated words spoken by the TTS voice across three separate blocks (one per physical form). Vocal accommodation was assessed using an AXB perceptual similarity rating task. **A** (the participant's pronunciation before exposure to the TTS voice), **X** (the TTS model's pronunciation), and **B** (the participant's pronunciation after exposure). Their task was to determine whether **A** or **B** sounded more like **X**. The independent raters determined whether participants' post-exposure pronunciation more closely resembled the TTS voice than their pre-exposure pronunciation, providing a global measure of speech adaptation rather than analyzing individual acoustic features (e.g., pitch contour, vowel articulation, speech rate, consonant duration, etc.) [38]. A similar study [126] examined how social and generational factors influence speech adaptation in both human-human and human-AI interactions. It investigated speech alignment across age (older adults: 53–81 years, younger adults: 18–39 years) and gender using a word shadowing task with human and voice-AI (Apple's Siri [9]) voices (one male, one female for each category). Alignment was measured via an AXB perceptual similarity rating task, where independent raters determined whether participants' post-exposure pronunciation more closely resembled the model talker than their pre-exposure pronunciation. It is to be noted that this study focused on the word shadowing task. In our explorations, we will explore some open-ended tasks as well.

With the long-term goal of advancing adaptive conversational interfaces, [42] explores how elementary-school children adapt their response latencies when interacting with animated characters using extrovert and introvert TTS voices, examining the influence of child-age, gender, and TTS voice gender on conversational adaptation. Results indicate that children dynamically adjust their response timing by lengthening their conversational latencies when interacting with slower-paced, introverted TTS voices and shortening them when interacting with faster-paced, extroverted TTS voices. Another study [26] suggests that speech accommodation occurs even in the absence of interactional goals, as partici-

pants in a virtual reality environment adjusted their speech rate to match that of a virtual agent, VIRTUO, despite its inability to understand or respond flexibly to their linguistic behaviour. In a different context, lexical entrainment was observed in [67], where people and robots come to use the same referential terms when repeatedly discussing the same object, with faster entrainment for distant or unfamiliar objects, which helped improve object recognition.

### 2.3.2 Speech Therapy and Fluency Shaping Techniques

Stuttering in children is related to a lack of language facility, articulatory proficiency, harmony, and structure in the environment [12]. Research suggests that modifying a child's surroundings can promote fluency, as highlighted in various studies and discussed in a stuttering therapy handbook [62]. A goal-directed approach utilises home and school resources to facilitate fluency with key strategies such as slow speech, simple language, frequent pausing, indirect fluency management, and encouragement. Effective intervention requires a coordinated effort among parents, teachers, and clinicians, combining environmental management and fluency facilitation to support speech development [12]. Additionally, a family-focused treatment approach for preschool children who stutter, emphasises parent education, communication modifications, and supportive responses to improve fluency. A preliminary evaluation showed significant fluency improvements, with most children benefiting from parent-focused strategies alone and parents reporting high satisfaction with the treatment [122]. Among adults who stutter, speech restructuring methods—particularly prolonged speech and gentle onset techniques—are strongly supported by empirical research for improving fluency, while comprehensive programs also integrate cognitive and anxiolytic strategies to address speech-related anxiety and avoidance behaviours [20].

A multifaceted fluency intervention [58] for preschoolers emphasises the role of caregivers, educators, and daycare staff in modifying environmental demands through education, emotional support, and behavioural adjustments. Sawyer et al. [106] investigated a parent-focused slow relaxed speech intervention for preschoolers who stutter, demonstrating that caregivers could be trained to reduce articulation rates, with most children showing fluency improvements. However, no direct relationship was found between caregivers' speech rate and children's fluency gains, suggesting that language simplification, reduced vocalization length, minimised simultalk (overlapping speech between speakers), and increased pausing may also play a role. It was found that when mothers substantially slowed their speech rates (by  50%), five out of six typically developing 3-year-old children also significantly reduced their speech rates [63]. These findings indicate that caregivers'

speech modifications can influence children's speech patterns, supporting their potential role in speech therapy for fluency development.

Social robots can serve as assistive tools in fluency interventions, providing structured and interactive therapy that supports speech rate control, articulation, and social interaction to enhance fluency and reduce anxiety [74]. Further, a review [53] of the potential of socially assistive robots in speech and language therapy, emphasizing their structured, engaging, and AI-enhanced capabilities. Given the influence of caregiver speech patterns on children's fluency, social robots could supplement interventions by modelling slow, relaxed speech in a consistent and adaptive manner.

In the next chapter, we will discuss my study on greeting preferences among two different cultural groups recruited from the university population in a remote study. In this study, the two groups are greeted with different ethnic greetings in different locational contexts. This chapter will highlight the ethical challenges and careful considerations needed for cultural integration in HRI. When it comes to adaptations, our study brings about the need for a discussion on the balance of autonomy between humans and intelligent agents.

# Chapter 3

# Study 1: Greeting Preferences in a Hospitality Context: A Cross-Cultural Study with a Social Robot

In this chapter, I begin my journey of finding applications for social robots as tools to promote more inclusive communication. Through the study discussed in this chapter, I designed an online study to examine participants' preferences for cultural greetings in two different culturally themed restaurant backgrounds. The primary goal was to explore the impact of location and context that should be taken into account when implementing cultural adaptation systems in social robots. Using prerecorded videos of the Pepper robot with backgrounds representing Chinese and Arabic restaurants, and greetings associated with Chinese, Arabic, and popular Western culture, I gathered participants' preferences using the likability subscale of the Godspeed questionnaire [13]. The results of the study suggest that the accuracy of implementation, lack of alignment between the entire communication and the cultural themes, and participants' backgrounds can impact their preferences for cultural greetings. To recall, in this study, participants who identified with the greetings "Huān Yíng" or "As-salāmu alaykum" are referred to as **HY** and **ASA**, respectively.

## 3.1 Research Questions

This study aims to fill the gap in the current literature on cultural integration in HRI by exploring different robot greeting scenarios with participants from different backgrounds. We focus on answering two main research questions:

**RQ1**: Will individuals from HY and ASA groups prefer a Western greeting or their own ethnic greeting in a restaurant whose theme resonates with their ethnic greeting?

**RQ2**: In a restaurant whose theme does not resonate with their ethnic greeting or Western greeting, will people in HY and ASA groups prefer a Western greeting, one that matches the restaurant's theme, or their own ethnic greeting?

For **RQ1** we formulate a hypothesis based on [110]: Participants will prefer their ethnic greetings in a restaurant setting whose theme resonates with their own ethnic greeting **(H1)**. **RQ2** is exploratory in nature.

## 3.2 Method and Implementation

In this section, we provide an overview of our study, exploring how the theme of a restaurant might affect participants' preferences for different ethnic greetings. Conducted virtually using Qualtrics [96], participants viewed pre-recorded videos of Pepper, a humanoid robot (made by SoftBank robotics) [2] greeting them at two different ethnically-themed restaurants. After watching each video, participants completed a questionnaire about the likeability of the greetings. The study was structured to expose each participant to two main scenarios, as explained below, to gather insights into cross-cultural greeting preferences influenced by the restaurant's theme. For representation and analysis, the greeting categories are labelled as "Own" (greetings matching the participant's ethnic greeting), "Popular" (the popular Western "Hello"), or "Other" (greetings reflecting the restaurant's ethnic theme), cf. Figure 3.1.

**Scenario 1**: Participants were greeted by a robot in two ways in a restaurant that reflected a culture in which their ethnic greeting was used: (a) with their ethnic greeting (Own) gesture, and (b) with common Western greeting practices (Popular)

**Scenario 2**: Participants were greeted in three ways in a restaurant with a cultural theme that was neither Western nor reflected the culture in which participants' ethnic greeting was used: (a) with a greeting consistent with their ethnic greeting (Own), (b) with common Western greeting practices (Popular), and (c) with a greeting typical of the restaurant's cultural theme (Other).

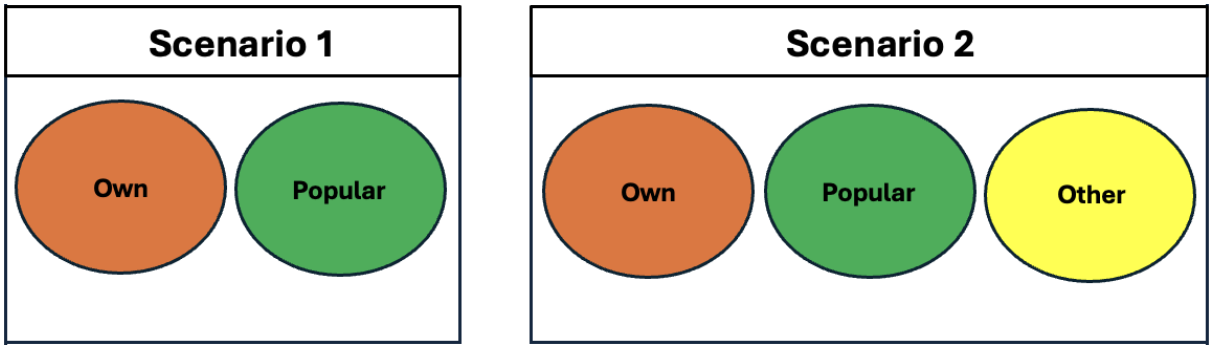



Figure 3.1: Scenarios and different greeting categories studied within the scenarios. The order of the scenarios and greeting categories was randomized.

There were two main reasons why we opted for a remote study. Firstly, the running time of the study was around 15 minutes and recruiting participants for this short duration for an in-person study in our lab did not appear to be the optimal choice. Secondly, creating a lab setting which resembled an authentic Chinese and Arabic restaurant was not feasible.

### 3.2.1 Artifacts

We used a Pepper robot due to its friendly and welcoming appearance which is desirable in the hospitality industry. We programmed the Pepper robot's greeting gestures using Choregraphe software [1] and recorded these gestures performed by the robot in front of a green screen in our lab. We took photos of an East Asian and a Middle Eastern restaurant in the Waterloo region with permission from the managers at the restaurants. These photos served as backgrounds in our videos. We replaced the green screen using Apple iMovie [1] in the videos as seen in Figure 3.2 and the outcome after video editing can be seen in the screen capture shown in Figure 3.3. These videos, each lasting around 27 seconds, were edited and uploaded to YouTube [2] for embedding in the Qualtrics survey platform along with questionnaires. We used the robot's English language package for all communication to keep the voice and language consistent.

[1] https://www.apple.com/ca/imovie/
[2] https://bit.ly/3wX1rbC




Figure 3.2: Pepper robot being recorded in front of a green screen



(a)



(b)



(c)

Figure 3.3: Snapshots of the Pepper robot performing various welcoming greeting gestures: (a) Huān yíng guāng lín in a Chinese restaurant (b) As-salāmu alaykum in an Arabic restaurant, and (c) Hello in a Chinese restaurant.

### 3.2.2 Procedures and Measures

The study was approved by the University of Waterloo Human Research Ethics Board. Participants were recruited from the University of Waterloo students. Note that the student population at the University of Waterloo is very diverse, with many international students, some of whom might only have recently moved to Canada. Recruitment strategies included posting flyers (see Appendix B) in specified campus areas, using university email lists, coordinating with departmental offices and the university's 'daily bulletin', and partnering with cultural student associations for the distribution of study materials and promotion via word of mouth. Participants interested in the study were sent an email containing a survey link. The survey began with a consent form, followed by a prompt asking participants to select a greeting that reflects their ethnic background from options "Huān Yíng", "As-salāmu alaykum" or "Other." If a participant chose "Other," the survey concluded, directing them to a page explaining the study's focus on participants who identify with one of those greetings. Participants who passed this step were then asked to complete demographic and the Ten Item Personality Inventory (TIPI) [57] to measure Big 5 personality traits. In the demographics section, participants were asked to share information on their Gender, Age, Ethnicity, Occupation, and Education level, entering their responses in the provided text boxes. The form also asked participants to list the countries in which they had resided for a minimum of six months, along with the duration of each stay. This specific data was intended to calculate the percentage of each participant's life that had been spent living in Western countries. Participants were informed that answering demographic questions was optional, and they could skip any they did not feel comfortable providing. We then assessed their closeness to their culture by asking them to rate their connection to both Western culture and their own culture on a bipolar continuous slider scale, with options ranging from "A little" to "A lot."

The study then proceeded to the main part. In a one-minute introductory video [2], the Pepper robot introduced itself and the experiment to the participants, to help the participants know what to expect in the study and get accustomed to the audio and visuals of the robot. As explained before, each participant went through the two scenarios with each greeting presented in a separate video.

**Scenario 1:** The participants in the HY group were welcomed in a Chinese restaurant with their ethnic greeting "Huān yíng guāng lín", and, separately, the popular "Hello" greeting. Similarly, the ASA participants were welcomed in an Arabic restaurant with their ethnic greeting "As-salāmu alaykum", and the popular "Hello" greeting.

**Scenario 2:** The participants in the HY group were welcomed in an Arabic restaurant with the ethnic greeting prevalent in Arabic culture, "As-salāmu alaykum". They were also,

in separate videos, greeted with their ethnic greeting "Huān yíng guāng lín" and the popular "Hello" greeting. Similarly, the ASA participants were welcomed in a Chinese restaurant in the ethnic greeting prevalent in Chinese culture, "Huān yíng guāng lín". They were greeted with their ethnic greeting "As-salāmu alaykum" and the popular "Hello" greeting.

All greetings were paired with culturally specific gestures ("Huān yíng guāng lín" with a bow and "As-salāmu alaykum" with a gentle bow and the right hand on the heart), performed by the Pepper robot. The cultural greetings were followed by the welcome message in English so that participants from both the HY and ASA groups could understand it. The English language package of the Pepper robot was used consistently across all study conditions for implementing the cultural greetings and the welcome message. To match the correct pronunciation of the cultural greetings, instead of writing the exact words for the greetings "Huān yíng guāng lín" and "As-salāmu alaykum", I broke them up into smaller chunks and worked with lab members who are familiar with these cultures to get the pronunciation and gestures close to the actual pronunciation of the cultural greetings. The order of greetings and scenarios was randomized via the Qualtrics portal's Randomizer feature, ensuring all permutations were evenly distributed to reduce order bias. Following each greeting video, participants' impressions were assessed using the Godspeed Questionnaire subscale for Likeability [13], which is a widely used measure in HRI. To evaluate participants' attentiveness during the online study, we included attention check questions following each likeability questionnaire. These questions asked participants to identify the type of restaurant where the Pepper robot greeted them, with four options. Participants could choose to replay the video, or proceed directly to the questionnaire. To maintain focus on the videos, the videos paused automatically if the participant navigated away from the active window or tab. An example of the user interface with the embedded YouTube video along with the Godspeed Likeability questionnaire and attention check question can be seen in Figure 3.4. To gain insight into the participants' thought processes behind their preference ratings, we asked two open-ended questions after they had watched all the videos concluded:

Question 1: "In a restaurant setting that matches your ethnic background (for example, a Chinese/Arabic-themed restaurant), would you prefer to be welcomed by a social robot that aligns with your ethnic background or follows popular Western culture customs, or do you have any other preferences? Please provide your choice and reasons to support your preference."

Question 2: "In a restaurant setting that does not match your ethnic background (for example, an Italian or Japanese-themed restaurant), would you prefer to be welcomed by a social robot that aligns with your ethnic background or follows popular Western culture customs, or do you have any other preferences? Please provide your choice and reasons to

support your preference."

At the end of the study, we thanked the participants for their participation. They received 6 CAD for dedicating around 15 minutes of their time to the study.

### 3.2.3 Participants

We advertised widely for several months across the university campus and communities. 103 participants took part in this study (HY = 66 and ASA = 37). To ensure data quality, participants who incorrectly answered at least three out of five attention check questions were excluded. Participants with exactly two mistakes were also excluded, except for those who had confused between Chinese and Arabic restaurants in one of the attention checks. The answer to all attention check questions was either "Chinese" or "Arabic" restaurant. In order to be considerate about confusing between the two options, we allowed participants who got two attention checks wrong, provided that at least one of the incorrect responses was either "Chinese" or "Arabic" restaurant.

After the filtering process, 94 participants remained: 58 HY and 36 ASA. For the HY participants, there were 26 males, 29 females, and 3 who did not indicate their gender. Their age was in [18, 30] with $M_{\text{age}} = 21.8$ and SD = 2.7. In the ASA group, there were 19 males and 17 females, age [18, 36] with $M_{\text{age}} = 23.51$ and SD = 5.09.

### 3.2.4 Statistical Analysis of Questionnaires

After watching each video, participants answered the Likeability subscale of the Godspeed Questionnaire to rate the observed robot greeting. We gathered their responses on continuous scales ranging from 0 (negative attributes) to 100 (positive attributes) such as *Dislike – Like, Unfriendly – Friendly, Unkind – Kind, Unpleasant – Pleasant, Awful – Nice* as seen in Figure 3.4. These ratings were a part of the Likeability Subscale of the Godspeed Likeability questionnaire [13]. The average of these scales was recorded as the "Overall Rating". We examined the distribution and normality of Overall Rating based on ethnicity and scenarios using histogram and density plots, see Figure 3.5. We can also refer the Q-Q plots in Figure 3.6. Visually, the plots suggest that the overall ratings may follow a normal distribution. However, to confirm the normality statistically, we performed a Shapiro-Wilk normality test. The results are presented in Table 3.1.

The Shapiro-Wilk test results indicate that the ratings given by HY participants in both scenarios and participants of the ASA group in a setting that is different from their

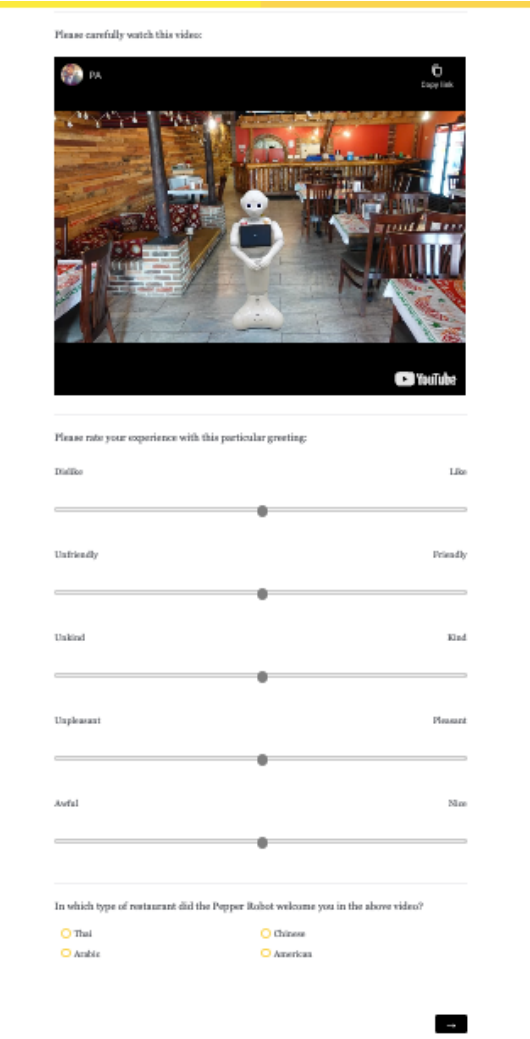



Figure 3.4: User interface with embedded video and GodSpeed Likeability questionnaire

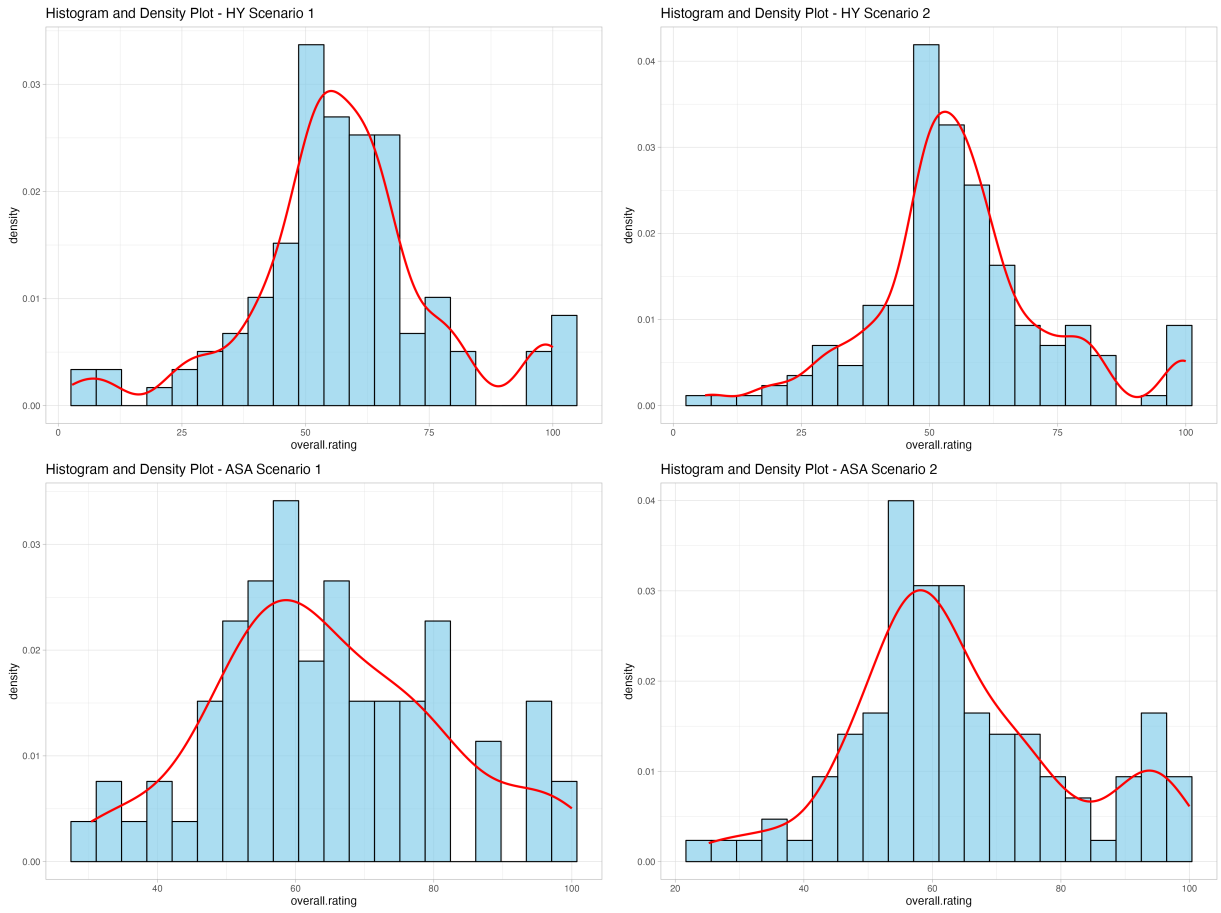



Figure 3.5: Histogram and density plots for Overall Ratings for HY and ASA in both scenarios

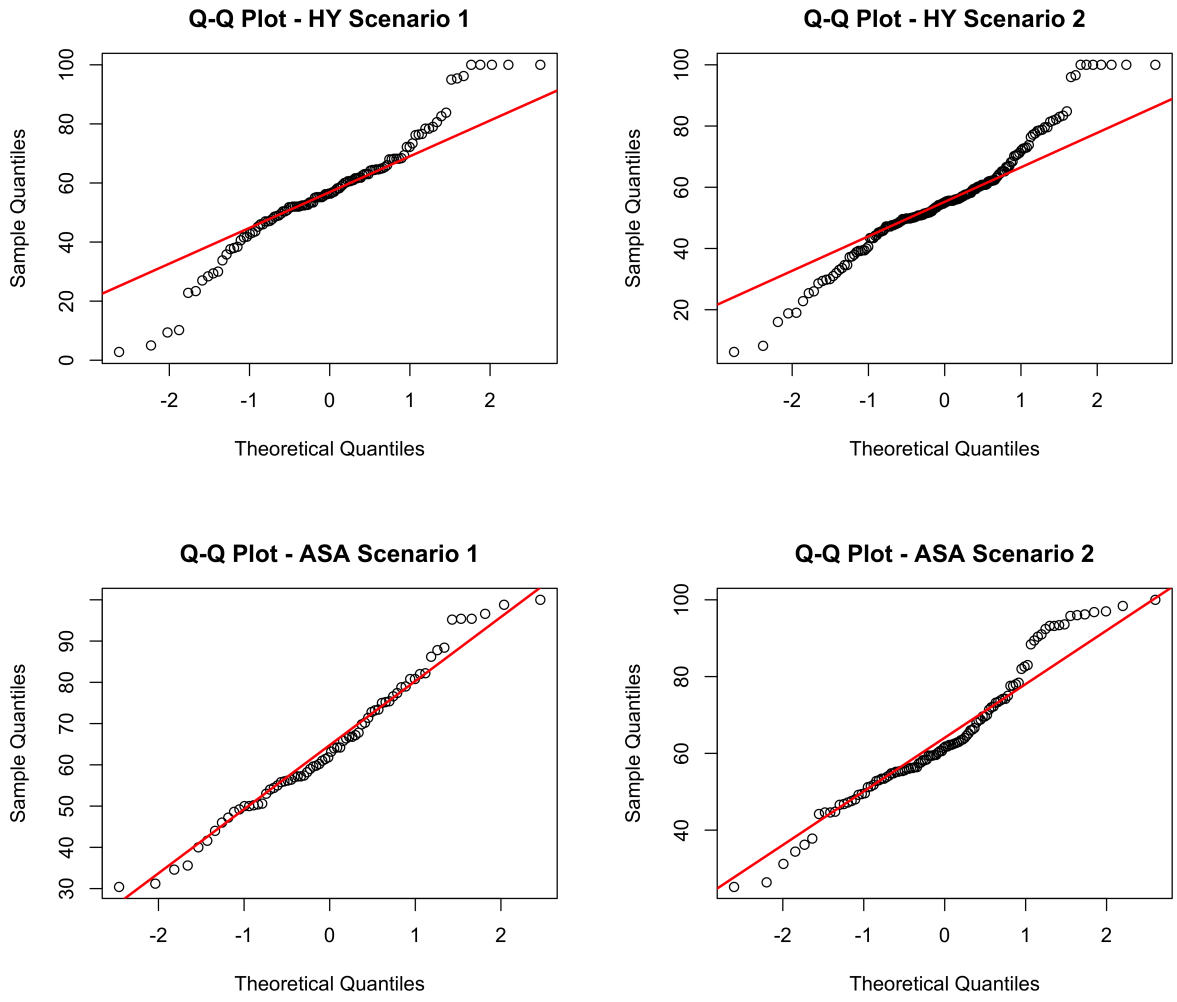



Figure 3.6: Q-Q plots for Overall Ratings for HY and ASA in both scenarios

| Category | W | p-value |
|---|---|---|
| HY (Scenario 1) | 0.95829 | 0.001 |
| HY (Scenario 2) | 0.96198 | 0.0001 |
| ASA (Scenario 1) | 0.98090 | 0.3434 |
| ASA (Scenario 2) | 0.95981 | 0.0024 |

Table 3.1: Results of the Shapiro-Wilk normality test for overall rating distributions.

ethnic background significantly deviate from normality ($p < 0.05$). However, the ratings for the ASA group for Scenario 1, where they were greeted in a restaurant setting that matches their ethnic background, follow a normal distribution ($p = 0.3439$). Although the Shapiro-Wilk test suggests that normality is violated in most of our cases, this does not necessarily rule out the use of linear mixed-effects models (LMMs) [15]. Since LMMs assume normality of residuals rather than the raw data, we will assess the residuals of our models in both Scenario-1 (Own Category) and Scenario-2 (Other Category) for the HY and ASA groups. We will check this using histograms, Q-Q plots, and residuals vs. fitted plots.

LMMs were used to investigate the effect of greeting category (Own, Other or Popular) in different scenarios. We included several potential confounding factors such as the Percentage of life lived in Western countries ((%) Life in Western Countries), Closeness to Western Culture, Closeness to Own Culture, Age, Gender, and the Big 5 personality traits of each participant. To reduce the Akaike's Information Criterion (AIC) [21], only a subset of these factors were kept in the models. We built four separate models, one model for each of the two groups in each of the two scenarios. We validated the models by standard diagnostic tests, including the examination of residuals through histograms and Q-Q plots for assessing normality, and scatter plots of residuals versus fitted values to check for homoscedasticity, ensuring that the assumptions underlying linear regression were met. As seen in Table 3.2, the Shapiro-Wilk test results indicate that the residuals for HY in Scenario 2 significantly deviate from normality ($p < 0.05$), whereas the residuals for HY in Scenario 1, ASA in Scenario 1, and ASA in Scenario 2 follow a normal distribution ($p > 0.05$). We can also perform the assumption checks of our models visually by assessing the normality and distribution of residuals in the Appendix C.

| Category | W | p-value |
|---|---|---|
| HY (Scenario 1) | 0.99086 | 0.757 |
| HY (Scenario 2) | 0.97102 | 0.002 |
| ASA (Scenario 1) | 0.98895 | 0.839 |
| ASA (Scenario 2) | 0.98988 | 0.602 |

Table 3.2: Results of the Shapiro-Wilk normality test for residuals of final models.

## 3.3 Results

The overall likability of the robot is plotted in Figure 3.7, displaying the overall ratings by both communities in both scenarios. As observed in Figure 3.7, in the first scenario, a tendency was noted among the ASA participants to prefer their own ethnic greeting, whereas the HY participants showed a preference for the "Popular" greetings. In the second scenario, neither of the communities appeared to have any strong preferences for any particular greeting category. With different trends observed, a more in-depth analysis of the four different situations (HY and ASA participants in Scenarios 1 and 2) was conducted separately. The final models are reported in Tables 3.3 and 3.4 for the two scenarios. The results of the Thematic Analysis, based on participants' responses to two open-ended questions mentioned above, are presented later in this section.

### 3.3.1 Likability of Robot Greeting in Scenario 1

**HY Participants greeted in Chinese Restaurants**: As seen in Table 3.3, the "Popular" Western greeting significantly enhanced participant's likability of robot greeting compared to when robot used their ethnic greeting ($SE$ = 1.539, $t$ = 2.359, $p$ = .022). Participants who had spent more time in Western countries reported lower likability ($SE$ = 0.053, $t$ = -3.043, $p$ = .004). Participants' Emotional Stability ($SE$ = 1.453, $t$ = 3.096, $p$ = .003) and Conscientiousness ($SE$ = 1.610, $t$ = 2.731, $p$ = .009) significantly and positively improved ratings. However, increased Openness to Experiences was found to reduce ratings ($SE$ = 1.586, $t$ = -3.125, $p$ = .003).

**ASA Participants in Arabic Restaurant**: As presented in Table 3.3, the only effect observed was that ASA participants did not like the 'Popular' Western greeting as much as they liked their ethnic greeting ($SE$ = 2.993, $t$ = -2.236, $p$ = .033).

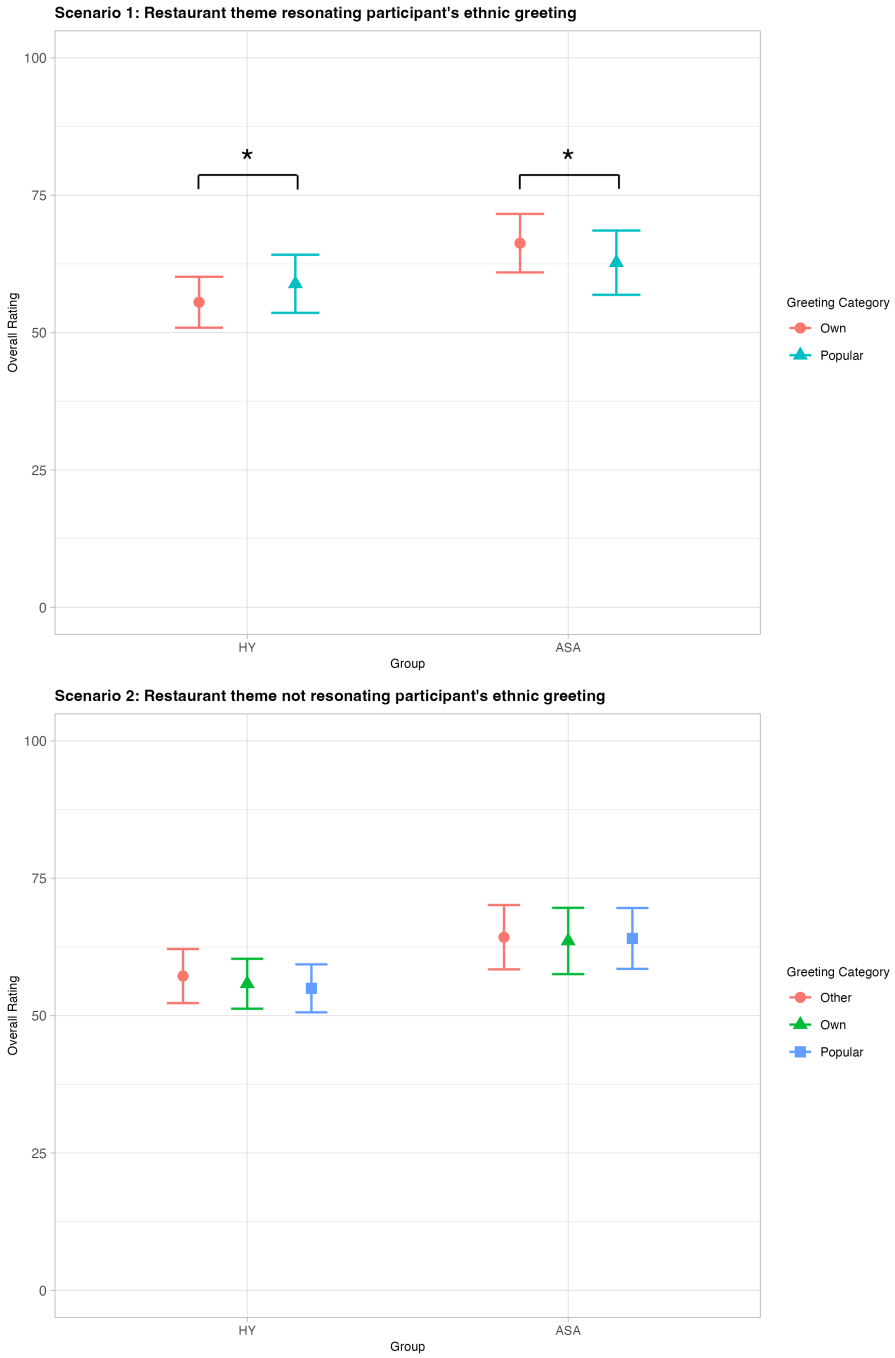



Figure 3.7: Participants' overall ratings by groups for both scenarios. *: $p < .05$.

Table 3.3: LMMs predicting participants' overall ratings in Scenario 1

| Covariate | Estimate | SE | $t$ | $p$ |
|---|---|---|---|---|
| **(HY Participants in Chinese Restaurant)** | | | | |
| Popular Western Greeting | 3.630 | 1.539 | 2.359 | * |
| (%) Life in Western Countries | -0.161 | 0.053 | -3.043 | ** |
| Age | -1.437 | 0.729 | -1.970 | . |
| Emotional Stability | 4.499 | 1.453 | 3.096 | ** |
| Conscientiousness | 4.398 | 1.610 | 2.731 | ** |
| Openness | -4.955 | 1.586 | -3.125 | ** |
| **(ASA Participants in Arabic Restaurant)** | | | | |
| Popular Western Greeting | -6.690 | 2.993 | -2.236 | * |
| (%) Life in Western Countries | -0.093 | 0.067 | -1.387 | |
| Popular Western Greeting: (%) Life in Western Countries | -0.057 | 0.051 | 1.118 | |

Participant's own ethnic greeting is the baseline.

.: $p < .1$;*: $p < .05$; **: $p < .01$; ***: $p < .001$

### 3.3.2 Likability of robot greeting in Scenario 2

**HY Participants in Arabic Restaurant**: For HY participants in an Arabic restaurant setting, Table 3.4, Age significantly and negatively affected their overall ratings ($SE$ = 0.710, $t$ = -2.791, $p$ = .014). No significant impact of the greeting category on the overall ratings was seen. Increased Emotional Stability was positively associated with overall ratings ($SE$ = 1.413, $t$ = 3.217, $p$ = .006). Openness had a significant negative impact, reducing the overall ratings ($SE$ = 1.542, $t$ = -3.136, $p$ = .005) while the percentage of life lived in Western countries was significantly correlated with decreased overall ratings ($SE$ = 0.051, $t$ = -2.446, $p$ = .021).

**ASA Participants in Chinese Restaurant**: In the setting of a Chinese restaurant, the greeting category did not have a noticeable impact on overall ratings by ASA participants, see Table 3.4. However, Agreeableness was significantly and negatively correlated with overall ratings ($SE$ = 1.853, $t$ = -3.028, $p$ = .005). Male participants in this group provided significantly higher overall ratings compared to female participants ($SE$ = 4.193,

Table 3.4: LMMs predicting participants' overall ratings in Scenario 2

| Covariate | Estimate | SE | $t$ | $p$ |
|---|---|---|---|---|
| **(HY Participants in Arabic Restaurant)** | | | | |
| Chinese Greeting | -1.506 | 1.560 | -0.965 | |
| Popular Western Greeting | -2.928 | 1.560 | -1.878 | . |
| (%) Life in Western Countries | -0.126 | 0.051 | -2.446 | * |
| Conscientiousness | 2.361 | 1.566 | 1.508 | |
| Age | -1.980 | 0.710 | -2.791 | ** |
| Emotional Stability | 4.547 | 1.413 | 3.217 | ** |
| Openness | -4.837 | 1.542 | -3.136 | ** |
| **(ASA Participants in Chinese Restaurant)** | | | | |
| Muslim Greeting | -0.683 | 2.165 | -0.316 | |
| Popular Western Greeting | -0.222 | 2.165 | -0.103 | |
| Agreeableness | -5.613 | 1.853 | -3.028 | ** |
| Gender (Male) | 11.816 | 4.193 | 2.818 | ** |
| Emotional Stability | 5.968 | 1.63 | 3.663 | *** |

Greetings aligned with the restaurant theme and Gender (Female) are the baseline. .: $p < .1$; *: $p < .05$; **: $p < .01$; ***: $p < .001$

$t = 2.818$, $p = .008$). Emotional Stability also had a significantly positive impact and increased the overall ratings ($SE = 1.63$, $t = 3.663$, $p < .001$).

### 3.3.3 Thematic Analysis of Open-ended Questions

To better understand why people preferred certain greetings in specific situations, we analyzed responses to open-ended questions using a six-step Thematic Analysis [23]. By analyzing responses to two distinct questions separately, we identified key likes, dislikes, and reasons for greeting preferences, uncovering broad themes and subthemes that reveal participants' overall motivations behind their choices. One researcher conducted the Thematic Analysis of all the data and another researcher independently reviewed a quarter of it. We checked how well the two analyses agreed by using Kappa's agreement score [77] to

measure the consistency.

The results of calculating weighted kappa were 0.926 (SE = 0.07, 95% CI = [0.787, 1]) and 0.939 (SE = 0.107, 95% CI = [0.729, 1]) for the two scenarios, which shows "almost perfect agreement" according to [77]. We assigned codes for themes and subthemes of Question 1 as [**CX**] and [**CX.Y**] and Question 2 as [**DX**] and [**DX.Y**], respectively, where X and Y are numbers. We have mentioned the percentage of responses that contained those themes and subthemes in parenthesis for both groups. The themes and subthemes along with these percentages have been summarised in Tables 3.5 and 3.6. In the tables 3.5 and 3.6 along with subthemes we have added an "overall" row for reporting percentages of overall occurrence of that particular theme. We have also quoted some comments that provided critical insights for analysis and discussion.

**Question 1:** In a restaurant setting that matched a cultural background where the participant's ethnic greetings were used, the following themes and subthemes emerged. A significant theme was the *Preference for Own Ethnic Greetings* [**C1**] (HY = 63.79%, ASA = 72.22%), where participants valued cultural consistency [**C1.1**] (HY = 27.59%, ASA = 36.11%), alignment with their ethnicity and the restaurant's theme to emphasise authenticity, respect for the culture, and consistency with the theme of the restaurant. This also included a familiar and homely feeling [**C1.2**] (HY = 17.24%, ASA = 8.33%), fostering a sense of home, familiarity, and friendliness, which is especially valued in foreign settings. Conversely, there was a *Preference for Western Greetings* [**C2**] (HY = 15.52%, ASA = 11.1%), appreciated for their neutral and universally understood nature, preferred for their familiarity and broad appeal. The importance of *Human Interaction Preference* [**C3**] (HY = 6.90%, ASA = 13.89%) highlights the demand for the warmth of interacting with humans instead of robots in hospitality settings. *"I prefer to not have a social robot, and rather, a real human. This makes the experience more personable. (**C3**)"*. Some participants expressed *No Strong Preference* [**C4**] (HY = 13.79%, ASA = 11.11%), indicating flexibility and openness to various greeting forms, while others were *Uncomfortable with Cultural Adaptation **[C5]*** (HY = 1.72%, ASA = 2.78%), feeling uneasy with cultural adjustments in greetings.

*The Restaurant Background* [**C6**] (HY = 1.72%, ASA = 2.78%) also had an influence on greeting preferences, with factors like the demographic location [**C6.1**] (ASA = 2.78%) and ambiance [**C6.2**] (HY = 1.72%) of the restaurant playing crucial roles. The location impacted cultural expectations and preferences, while the ambiance, whether high-end or casual, shaped the appropriateness and reception of different greeting types. *"I think it depends on the restaurant. I've grown accustomed to restaurants in a Western country greeting by Western customs. Also depends on how high-end the restaurant is. If it's just a fast food chain, im not expecting any thing. But if it's high end I'm expecting a very cultural*

| **Theme** | **Subtheme** | **HY %** | **ASA %** |
|---|---|---|---|
| Preference for Own Ethnic Greetings | Cultural Consistency | 27.59% | 36.11% |
| | Familiar and Homely Feeling | 17.24% | 8.33% |
| | Overall | 63.79% | 72.22% |
| Preference for Western Greetings | Overall | 15.52% | 11.1% |
| Human Interaction Preference | Overall | 6.90% | 13.89% |
| No Strong Preference | Overall | 13.79% | 11.11% |
| Uncomfortable with Cultural Adaptation | Overall | 1.72% | 2.78% |
| The Restaurant Background | Demographic Location | - | 2.78% |
| | Ambiance | 1.72% | - |
| | Overall | 1.72% | 2.78% |
| Enhanced Implementation | Voice | 18.97% | 11.11% |
| | Language | 3.45% | 2.78% |
| | Gesture | - | 2.78% |
| | Appearance | 1.72% | - |
| | Overall | 25.86% | 16.67% |

Table 3.5: Greeting Preferences in Scenario 1

*experience (**C6, C6.2**)"* Additionally, requests for *Enhanced Implementation* [**C7**] (HY = 25.86%, ASA = 16.67%) were identified, including areas of voice [**C7.1**] (HY = 18.97%, ASA = 11.11%), language [**C7.2**] (HY = 3.45%, ASA = 2.78%), gesture [**C7.3**] (ASA = 2.78%), and appearance [**C7.4**] (HY = 1.72%), pinpointing the need for improvements in tone and accent of robot's voice, multilingual support, naturalness of movements, and better appearance of the robot. *"I prefer the welcoming message to align with the culture (**C1**), however, the sound of the robot was weird, and the accent as well made it a little confusing. (**C7, C7.1**)"*

**Question 2:** In a parallel examination of greeting preferences in a restaurant setting of some other community, the *Preference for Other Ethnic Greetings* [**D1**] (HY = 43.10%, ASA = 58.33%) revealed an appreciation for ethnic greetings that align with the restaurants' cultural backgrounds and the restaurant's theme, stressing cultural consistency [**D1.1**] (HY = 25.86%, ASA = 25%) and cultural learning [**D1.2**] (HY = 6.90%, ASA = 13.89%). The *Preference for Western Greetings* [**D2**] (HY = 34.48%, ASA = 16.67%) illustrated topics ranging from lack of willingness to explore new cultures and the universality of Western greetings, through the subthemes of cultural disinterest [**D2.1**] (HY = 3.45%, ASA = 2.78%) and familiar comfort [**D2.2**] (HY = 10.34%), suggesting a preference for universally compliant interactions. The *Human Interaction Preference* [**D3**] (HY = 6.90%, ASA = 8.33%) and *No Strong Preference* [**D4**] (HY = 13.79%, ASA = 19.44%) echoed the themes identified earlier, reinforcing the value placed on genuine human connections and flexibility in greeting preferences. A *Preference for Own Ethnic Greetings* [**D5**] (HY = 1.72%, ASA = 11.11%) further emphasised personalised ethnic greetings in enhancing customer experiences. *"If I am not familiar with the culture I would not have any preference because I am ignorant (**D4**). But it would be nice to explore the culture with a robot that aligns with the culture so we can learn more about it (**D1, D1.2**)"*

The *Restaurant Background* [**D6**] (HY = 6.90%), including the demographic location [**D6.1**] (HY = 5.17%) and ambiance [**D6.2**] (HY = 1.72%) of the restaurant, and *Enhanced Implementation* [**D7**] (HY = 10.34%, ASA = 19.44%), covering voice [**D7.1**] (HY = 3.45%, ASA = 5.56%), language [**D7.2**] (HY = 8.62%, ASA = 13.89%), also were mirrored in earlier themes, underlining the influence of setting on greeting preferences and identifying similar areas for improvement in cultural integrated systems. Lastly, *Uncomfortable with Cultural Adaptation* [**D8**] (HY = 13.79%, ASA = 8.33%) introduced concerns over unsettling feeling [**D8.1**] (HY = 3.45%) and privacy concerns [**D8.2**] (HY = 10.34%, ASA = 8.33%), reflecting discomfort with the adaptation process and ethical considerations in culture detection and adaptation. *"I would prefer for the social robot to align with the cultural background of the restaurant (**D1**), or to just follow popular Western culture customs (**D2**). I would find it somewhat strange/overassuming if it followed my cultural*

| **Theme** | **Subtheme** | **HY %** | **ASA %** |
|---|---|---|---|
| Preference for Other Ethnic Greetings | Cultural Consistency | 25.86% | 25% |
| | Cultural Learning | 6.90% | 13.89% |
| | Overall | 43.10% | 58.33% |
| Preference for Western Greetings | Cultural Disinterest | 3.45% | 2.78% |
| | Familiar Comfort | 10.34% | - |
| | Overall | 34.48% | 16.67% |
| Human Interaction Preference | Overall | 6.90% | 8.33% |
| No Strong Preference | Overall | 13.79% | 19.44% |
| Preference for Own Ethnic Greetings | Overall | 1.72% | 11.11% |
| Restaurant Background | Demographic Location | 5.17% | - |
| | Ambiance | 1.72% | - |
| | Overall | 6.90% | - |
| Enhanced Implementation | Voice | 3.45% | 5.56% |
| | Language | 8.62% | 13.89% |
| | Overall | 10.34% | 19.44% |
| Uncomfortable with Cultural Adaptation | Unsettling Feeling | 3.45% | - |
| | Privacy Concerns | 10.34% | 8.33% |
| | Overall | 13.79% | 8.33% |

Table 3.6: Greeting Preferences in Scenario 2

*background at a place that was not Chinese just because I am Chinese. Personally, I would find it conflicting with the cultural background of the restaurant I'm at (**D1.1**), as well I would be confused as to why it automatically referred to Chinese cultural customs only based on my looks and not my preference.(**D8, D8.2**)"*

## 3.4 Discussion

Our findings shed light on people's likability and their preferences towards ethnic greetings of a humanoid robot with respect to locational context. We explored how participants' identity (HY or ASA), percentage of their life lived in Western countries, personality, the theme of the restaurant, and implementation of cultural elements could shape these preferences. As I discuss the results, it is important to note the use of the English language package for implementing the greetings using the Pepper robot. Additionally, the design choice to keep the rest of the communication in English was made so that both the HY and ASA communities could understand the entire welcome message in both the Chinese and Arabic restaurant settings. This may have influenced participants' preferences for cultural greetings in both the scenarios (see Figure 3.1). The results reported and discussed in this study can be analyzed with these limitations in mind. The thematic results [**C7.1, C7.2, D7.1, D7.2**] of the open-ended responses support the impact of this limitation.

**Greeting Preferences in Scenario 1**: We discuss here how our findings develop insights around **RQ1**. In **Scenario 1**, when participants experienced the greetings in a restaurant whose theme resonated with their ethnic greeting, unlike the initial hypothesis **H1** predicting that all participants would prefer their ethnic greetings in this scenario, our results indicate a more complex picture. In particular, HY participants indicated a preference for "Popular" Western greetings in such scenarios. This unexpected preference highlighted the impact of the percentage of life lived in Western countries on individual cultural preferences and the broad appeal of Western greetings, as will be explained. ASA participants, however, showed an alignment with our hypothesis, i.e. preferring greetings consistent with their own ethnic greeting in Scenario 1. Thematic Analysis results concerning the aspects of cultural implementation [**C7, C7.1, C7.2**] indicated that the aspects of tone and not having entire communication in their native language had a consequential impact on HY participants' preference. When individuals were dissatisfied with aspects of the ethnic greeting's implementation, which was more common in HY participants, they often opted for the widely recognised Western behaviour in social robots, as it could provide a more seamless and genuine interaction experience. Therefore, the difference in the preference for greetings between the two experimental groups in Scenario 1 could be due

to the perception of the voice and language used by the robot and the percentage of life lived in Western countries by the participants.

**Greeting Preferences in Scenario 2**: In a restaurant whose theme did not align with their own or Western greetings, **Scenario 2 and RQ2**, neither group showed a strong preference for any particular greeting, highlighting the complex factors influencing greeting preferences. In this scenario, we even observed in the thematic analysis that both groups had a lower preference for their ethnic greetings [**D5**]. Our results suggest that participants might value the authenticity of the experience or the aspect of learning about a new culture over the comfort of familiar greetings [**D1, D1.1, D1.2**] in unfamiliar cultural contexts. This insight points to the potential for social robots to serve as cultural ambassadors, facilitating cultural exchange and understanding in diverse settings.

**The Role of Personality Traits and Demographics**: Our analysis revealed a significant role of certain personality traits, such as Emotional Stability and Conscientiousness, and demographics, including age, in shaping participants' greeting preferences. These findings indicate that individual psychological traits and life experiences critically influence perceptions of social robot greetings. For example, the fact that older HY participants were less satisfied with greetings in an Arabic restaurant setting suggests that age might have an impact on their openness to other cultures or their willingness to adapt to new situations. However, other explanations for this finding are possible and need to be explored further.

**Implications for Social Robot Interaction Design**: The participants' feedback on areas for enhanced implementation, including improvements in voice, language, gesture, and appearance, highlights the challenges in creating social robots that are not only culturally adaptable but also respectful of cultural nuances. The tone and accent of the robot's voice impacted the likability of the greetings [**C7.1**], especially in the context of interactions happening in a setting where there was an alignment with the participant's background. Designers and developers of social robots should consider these factors to ensure that robots can handle global cultural diversity effectively.

Results from our thematic analysis raised important ethical considerations, particularly regarding the discomfort some participants felt with cultural adaptation by robots. These issues highlight the need for cautious deployment of social robots to respect cultural diversity and personal preferences while safeguarding privacy and autonomy [**C5, D8**]. Not everyone will appreciate it if the robot assumes their ethnic greeting and starts adapting. I realised something similar after reading about cultural adaptations through complex greeting gesture selection systems in social robots in [117] and that led me to pursue this study.

## 3.5 Limitations and Future Work

- The videos shown to participants cannot fully capture the unique atmosphere and subtle aspects of experience of interacting with real robots in real-world restaurant applications. This could potentially affect the validity of our conclusions.
- Focusing primarily on university students limits the diversity of perspectives on cultural integration in social robots.
- The number of participants was unbalanced, with fewer individuals from the ASA group.
- The greetings in our study were delivered using the same voice and accent due to technical constraints. As mentioned earlier, the greetings were implemented using the English language package of the Pepper robot that could induce an English accent in pronunciation.
- Continuing communication in English after the initial cultural greeting could diminish the authenticity of the greetings, making the interaction appear more as a formality or a forced effort.

Future research may address these issues by including participants from diverse backgrounds and ages, ensuring balanced representation among all groups, and exploring a wider range of settings such as hotels, stores, and airports. Additionally, robot greetings could be enhanced by incorporating specific cultural accents and phonetics to create a more authentic interaction. Further studies should investigate how people's preferences evolve as social robots become more common across different cultures and how long-term exposure influences perceptions of authenticity and acceptance. These explorations will provide valuable insights into the role of social robots in diverse environments and contribute to their effective cultural integration.

We observed how sensitive cultural integrations could be in HRI and how they impact communication, which eventually affects the perception of social robots. This chapter marks the end of our segment on cultural integration research in HRI under our broad theme of "Inclusivity in Communication". On the other hand, it also marks the starting point where, from the next chapter, I will report my studies around finding applications of social robots in speech therapy.

# Chapter 4

# Study 2: Design Considerations for Applications of Social Robots in the Stuttering Clinic

Social robots have been used in HRI for speech therapy [27, 100, 123], but there is a lack of foundational groundwork on which new research can be built. In other words, if we aim to study different aspects of HRI in the context of speech therapy—such as non-verbal gestures, gaze, cultural integration, empathy, and emerging technological advancements in sensing and detection to improve interactions—we do not yet know which applications would be useful or what design considerations should be taken into account.

To provide context for this chapter, it is important to note that SLPs (Speech and Language Pathologists) are professionals who conduct speech therapy sessions. In this thesis, we observed stuttering therapy sessions for preschoolers with SLPs. For this study, I analyzed past research in this field conducted by members of the Social and Intelligent Robotics Research Lab (SIRRL)[1] [28,29,64], combined with in-person shadowing and interviews with SLPs, to ground my findings in the real-world challenges that SLPs face. Based on these insights, my goal was to explore the applications of social robots in stuttering clinics for their use by SLPs through a co-design approach.

Following the Constructivist variant of GTM [31], I analyzed the collected data to determine whether the emerging solutions are viable for implementation in the context of stuttering therapy for preschoolers, given the technological limitations provided by the

[1] https://uwaterloo.ca/social-intelligent-robotics-research-lab/

setting and stakeholders. As a result of this study, I propose the types of applications that can be built using social robots, how these applications could be designed, and how the robot may communicate with stakeholders in these applications. Broadly, the results suggest that the robot could serve as a toolbox with various interactive games or could help model the speech of the child or parent, depending on the use case. The applications should be interactive, engaging, and structured, and should involve SLPs or other accompanying caregivers. The results suggest that the robot should speak in short sentences, minimise questions and rather make supportive comments, should not interrupt when the child speaks, and reduce stressors for the child during these interactions. The applications should be designed to provide opportunities to increase utterances from the child so that the SLPs can collect speech samples to track the child's progress during the sessions.

## 4.1 Research Objective

Our research objective in this study was to explore the applications of social robots in stuttering clinics and propose design considerations for these applications intended for use by SLPs through a co-design approach. The study is an extension of our previous work by other researchers at SIRRL [28, 29, 64], which initiated the research direction of exploring the feasibility of using social robots for stuttering interventions. A Double Diamond design process [46] was chosen to find applications of social robots in speech therapy since our previous works at SIRRL, as it provided a structured path from exploration to execution. The Double Diamond design suggests divergent (broad exploration) and then convergent (focused decision-making) approaches in each diamond. The process has four phases according to this framework: Discover (diverge), Define (converge), Develop (diverge), and Deliver (converge) phases. Discover and Define phases are in the first diamond where we explore problem space. Develop and Deliver phases are in the second diamond which is focused more on the solution. This current research focuses on discovering and defining (first diamond) potential considerations for developing applications on social robots to be used by SLPs in their clinic in the future (second diamond). We actively involve SLPs in the Definition phase (see Figure 4.1). By incorporating insights from the SLPs, we discuss applications that will facilitate the adoption of social robots in speech therapy practices.

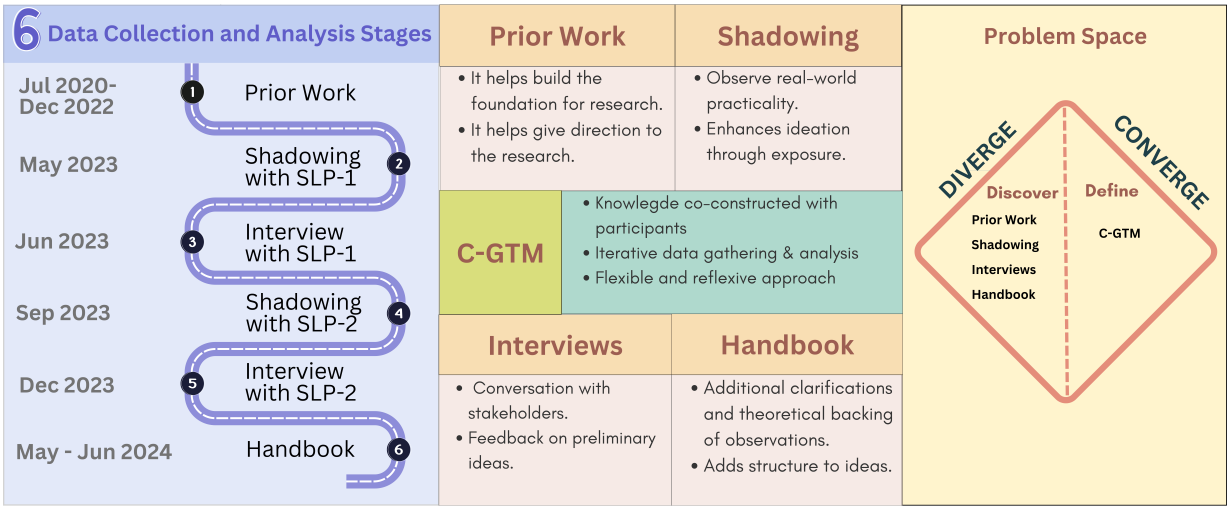



Figure 4.1: Data Collection and Analysis Stages. The image highlights the first phase ("first diamond") of the Double Diamond Process Model [46], as applied in this study. This work is informed by prior research ( [28, 29, 64]) and the Handbook on stuttering ( [62]).

## 4.2 Methodology

We followed a co-design approach and employed the Double Diamond Design Process Model that was used in past research ( [29]) at SIRRL, focusing on the problem definition phase: Discovery and Definition (See Figure 4.1). As we wanted to build solutions to meet the actual needs and preferences of stakeholders, it was crucial to involve them actively in the core phases of the design process. This collaborative approach ensures that the SLPs contribute to the design of solutions. I followed a Grounded Theory Methodology approach for data collection and analysis to address the research objectives. I will discuss this approach in this section and the outcomes of the analysis in the following sections.

### 4.2.1 Grounded Theory Methodology

In this study, I adopted data collection and analysis strategies from Charmaz's Constructivist Grounded Theory Methodology (C-GTM) [31]. The core processes—open coding, axial coding, and selective coding—allow researchers to develop a theory grounded in the data itself, making it very effective for developing explanatory conceptual theories in emerging academic fields [39]. There are several key aspects of C-GTM: it recognises that

knowledge is co-constructed by the researcher and participants rather than simply emerging from the data [39]. C-GTM accounts for the researcher's background and perspectives along with interactions with participants and the data will have an effect on the theory development process. Much like classic GTM, in C-GTM there are iterative cycles of data gathering and analysis where each cycle informs the next steps. Lastly, C-GTM encourages researchers to remain flexible and reflexive throughout the research process.

C- GTM enabled me to contextually ground our results in the data primarily gathered through direct SLP engagement, offering a flexible framework for identifying the practical applications of social robots rather than formulating an abstract theory. I collected data over six stages (denoted as DC-1 through DC-6), with each stage informed by the preceding stage's data analysis, see Figure 4.1.

**Prior work (DC-1):**

For establishing a base, in the first stage of our data collection (DC-1), I analyzed previously published exploratory work by other researchers at SIRRL regarding applications of social robots in speech therapy [28, 29, 64]. Gupta et al. [64] summarised stuttering treatment approaches over the past two decades and Chandra et al. [28] proposed higher-level conceptualizations for incorporating social robots in stuttering interventions. The proposals were to be refined through co-design and feedback from other stakeholders during empirical research [28]. In [29], the authors employed co-design methods, including surveys, individual expert interviews, and focus groups, to identify common challenges faced by speech-language pathologists and to explore how social robots could address these challenges. The study included four SLPs from the Institute of Stuttering Treatment and Research (ISTAR) in Canada. The participants, all female, age range [29 52] ($M_{\text{age}} = 38.5$, SD = 10.08). The SLPs had between 4 to 28 years of professional clinical experience with stuttering. All participants were registered with the Alberta College of Speech-Language Pathologists and Audiologists. None had previously interacted with robots. The questions posed to the SLPs focused on preschool and school-age children. The study suggested that social robots could be integrated into stuttering therapy, particularly serving as engaging and motivational partners, providing consistent speech models, reinforcing social and speech skills, acting as homework partners, and aiding in the transition through different treatment phases, such as establishment, transfer, and maintenance. DC-1 was conducted virtually due to the COVID-19 lockdown. This study builds on the foundational prior work through in-person data collection, informing the project's future direction to assess the practicality of the suggested solutions through shadowing.

**Shadowing SLP-1 (DC-2):**

To obtain the specific details needed to address the gap in existing research and following convenience sampling, we decided to conduct in-person shadowing and interview sessions with the SLPs (referred to as SLP-1 and SLP-2, not involved in DC-1) at KidsAbility[2] in Waterloo, Ontario, Canada, starting with shadowing sessions with SLP-1. Shadowing is a qualitative user experience research technique, where the researcher acts as an observer only and records their observations simultaneously without interfering with the ongoing activities [68]. Shadowing helps HRI researchers to learn about real-world contexts and understand user preferences. The goal was to be present in the therapy room during the sessions to gather practical insights into the available opportunities for addressing our research objective. The first and second authors participated in four shadowing sessions with SLP-1 and their clients and took detailed notes about their observations. The detailed notes from the shadowing sessions were analyzed to brainstorm potential applications of social robots in the stuttering sessions of preschoolers. This step, DC-2, was crucial in exposing us to a range of games that are used to engage the child in the session. The nature of games, cognitive levels of the games, interaction between participants in the game scenarios and outcomes expected from the child through the games were some key observations. The children shadowed were two boys and two girls, aged 3 and half years to 5 years with $M_{\text{age}} = 4$ and SD $= 0.612$. Three of the four children were non-severe stutterers with English as their first language. One child, who was a severe stutterer, did not have English as their first language. Each session included the SLP, the child, and one or more family members, which could consist of the mother, father, both parents or siblings.

The shadowing sessions started with the SLP taking updates from the parents, and asking for the severity ratings [3] of the child's speech since their last visit, while the child was given a toy to play with. The severity rating refers to the assessment of the child's stuttering in their daily communication, rated by parents or SLPs. These ratings typically range from 0 (no stuttering) to 10 (extremely severe stuttering), helping to guide treatment decisions and track progress. The session included the SLP providing the child with toys and games to play with, considering the child's preferences. In Table 4.1, I have summarised some of the games I observed being used to engage children during the sessions. I listed

[2] https://www.kidsability.ca/

[3] https://www.uts.edu.au/sites/default/files/2022-12/Child%20Stuttering%20Severity%20Chart%20v2%20-%20Dec%202022.pdf

the characteristics of games based on whether they involved tangible toys or were screen- or tablet-based. All of the games during the shadowing sessions were not screen- or tablet-based. I also noted whether the games included physical activities that required children to get up and move around the room. Additionally, I recorded whether turn-taking was a part of the game and required others, such as parents or SLPs. The child was allowed to play with a toy/game for 5-10 minutes and then a new toy was introduced. There were around a total of 4-6 different games per therapy session. The games also included participation from the SLP or the accompanying family members. The insights gained from the analysis at this stage informed the questions asked in the next stage. The insights also helped brainstorm potential application scenarios and identify any concerns raised by the SLP after being exposed to a social robot. These opportunities and considerations were explored in a semi-structured interview with SLP-1.

**Interview SLP-1 (DC-3):**

After the shadowing sessions, we developed initial ideas of the prospective applications of social robots in clinical settings. However, an in-person interview with the SLP provided us with the opportunity to gain their perspective on their challenges and usage of social robots in stuttering sessions along with their feedback on our initial ideas. The SLP was a female of 56 years with 29 years of clinical experience with stuttering, primarily with preschoolers. She was registered with CASLPO (College of Audiologists and Speech-Language Pathologists of Ontario) and had not interacted with a social robot before. The interview format was semi-structured, and I was the main interviewer alongside two other SIRRL researchers who also participated in seeking clarifications and getting their queries addressed. Our queries were centred around our preliminary ideas of the use cases of a social robot in the stuttering therapy sessions. The interview aimed to explore the challenges faced by clinicians and their clients from the clinician's viewpoint. We asked five questions, each addressing different challenges, which were adapted from [29].

1. What are the most common challenges that you face with your clients?
2. What do you think are the major challenges that might prevent the children from making progress during a therapy session?
3. What major challenges do you encounter with parents throughout the entire therapy process?
4. What role do you think the robot could fit into at this moment, and do you believe it would be useful in some sessions with certain children?

Table 4.1: Description of a Subset of Games Observed During Shadowing

| **Game** | **Description** | **Characteristics** | **Game Partners** |
|---|---|---|---|
| Plastic Toys | The child plays with plastic toys such as Lego, model trucks and helicopters. | Tangible toys, turn-taking | Child - Parents |
| Knock-Knock | The game involves revealing cat pictures from a box. | Tangible toys | Child - SLP |
| Bubbles | The SLP blows soap bubbles, and the child tries to burst them. | Tangible toys, physical activity | Child - SLP |
| Cube Animals and Food | The child places food in the mouths of animals drawn on the sides of a cube. | Tangible toys, turn-taking | Child - Parents |
| Zoo Toy Set | The child plays with various toy animals and toy slides, similar to those in playgrounds. | Tangible toys, turn-taking | Child - Parents - SLP |
| Catch Some Bugs | The child catches bugs using a fishing stick with a magnet on a thread, with bugs placed on a puzzle-like board. | Tangible toys, turn-taking | Child - Parents |
| Sea Animal Stickers | Players stick sea animal stickers onto a board. | Tangible toys, turn-taking | Child - Parents - SLP |
| Animal Bingo | A matching game where players find matching sea creatures on cards, similar to classic bingo. | Tangible toys, turn-taking | Child - Parents |
| Flashcards with Objects | Flashcards with pictures of objects are used along with plastic versions of the objects, which are matched and placed in a bag. | Tangible toys | Child - SLP |
| Rocket and Pump | A rocket is placed on a launch pad and launched by jumping on a balloon to use air pressure. | Tangible toys, physical activity | Child - SLP |

5. What concerns would you have if you were to use social robots in therapy sessions?

Before asking the fourth question, we showcased a demonstration of the QT robot[4], highlighting its customizable speech controls for speed, pitch, and voice and showcasing various games developed in collaboration with the Learning Disabilities Society in Vancouver. We used the QT robot for our demonstration because it is customizable and non-threatening, designed for children with special needs education. Developers and therapists can adjust the QT robot's behaviour for effective therapy, and it has been used in related projects by SIRRL researchers( [10, 11]). We also shared with the SLP experiences from previous studies SIRRL researchers had conducted on using the QT robot as an educational tool for students with learning disabilities, which demonstrated positive impacts on student engagement and reduced off-task behaviours ( [10, 11]). The purpose of this demonstration was to illustrate the potential applications of the robot for children with disabilities and to gather feedback from the SLP. This was followed by brainstorming with the SLPs, prompting additional questions about possible applications, and note any concerns they brought up. Toward the end of the interview, the SLP recommended that we also engage with another SLP at the same institution, who works with preschoolers who stutter, to gain their additional perspective on the topic. The interview audio data was converted to text using Vosk[5], an offline open-source speech recognition toolkit for Python. To verify the accuracy of the transcriptions, I reviewed each recorded audio and cross-checked the transcriptions before conducting the analysis.

**Shadowing SLP-2 (DC-4):**

In addition to following the recommendation of SLP-1, the goal of interviewing a second SLP was to compare experiences and perspectives, identify common patterns and look for any major variations in the codes of collected data that could lead to the refinement of emerging concepts. In stage four, we shadowed one client with SLP-2, following the same steps as in DC-2. This provided further insights into the practicalities and challenges faced in therapy sessions. The child in the shadowing session was a 4 years 9 months old boy with non-severe stuttering and English as his first language.

[4] https://luxai.com
[5] https://pypi.org/project/vosk/

**Interview SLP-2 (DC-5):**

Similar to DC-3, we conducted a semi-structured, forty-minute interview with SLP-2. The SLP was a female of 55 years with 28 years of clinical experience with stuttering in both preschoolers and school-aged children. She was registered with CASLPO and had not interacted with a social robot before. I interviewed the SLP, and the format mirrored that of the interview with SLP-1.

**Extant text: Handbook for Clinical Speech Therapy (DC-6):**

In addition to the shadowing sessions and interviews that were more observational and practical, I refer to an additional data source to get a theoretical understanding of of what is expected in stuttering sessions by referencing Chapters 12-14 (related to the treatment of younger and older preschool children) of the handbook *Stuttering: An Integrated Approach to Its Nature and Treatment*, [62]. GTM allows for various data types, including individual and focus group interviews, document collection, observations, and mixed methods [16]. Specifically, "Extant Texts" are mentioned as a valid data source [31], e.g. literature, public records, etc. In our project, I selected a handbook as a data source and created notes on established clinical practices, focused on the treatment programs, and provided details on the SLP and caregivers' interactions with the child.

After each stage of data collection, I generated memos based on the initial coding of the data gathered at that stage. These memos were then discussed with the research team to inform decisions about subsequent data collection, ensuring alignment with our research objectives. In the early stages, preliminary codes were derived through open coding, which involved tagging relevant portions of various data sources (e.g., prior publications, notes from shadowing, interview transcripts, and extant texts). Importantly, this process was iterative and ongoing throughout data collection, with new codes being constantly compared against existing ones. By the end of the iterative open coding phase, we had thirty-seven preliminary codes from the data collected in stages DC-1 through DC-6 (as illustrated in Figure 4.3). In line with the co-construction concept of C-GTM, three HRI team members engaged with the data using this flexible framework to reflexively interact with the codes, selectively dropping certain codes during the analysis process. Five codes were dropped during iterative data collection and analysis due to the current technical infeasibility of applications that would need precise real-time sensing and response of the robot in noisy environments, as identified through observations during shadowing sessions and discussions with SLPs during interviews (for details, see Subsection 4.4.1). Dropped codes related to certain applications of social robots, e.g. suggestions that they could serve as conversation

partners, homework assistants, increase engagement through personalization, or reinforce social and speech skills in the child. During the focused coding phases, I clustered and related codes to one another, refining them into higher-level conceptual categories and related subcategories. This process resulted in the development of three core categories, each with two to five subcategories, as shown in Figure 4.3. Each subcategory addressed related concepts, encompassing particular properties of designing the application of social robots in the stuttering clinic. In the final selective coding phase, I integrated and refined the core categories that capture our research objectives, developing our proposal for the applications of social robots in stuttering clinics and the design considerations for these applications.

## 4.3 Results

The core categories and subcategories along with key comments from the interviews with the SLPs are detailed in this section. When interpreting the core categories, we realised that, in addition to answering what the possible applications of social robots are in the context of stuttering therapy for preschoolers, it also helps us answer how these applications should be designed during interactions. It further informs how the robot should communicate during these interactions. The core categories and the questions it helps us answer are highlighted in the Figure 4.2.

### 4.3.1 Core Category 1: Practical Applications of Social Robots:

The applications of social robots encompass both their potential use cases and the scope of their application. For instance, the robots could act as a versatile toolbox, offering a variety of games and serving as a consistent speech model during interactions with children and parents (see Figure 4.3). ***SLP-1****: "I think it [robot] can be used in very defined activities, it can be used as a language sample, possibly it can be used with parents. I suppose to model some speaking."* It could also demonstrate different levels of stuttering to help parents understand various severity ratings. Through shadowing and interviews, it became clear that the robot could not replace a therapist but rather shall serve as a complementary tool that supports the therapist's role.

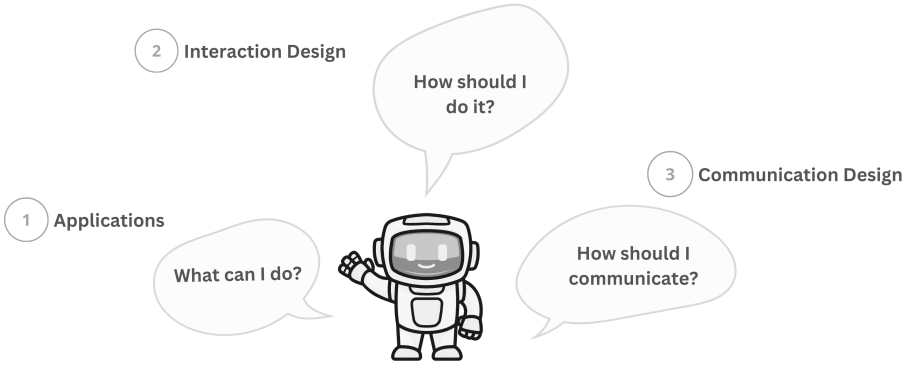



Figure 4.2: Core categories identified in our GT Analysis

### 4.3.2 Core Category 2: Interaction Design Considerations:

The previously identified applications are now further explored to determine *how* the robot should interact with participants during stuttering sessions. Play-based interactions stand out as a central key concept, i.e. designing gameplay scenarios that are engaging, entertaining, interactive, involve turn-taking, reward-based, and flexible. I observed around 24 sets of different games used by the SLPs during the sessions I shadowed. Additionally, cognitive ease is important, meaning interactions should be structured and impose a low cognitive load on the children, as well as not burden the SLP. Stuttering therapy aspects should be supervised by the SLPs. The interactions should help SLPs' aims to achieve specific speech goals, and encourage the child's speech production (see Figure 4.3). ***SLP-2:*** *"It's important to not only identify and make the sounds of animals but also to consider the variety of words we can use. Instead of just making the animal sound or naming the animal as a noun, we should talk about what the animal can do. For example, a horse can run, gallop, jump, and be ridden. You're looking at the variety of words in the English language, including nouns, verbs, adjectives, and locations. This approach can enhance vocabulary by following directions, such as putting the horse under the table or making it jump over the fence. It makes the interaction more engaging and educational, rather than simply labelling the animal and its sound."* Including game partners such as the SLPs and accompanying family members can encourage human-human connections. The gameplay should incorporate tangible props such as soft toys, plastic cars, trucks, blocks, and Lego sets to be used by the children or SLPs (see Table 4.1 for more details). The use of tablets and other screens should be minimised to engage children in the sessions, and instead, they should be used by the SLPs as part of gameplay with the robot. For instance, a tablet could feature a "spin the wheel" game displaying images of fruits, where the child must

show the robot the correct fruit when the wheel stops. The robot can take on various roles in the sessions, such as a peer, friend, or co-therapist, depending on the design and goal of the application. It is important to consider that the robot's presence or absence in some sessions may cause distractions or reduced participation in the remainder of the session, and could even upset the child.

### 4.3.3 Core Category 3: Communication Design Considerations:

Through data collection, particularly during shadowing sessions, I observed that both verbal and non-verbal communication play a crucial role in stuttering therapy sessions with preschoolers. We have established what a social robot could do in a stuttering session for preschoolers and how its interactions should be designed (see Figure 4.3). Now, we need to address *how the robot should communicate* during these sessions. Communication design considerations include the content and sound of the speech, cultural elements, and good listening practices. The speech content should be non-generic and personalised, and short sentences should be used. The words used should describe gameplay scenarios. The robot's speech should primarily consist of comments rather than questions directed at the child. Repeating simple phrases and incorporating gestures can add more structure to the interactions. HRI researchers should avoid making the robot's speech sound robotic and instead use slow and relaxed or prolonged and syllable-timed speech [62] depending on the fluency shaping strategies [114] being considered. ***SLP-1:*** *"Changing how they (parents) speak to the child because we're asking them to slow the rate down or asking them to model shorter sentences and use turn-taking, good listening, using more comments. So you know you're trying to change your own speech, it's hard."* In the scenarios, the robot should have a positive and encouraging tone. Cultural elements could be incorporated, including culturally relevant speech content (e.g. names of food or animals, music, cultural practices, etc) and greeting gestures, if the child is familiar with them through interactions at home, and importantly, if parents would like to introduce such cultural references. It is essential to demonstrate non-verbal gestures that indicate attentive listening and avoid interrupting when the child speaks.

## 4.4 Discussion

Aligning the data collection and analysis approach with C-GTM, I iteratively collected and analyzed data until additional data collection had no significant impact on the evolution

| Category | Subcategory | Properties |
|---|---|---|
| Practical Applications of Social Robots | Toolbox | 1) Variety of games, 2) Consistent speech model, 3) Demonstrate stuttering examples for parents, 4) Story-teller |
| | Complementary role | 1) Complements therapist's role, not replaces it |
| Interaction Design Considerations | Play-based Interactions | 1) Engaging, 2) Entertaining, 3) Interactive, 4) Turn-taking, 5) Rewarding, 6) Flexible |
| | Cognitive Case | 1) Structured interaction, 2) Low cognitive load |
| | Stuttering therapy aspects | 1) Supervised by SLPs, 2) Achieve speech goals, 3) Encourage speech production, 4) Facilitate human-human connection |
| | Tangible Objects | 1) Include Physical items, 2) Avoid screens |
| | Interactions with Robots | 1) Minimise distraction, 2) Varied robot roles |
| Communication Design Considerations | Speech content | 1) Non-generic speech content, 2) Descriptive comments, 3) Short sentences, 4) Personalized conversation, 5) Fewer questions, 6) More comments, 7) Use repeated phrases/gestures |
| | Speech sound aspects | 1) Non-robotic speech, 2) Slow and relaxed speech or prolonged and syllable-timed speech, 3) Enthusiastic/encouraging tone |
| | Cultural elements | 1) Cultural speech content (e.g., language-specific greetings, object names), 2) Gestures, if the child is familiar |
| | Good listening | 1) Non-verbal gestures to demonstrate good listening, 2) Not interrupting the child |

Figure 4.3: Categories, subcategories, and properties identified through GT Analysis.

of categories and subcategories. By the final stage, we developed a comprehensive set of applications and considerations for human-robot interactions in the stuttering clinic, which can serve as guidelines for other researchers exploring this domain. In this section, I discuss the emergent properties that constitute the subcategories, which in turn informed the development of the core categories presented in the results section of this chapter. These properties were iteratively constructed through C-GTM, drawing from field notes taken during analyzing prior work, shadowing sessions, interview transcripts, and analysis of extant texts. While in sociological research a 'theory' is the desired end result of using GT, our results are instead grounded design considerations and applications of social robots where we focus on identifying practical, actionable insights rather than abstract theoretical constructs. Note, that our findings are specifically limited to preschoolers who stutter.

### 4.4.1 Practical Applications of a Social Robot

Practical applications of a social robot in the stuttering therapy setting was a core category that was observed in the process of applying C-GTM. While exploring the use of a social robot in a clinical setting for preschoolers who stutter, we faced challenges that made many initial ideas infeasible as we gathered more data. During shadowing sessions, we observed that therapy with preschoolers often involves playing various games, which creates significant background noise. For example, in a session, a "Zoo toy set" game, where the child, parents, and SLP took turns playing with various animal toys and toy playground slides, and a "Catch some bugs" game, involving the child and parents using a magnet to catch bugs on a puzzle-like board, were played. This noisy environment, combined with the preschoolers' incomplete sentence formation, makes it difficult, if not impossible, for robotic/software applications to work reliably and effectively when trying to detect and understand speech, including pre-schoolers' stuttered speech. Applying artificial intelligence tools is particularly challenging in such an environment where multiple individuals are talking, and the child's speech may not be well-formed. Additionally, upon referring to the handbook [62] we recognised the importance of reporting a child's stuttering severity ratings as accurately as possible since they are crucial decision-making factors for evaluating treatment progress and determining future therapy directions. Similarly, for a robot to accurately provide reinforcement and feedback, depending upon the speech skills of the children, is very challenging. Moreover, in interviews with the SLPs, we learned that using a robot to provide feedback, appreciation, or corrections based on the SLP's input via a tablet (to avoid difficulties in automatic speech and stuttering recognition) would in fact *increase* the SLP's workload and delay feedback, making the child-SLP interaction unnatural. This point highlights the complexities of integrating social robots in the stut-

tering clinic for preschoolers, which should lighten and not worsen SLPs' workload, and not interrupt the flow of SLP-child interaction in the therapy session.

Despite these challenges, a social robot has several useful applications in speech therapy, enhancing the therapeutic process in various ways. It can serve as a versatile tool, offering a range of engaging games tailored to different speech therapy needs. The game scenarios, which can be very motivating and entertaining for the children, may provide SLPs with the opportunity to take more samples of the child's speech, so the robot should afford opportunities that encourage speech production from the preschoolers. Additionally, the robot provides a reliable model for consistent speech patterns, acting as a stable reference for clients for various fluency-shaping strategies [62, 114]. Stuttering therapy includes modifying the child's environment (home, school, community, and any other places they frequently spend time) to ensure verbal interactions with the child involve a slow rate and pauses [62]. The robot can also demonstrate varying levels of stuttering, which may help parents understand and more accurately report the severity of their child's stuttering. The robot can be used in storytelling activities, making the practice of using speech both entertaining and educational for children. Importantly, while the robot is a valuable tool, it is designed to complement and support the therapist, rather than replace them, thereby enhancing the overall therapy process.

### 4.4.2 Interaction Design Considerations

This subsection discusses the core category of interaction design consideration that was the outcome of following the C-GTM approach. Shadowing and interview sessions with the two SLPs enabled us to compile an informed list of essential aspects for designing effective interactions with social robots. One crucial aspect was incorporating tangible objects. In the games I observed, I saw tangible objects such as toy cars, building blocks, stickers, and other activities that involved the child's gross motor skills. Having tangible objects not only allows the child to engage with real-world objects but also reduces screen time by avoiding tablet-based games. The interaction should be designed to be interactive and flexible towards the child's needs. Engaging activities ensure sustained interest and participation of the child. Additionally, the sessions should be well-structured to provide clear guidance and instructions about the interactions, with SLPs supervising to ensure effectiveness. The design of activities using a social robot should be such that they reduce the workload of SLPs and *enable them to get more utterances from the child* quickly and easily. These speech samples are critical for SLPs to evaluate the progress of the child in these therapy sessions.

The application of the robot and the designed interaction should be such that they provide more opportunities for SLPs to collect speech samples from the child by including turn-taking activities to extend conversations. The interaction should also promote follow-ups and more extended conversation opportunities, facilitating human-human connections by involving other interaction partners, such as parents. The interaction with the robot should avoid adding stressors to the child's environment and be structured to support the smooth attainment of speech goals by the end of the sessions.

### 4.4.3 Communication Design Considerations

In this subsection, I will discuss the core category of communication design considerations that were arrived at after applying C-GTM. When integrating a social robot into speech therapy sessions, it is essential to align the robot's verbal and non-verbal communication style with the standard expectations set for SLPs and parents to improve therapy outcomes. The robot can use slow and relaxed speech to aid in clarity and reduce pressure or can follow any specific fluency-shaping strategy [62, 114] for practice with the child. Demonstrating good listening skills by the robot through non-verbal gestures (backchannel signalling, e.g. through head nods [83]) can make the child 'feel heard' and could motivate them to speak more [62]. The robot's speech content should be specific and relevant to the context of the activity rather than generic as this ensures that the interactions are not boring and vague. Similar to the activities, even the speech should be natural and engaging, so that the child should feel comfortable during the sessions and wants to continue participating in further activities. Moreover, efforts need to be made to avoid a robotic tone as it is not the communication style that one wants the child to pick up from speech therapy. Encouraging and positive comments can boost the child's confidence and willingness to engage actively in the treatment [62]. Descriptive comments enhance the richness and educational value of the interaction for the child. Incorporating cultural elements familiar to the child (e.g. related to the language their parents speak at home) can make communication more relatable and comforting, helping to build a sense of connection. It is important to ensure that the child is exposed to these cultural aspects at home, allowing them to find comfort and familiarity in them.

Short and clear sentences ensure that the child comprehends the instructions easily and does not add stressors to the child. Finally, we recommend personalised aspects in conversations that help building rapport, making the interaction more meaningful and tailored to the child's individual expectations. For example, if the child had a birthday coming up, then the SLP started the conversation by talking about birthdays and more topics around the celebration of the event (DC-2). User profiles can be created for games

on the robot to collect specific information, such as names, birthdays, and best friends' names, etc. This information could then be inserted into template sentences used by the robot in games to enhance personalization. Note, the choice between personalization and customization needs to be considered carefully by HRI researchers, as I highlighted in the cultural context in the first study in Chapter 3.

## 4.5 Limitations and Future Work

In this section, I will discuss the limitations followed by future work that could be undertaken in the domain of applications of social robots in stuttering therapy of preschoolers. Note, the primary limitation was having few institutions to partner with in commutable distance, this led to some additional related limitations.

- Recruiting a sufficiently large sample of preschool children who stutter was challenging, along with obtaining parental and SLP consent to observe therapy sessions within a practical geographic proximity.
- The SLPs in this study conducted therapies exclusively in English, limiting the generalizability of the findings to non-English-speaking populations.
- The majority of the children observed were non-severe stutterers, which could have limited the amount of extra information extracted from the shadowing sessions. There is a chance we could learn more design considerations from sessions of preschoolers with severe stuttering. It is possible that it will be more difficult to get utterance from severe stutterers for speech samples for analysis and it would have been insightful to see how SLPs manage get the utterances in this case.
- The primary stakeholders for this stage of our project were the SLPs, with a focus on the strategies and different kinds of interactive games they included in their sessions. Requirements could change depending on the stakeholders the researchers are targeting to work with.
- The findings of our study need to be tested in practice to determine their effectiveness in improving the therapy experience and outcomes for both SLPs and their clients.
- Since our study focused on preschoolers, the guidelines may not be applicable to older children, as speech therapy techniques vary across different age groups.

In our future work, we plan to prototype games and applications based on the established design considerations and gather feedback from SLPs. After completing the co-design process, we will conduct studies involving preschoolers who stutter and SLPs integrating our solutions into their sessions. These evaluations will not only validate the findings of this study but also instill greater confidence in the continued exploration and refinement of these design considerations.

Through this chapter, we transitioned from inclusivity through cultural integration in HRI to promoting inclusivity by finding applications where social robots can be used in speech therapy. Moving forward to my third study in the next chapter, I would like to note that analyzing the data for the second study discussed in this chapter sparked the idea for the next study. During our interviews, ***SLP-1*** mentioned that they asked parents to slow their speech rate down when communicating with their child. The SLP also mentioned that changing our speech is difficult. This has been quoted in the Results section of this chapter. Slowing and relaxed speech for caregivers is also recommended in the handbook [62] that was referred to in extant texts in (DC-6). Given those insights, I started researching how a social robot can be used to slow the speech rate of adults, which is discussed in detail in the next chapter.

# Chapter 5

# Study 3: Investigating Human Speech Rate Alignment in Response to Slow Robot Speech

The primary motivation for this study was to explore the applications of social robots that could ultimately benefit preschoolers by helping their caregivers acquire slow speech. This focus helped me refine the research scope by examining recommendations to parents or caregivers by SLPs, as well as general guidelines for family interactions to facilitate fluency outlined in the stuttering therapy handbook [62]. Caregivers are advised to speak in a slow and relaxed manner with comfortable pauses in between to help change the environment for the child who stutters to promote smooth or fluent speech. In the literature, Mr. Fred Rogers is cited as an example of a person with a slow speaking rate [62, 122]. For instance, in the handbook on stuttering therapy [62], it was suggested to watch Fred Rogers' 2002 Dartmouth College Commencement Address[1] on YouTube as a reference. For this study, I also drew inspiration from Mr. Rogers for the Voice System's speaking style, and I tried to replicate his calm and controlled manner of speech. A slower, more relaxed speech pattern reduces stress and provides the child with more time to respond if there are sufficient pauses. However, altering one's natural speaking style can be challenging. To address the challenges of modifying one's speaking style, I explored whether the concept of Speech Entrainment (SE) could be utilised to induce slower and more relaxed speech in adults.

I conducted an in-person study in a lab with two sessions per participant, containing four and five tasks respectively. The tasks involved obtaining verbal responses from the

[1] https://www.youtube.com/watch?v=907yEkALaAY&t=442s

participants. The Voice System program that I had developed spoke in a slow and relaxed speech throughout these tasks. The tasks involved long-form open-ended responses, sentence repetitions, and three-sentence picture descriptions, and my primary goal was to study whether different tasks had different impacts on the speech metrics, and whether repeated sessions had any impact. The study had two conditions, with the Voice System program running either on the QT Robot and on a laptop (between-participant variable).

The speech audio data, questionnaire responses related to participants' experiences during the study, and their exposure to and usage of the English language were collected, along with a semi-structured interview at the end of the study. The results suggest that sentence repetitions have the most significant impact on speech metrics (articulation rate, pause rate, and pause ratio). Repeated sessions also had an increased impact on the speech metrics. Valence, Arousal and Control collected using the Self-Assessment Manikin questionnaire [22] before and after each session showed a decrease in the post-session rating of the second session compared to the post-session ratings of the first session. Additionally, responses to the interview questions about participants' opinions on the impact of slow speech and its usefulness suggest that slow speech could reduce engagement or have the reverse effect where participant starts speaking faster out of irritation. Novel application areas such as relaxation and announcement systems, as well as use for individuals who are non-native speakers or have speech impairments, could potentially benefit from slow and relaxed speech.

## 5.1 Research Questions

I frame my research questions to study SE in participants in a variety of verbal tasks across two sessions. I wanted to know if a particular kind of task had more impact than others or not. Additionally, I wanted to study the impact of multiple sessions on the speech metrics, so each participant took part in 2 sessions. Another goal was to compare the robot with a non-robotic Voice System, i.e. a laptop. To the best of my knowledge, the current research on SE in HRI or HCI has not compared laptops and robots as agents to induce slow speech. Not much work has been done to study how the nature of verbal tasks can impact SE. Knowing these are important for designing future applications and research studies to evaluate the effectiveness of social robots as models for fluency-shaping techniques, my objective was to address the following research questions

**RQ1:** How do human speech patterns evolve *between sessions* when interacting with a social robot or a laptop speaking at a slow speech rate?

**RQ2:** Is there a difference in the extent of human speech alignment when interacting with a *social robot versus a laptop* speaking at a slow speech rate?

**RQ3:** How does interacting with a social robot or a laptop speaking at a slow speech rate influence human speech patterns *across different types of speech tasks* (sentence repetition, three-sentence picture descriptions, and open-ended responses) within a session?

**RQ4:** Does the participant maintain a consistent speech pattern during a 'sales pitch', i.e. the task of "selling" an object to the social robot or laptop at the end of the second session, creating a degree of subtle pressure of the task?

**RQ5:** How do the interaction context (social robot vs. laptop) and slow speech rate influence participants' emotional responses (valence, arousal, and control)?

Effects of sound have been studied on humans [85]; when humans find a sound unpleasant, they tend to have negative reactions. Christenson et al. [35] explored how speech rate affects the likability of digital assistants. These studies inspired me to formulate RQ5.

For the most part, this study is exploratory in nature. For RQ1, we have a hypothesis (**H1**) based on existing research on speech entrainment and HCI cited in 2.3.1.

**H1:** Participants will slow down their speech rate after interacting with the voice system, which speaks in a slow and relaxed manner. The sentence repetition task is expected to have the strongest effect on speech entrainment, while possible effects of the other tasks are more exploratory in nature.

## 5.2 Method and Implementation

This section introduces the steps I took to evaluate the effects of the Voice System's slow speech on the participants' speech patterns. The design of this study intended to create scenarios where the participants were exposed to the Voice System's speech and engaged in neutral topics such as food and music to collect natural speech samples of the participants. The reasons for choosing these particular topics as themes are explained later in this section. To investigate speech entrainment between participants and the Voice System, I designed various tasks—including repeating a sentence spoken by the Voice Systems, describing a picture in three sentences, open-ended responses for a list of questions posed by the Voice System, and a sales pitch task. These different tasks were designed to keep participants engaged by incorporating interactive elements and variety in tasks without relying on advanced Large Language Models (LLMs) to hold actual conversations. Also, LLMs do not engage in conversation with a genuine understanding and rather generate statistically

likely continuations of text rather than meaningfully grounded responses [109]. So I did not consider LLM models due to the unpredictability of their responses, and also because they fell outside the scope of our research questions' primary focus.

This study used a mixed factorial design with one between-participant factor, where participants were assigned to one of two experimental conditions (Robot or Laptop) and a repeated measure within-participant design whereby data was collected in two sessions (Session 1 vs. Session 2). This allowed for a within-participant comparison across two sessions to study a possible change in their speech over time. The first four tasks in both sessions were the same, though the overall theme (food and music) of each session differed. Additionally, Session 2 included a fifth task, which will be discussed later in this section.

This study was designed to collect speech samples for each task without relying on any real-time detections, such as speech or facial expression analysis, during the session. To achieve effective speech sample collection, it was crucial that the user interface was intuitive and the instructions about the study were clear to participants. I built a web user interface with colour-coded sections—green, yellow, or red— with labels "Speak", "Prepare", or "Listen" to visually inform participants of the expectations. The Voice System also verbally explained the tasks in detail and repeated the instructions, along with the meanings of the different colours on the screen, before each task. This ensured that participants understood when to speak and when to stop and listen to the Voice System. It also helped prevent interruptions while ensuring participants had sufficient material to respond to the verbal tasks. The design of the web interface that can be seen in Figure 5.4 will be discussed in further detail when I explain the different tasks in the two sessions in section 5.2.2.

### 5.2.1 Artifacts

The Voice System program developed by me ran on either the QT Robot[2] or a Lenovo laptop with the Ubuntu operating system, representing the two experimental conditions in our study. I used the two conditions to study whether the embodiment of the Voice System has any effect on slow speech entrainment. The QT robot is commonly used in studies and applications in therapy contexts [10, 40, 116]. I additionally tried to explore whether we can achieve the same effect of verbal tasks using a laptop. If they both have similar effects, then future applications can also be designed considering that laptops are currently more common in households than robots.

[2] https://luxai.com/#LearnMore

The Voice System program used audio files generated for each speech output of the QT Robot, which were saved as WAV files on the robot. These files were then exported and stored on a laptop for use in the laptop condition. This ensured that the exact same audio was used in both conditions. I used an additional monitor in the robot condition to display information of the web interface. The laptop screen was comparable to the screen of the monitor that was used with the QT Robot. I opted for screens larger than tablet size, as clear visual instructions on the web interface were essential for the system to effectively communicate with participants during the various tasks (the experimenter was not present during the sessions). Audio data of the entire session was recorded. For enhanced and efficient audio sample collection, I used a DPA 4088 headset microphone[3], connected via a Fireface UCX II USB audio interface[4]. In Figure 5.1 we show some of the equipment that were discussed above. The audio interface was connected to a Microsoft Surface Book, which ran Audacity[5] to record the entire sessions of each participant. Using Audacity, I stored audio files in WAV format, sampled at 44.1 kHz in 24-bit mono channel audio. This setup ensured that the collected audio samples were clear, with minimal background noise and echo impact. Additionally, the sessions were recorded using a video camera for documentation and verification purposes, and might be used in future for further analysis. Qualtrics[6] was used to prepare the questionnaires and collect the responses. All questionnaires were digitally administered. I used my laptop to allow participants to fill pre-session, post-session and post-study questionnaires, as well as the remuneration form in the lab. The setup of all equipment can be seen in Figure 5.2.

### 5.2.2 Procedures and Measures

In this section, I will discuss in detail the session structure, various tasks in these sessions and the data we collected.

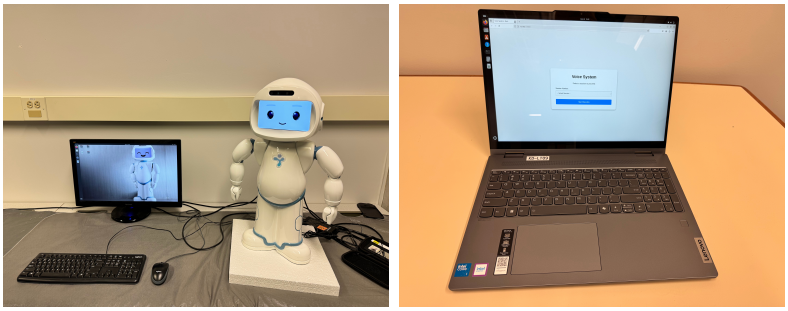
#### Pre-Study Questionnaires

Prior to the study, participants were provided with an online Information Sheet and Consent Form. Before the initial survey, participants were assigned to one of two conditions: either the robot condition or the laptop condition. They were also given a participant ID,

[3] https://www.dpamicrophones.com/microphones/headset-and-earset/4088?variant=17
[4] https://rme-audio.de/fireface-ucx-ii.html
[5] https://www.audacityteam.org/
[6] https://www.qualtrics.com/



(a) Headset microphone (b) Video camera (c) USB Audio interface



(d) QT Robot (e) Lenovo Laptop

Figure 5.1: Images of some of the equipment used in the study



Figure 5.2: The setup for Session 2 in Robot Condition. The setup for Session 1 was the same, except there were no objects on the table on the right.

which they used to complete all questionnaires in this study. This ensured that participant names were not linked to the collected data. Those who agreed to participate were redirected to a web form that collected demographic information, which asked about their age and gender. The questionnaire also included the widely used Ten-Item Personality Inventory (TIPI) [57], designed to assess the Big Five personality dimensions, see Table 5.1. The responses were recorded on a 7-point Likert scale ranging from 1 (strongly disagree) to 7 (strongly agree).

### Scheduling of Sessions

Participants were scheduled for two in-person sessions. Both sessions were held on the same day with a minimum of four hours gap in between. We decided to have both sessions on the same day because we wanted to retain the experience of the slow speech from the first session, rather than having a fresh start on a different day. We assumed that this design would be more effective in observing the possible effects of slow speech. This gap was included to prevent participants from getting bored due to the (almost) identical nature of the sessions and to avoid possible fatigue from listening to the Voice System's slower

Table 5.1: Ten-Item Personality Inventory (TIPI)

| Item | I see myself as: |
|---|---|
| 1. | Extraverted, enthusiastic. |
| 2. | Critical, quarrelsome. |
| 3. | Dependable, self-disciplined. |
| 4. | Anxious, easily upset. |
| 5. | Open to new experiences, complex. |
| 6. | Reserved, quiet. |
| 7. | Sympathetic, warm. |
| 8. | Disorganized, careless. |
| 9. | Calm, emotionally stable. |
| 10. | Conventional, uncreative. |

than normal human speech. The four-hour gap was chosen for practical reasons as all participants were students. This allowed the students to attend a class or complete a task during the break while also making scheduling sessions more convenient.

**Session Structure:** Each session followed a similar structure, with the first session consisting of four tasks and the second session including five tasks. The first four tasks were identical across both sessions. Apart from collecting audio and video data during the sessions, we also asked participants to complete a pictorial Self-Assessment Manikin (SAM) [22] questionnaire both before and after each session (see Figure 5.3). It was used to evaluate affective dimensions: valence, arousal, and control. Participants responded to the following questions: "How do you feel in general right now?" (valence, top panel), "How energised do you feel?" (arousal, middle panel), and "How in control do you feel in the current situation?" (control, bottom panel). For each question, participants selected one option from A to E in the corresponding panel. SAM was chosen for its intuitive, language-independent format that allowed participants to quickly and easily report their emotional state without relying on verbal descriptions [22].

The two sessions of the study were based on a specific theme. Food was chosen as a theme for the first session, and music for the second. These specific themes were selected because participants can be expected to have an interest in these topics. Also, they are neutral, uncontroversial, and engaging while also providing an opportunity to frame questions or tasks that I anticipated would encourage participants to produce more utterances for the speech samples during the sessions. We wanted the participants to speak naturally as much as possible for the entire duration allotted to them.

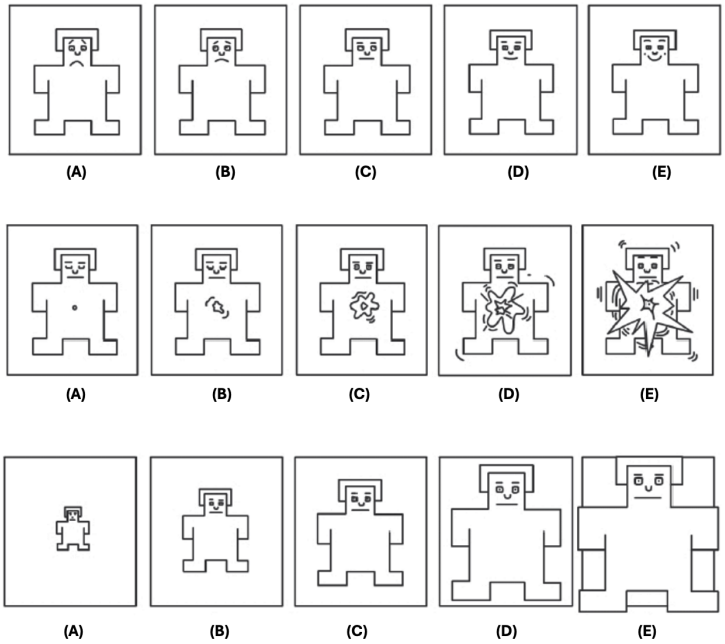



Figure 5.3: The Self-Assessment Manikin - valence (top panel), arousal (middle panel), and control (bottom panel).

The Voice System provided clear verbal prompts throughout all the tasks to indicate whether it was time to prepare, speak or listen. To reinforce these instructions visually, I also used a screen that displayed a colour-coded section. Red with the phrase 'Listen,' yellow with 'Prepare' and green with 'Speak' written at the centre of the colour-coded section were used to signal to participants whether they should speak, listen, or prepare and answer. The Voice System thoroughly explained these colour states before each task started, ensuring participants understood the workflow. During the test runs of the study, I received feedback from SIRRL labmates that they felt annoyed when the Voice System verbally interrupted them after their speaking time slot for the task had ended. A slight change was therefore made to the design, adding a five-second delay for tasks 1, 3, 4, and 5 and a two-second delay for task 2 between the moment the screen turned red—signalling to the participant to stop speaking—and when the Voice System instructed the participant to stop speaking. With this approach, the participant is not verbally interrupted by the Voice System and has some time to realise they need to stop speaking. For the entire script containing what the Voice System spoke during different tasks in both sessions, please refer to the Appendix D.

### Task 1: Introduction and Initial Speech Sample

At the beginning of the first session, I gave participants a brief explanation of the study to remind them of the procedure they had previously read in the information sheet. In the first task I handed them a sheet with the instructions for the task and the following questions to get their initial speech sample for that session to serve as a baseline. The instruction sheet for both sessions can be found in the Appendix D.

**Session 1:** *What's your name, and where did you grow up? Could you tell me a bit about the family members or friends you spent time with as a child? I'd love to know about the schools you went to and any favorite places you enjoyed exploring. Were there any special dishes or snacks you loved that remind you of home? Take a moment to think about these memories, from the people who shaped your world to the places and flavors that made it special.*

**Session 2:** *What program are you enrolled in, and which term are you in? Could you tell me a bit about the courses you're taking this term and whether any of them are particularly interesting or challenging? Now, here's something a bit more playful—if you could step into the life of any celebrity, who would you choose to become? Take a moment to think*

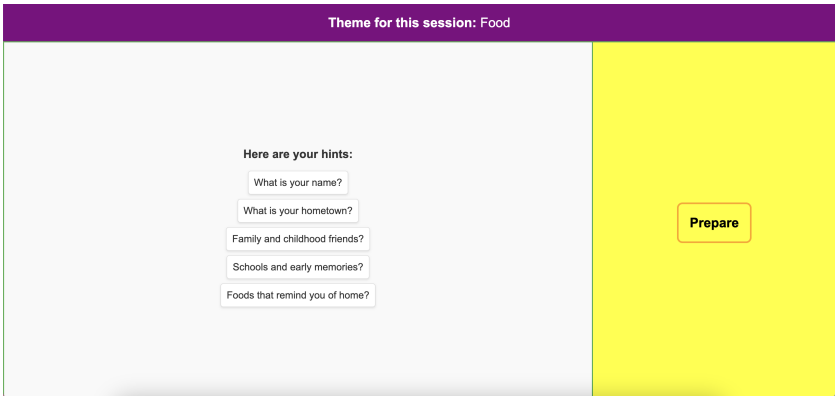



Figure 5.4: Web User Interface for Session 1 Task 1

*about what draws you to them, from their personality or accomplishments to the values they represent or the life they lead.*

The sheet also provided information about different colour schemes, their meanings, and the preparation and response times. Afterward, participants engaged with the Voice System, which delivered structured interactions. Once a participant had finished reading the Task 1 sheet and completed the pre-session questionnaire, I turned on the video and audio recording devices. A test audio sample was collected before I started the Voice System program and then left the room where the study was conducted. I waited in the nearby corridor, and I could hear the Voice System's voice from outside. The program ran for approximately 24 minutes in the first session and 28 minutes in the second session. I re-entered the room only after I heard the Voice System say the last sentence for that session while I waited in the corridor.

At the start of the session, the screen displayed a yellow background with the word "Prepare" written on it. Participants were instructed to prepare their responses for 60 seconds and then speak for 90 seconds. The screen showed short versions of the questions that participants had already read on the sheet, so they didn't have to remember them while preparing or answering. The web interface with short hints for the questions to respond to, along with the colour-coded section, can be seen in Figure 5.4. After the 60-second preparation period, the Voice System verbally instructed the participant to begin responding to the questions from the sheet. In the robot condition, each verbal output was accompanied by a suitable gesture based on the speech content.

### Task 2: Sentence Repetition

In the second task, participants were instructed to repeat ten six-word sentences within 15 seconds after the Voice System finished speaking each sentence. I chose six-word sentences because they are easy to remember. This was inspired by Miller's Law [84], which states that a person can keep $7 \pm 2$ objects in their working memory at a time. A sample sentence for each theme would be, "Fresh apples are perfect for quick snacks" and "Soft melodies help clear your thoughts". The entire list of sentences with information about different tasks can be found in Appendix D.

### Task 3: Picture Description

The third task involved a picture description activity, where images appeared on the screen, and the Voice System and participants took turns describing them in three sentences each. Participants were allocated 25 seconds for their turn. The task featured four unique pictures for the Voice System and four different pictures for the participants. I used DALL.E 3 [7] to generate the images and a sample of them for each theme can be seen in Figure 5.5. Sample prompts used for Session 1 with the theme of food are - "Friends sharing a pizza at a cozy restaurant" and "A bowl of hot chicken soup on a kitchen counter". Similarly, for Session 2 with the theme of music, the sample prompts were - "A boy sitting on a bench, focused on playing his guitar" and "A man sits on the grass, happily playing his guitar for a group of friends". Other prompts were of the same format, just food items or musical instruments, along with the location context, were changed.

Given the consistent central themes (food or music) and the gradual increase in how open-ended the responses needed to be from tasks 2 to 4, this design aimed to help participants feel more comfortable sharing their preferences in detail in Task 4. The session began with sample speech collection in Task 1 to serve as a baseline. Task 2 involved sentence repetition, followed by tasks that increasingly required participants to produce more open-ended speech.

### Task 4: Preference Discussion

In the fourth task, the Voice System shared its preferences related to the central theme for approximately 4 minutes and then posed questions to the participants about their own preferences. Participants were instructed to prepare their responses for 60 seconds before

[7] https://openai.com/index/dall-e-3/



Figure 5.5: Sample pictures shown for Session 1 and Session 2 on themes food and music.

speaking for another 90 seconds. Tasks 1 to 4 were the same in structure across the two sessions, but the content differed based on the theme of each session.

To help recall the questions, the screen displayed concise versions of the questions the Voice System had asked about their preferences (similar to Task 1). The goal was to analyze natural speech samples from the first to the fourth task to study speech alignment with the robot's slow speech. Tasks two and three were designed to be interactive and engaging, helping to put participants in a comfortable zone where they could speak more naturally about the central themes (food or music).

### Task 5 (Session 2 Only): Sales Pitch Task

The second session included an additional fifth task, where the Voice System instructed participants to prepare a sales pitch to sell one or more items placed on a table to their right. There were six objects on the table: an Apple iPhone XR (empty box), Apple AirPods Pro 2 (empty box), a WD SSD hard drive (empty box), an Alexa Echo Dot (empty box), a Tim Hortons Hot Chocolate Mix, and a packet of naphthalene mothballs. I chose objects which had a label (with ingredients or product specifications) behind them so that the participants had material to prepare the sales pitch.

This task was not disclosed in the information sheet beforehand and was designed to observe whether this impromptu, slightly stressful activity influenced participants' speech patterns and any speech alignment with the Voice System, if present, was retained. Participants were given 60 seconds to prepare their responses before speaking for another 90 seconds.

**Confidence Assessment**

Along with the SAM questionnaire, participants were asked to rate their confidence in responding to the Task 1, Task 4, and Task 5 (Session 2 only) questions after each session. We had asked the following questions:

- (Session 1) How confident were you in answering the questions about your personal background and early memories at the start of this session?
  - Continuous bipolar slider (0 = A little, 100 = A lot)
- (Session 1) How confident were you in answering the questions about your food preferences and dining habits at the end of this session?
  - Continuous bipolar slider (0 = A little, 100 = A lot)
- (Session 2) How confident were you in answering the questions about your academic journey and celebrity inspiration at the start of this session?
  - Continuous bipolar slider (0 = A little, 100 = A lot)
- (Session 2) How confident were you in answering the questions about your music tastes and aspirations at the end of this session?
  - Continuous bipolar slider (0 = A little, 100 = A lot)
- (Session 2) How confident were you in the sales pitch preparation challenge at the end of this session?
  - Continuous bipolar slider (0 = A little, 100 = A lot)

This step was important to assess the impact of their confidence during these tasks on their speech metrics. Responses were collected using a Qualtrics form.

**Post-Study Questionnaire**

After completing all sessions, participants were asked to complete a post-study questionnaire on Qualtrics designed to gather their perceptions of the Voice System, their self-reported language proficiency, their own speech perception, and additional factors, which are discussed in detail below. It should be noted that the numerical values on the bipolar sliders were hidden from the participants in the on-screen questionnaire.

The questionnaire included several components:

- How fast or slow did you perceive the Voice System's speech rate?
  - Continuous bipolar slider (0 = Very slow, 100 = Very fast)
- They evaluated the Voice System's speech using the Godspeed Likability Scale [13], where participants used a continuous bipolar slider ranging from Dislike (value 0) on the left to Like (value 100) on the right, as well as other paired descriptors such as Friendly to Unfriendly, Unpleasant to Pleasant, Unkind to Kind, and Nice to Awful. For analysis, I take the average of these ratings and store it as 'Overall Likability'.
- To assess the feeling of social presence during the interaction (borrowed from [79]), participants responded to questions such as:
  - How much did you feel as if you were interacting with an intelligent being?
    - Continuous bipolar slider (0 = Not at all, 100 = Absolutely)
  - How much attention did you pay to it?
    - Continuous bipolar slider (0 = Not at all, 100 = Absolutely)
  - How much did you feel involved in the interaction with it?
    - Continuous bipolar slider (0 = Not at all, 100 = Absolutely)

Participants provided information about their language proficiency. I asked them whether English was their first language, their fluency level, how long they had been speaking English and other similar questions. I asked:

- Is English your first (native) language?
  - Radio Button (Yes/No)
- How fluent are you in speaking English?
  - Multiple Choice:

  - Basic knowledge (Basic words and phrases; simple, short conversations)
  - Intermediate (Converses on everyday topics; some hesitations)
  - Advanced (Comfortable in complex conversations; minor limitations)
  - Completely fluent (Near-native proficiency; fluency in any context)

- How long have you been speaking English?
  - Radio Button (Less than 1 year, 1-2 years, 3-5 years, 6-10 years, More than 10 years)

- How often have others told you that you speak too fast?
  - Multiple Choice (Never, Rarely, Sometimes, Often, All the time)

- How often have others told you that you speak too slowly?
  - Multiple Choice (Never, Rarely, Sometimes, Often, All the time)

- In which contexts do you frequently speak English?
  - Checkbox (Select all that apply):
    - Formal (e.g., At work, In academic settings, In professional meetings)
    - Informal (e.g., Socially with friends, At home, Casual conversations)

I also asked additional questions about their experience with speaking English. These questions were based on their experience with speaking English, assuming that this could have impacted their speech samples, which were in English. I asked:

- How often do you think in English when speaking the language?
  - Multiple Choice (Never, Rarely, Sometimes, Often, All the time)

- Do you currently or have you ever had any speech disfluency?
  - Multiple Choice (Yes, currently / Yes, in the past / No / Prefer not to answer)
  - Text Box (if "Yes" selected, optional)

- How comfortable do you feel when speaking English?
  - Slider (0-100) (1 = Not comfortable at all, 100 = Extremely comfortable)

- How often do you feel nervous or anxious when speaking English?
  - Multiple Choice (Never, Rarely, Sometimes, Often, All the time)

- What is the main accent or dialect of English you have been exposed to?
  - Text Box
- Have you lived in an English-speaking country?
  - Radio Button (Yes/No)
- If yes, how long have you lived in an English-speaking country?
  - Radio Button (Less than 1 year, 1-2 years, 3-5 years, More than 5 years)

Note, the above questions were asked towards the end of the study since, if asked before the sessions, they might have made participants self-conscious of their own speech.

### Semi-Structured Interview

I conducted a semi-structured interview to understand participants' experiences with the Voice System, focusing on their experience during the study. The interview was audio-recorded for transcription and qualitative analysis.

Participants were informed that the system was intentionally designed to speak slowly, and questions were as follows:

- Should the Voice System change the way it speaks in terms of speed? If yes, then how and why?
- Do you think a slow and relaxed speech could be useful and why?
- To what extent do you think a Voice System's speaking style can impact how you speak (scale 0-100) and why?

### Debriefing and Compensation

After the interview, participants were debriefed using a prepared script (see Appendix D). At the end of the study, we thanked the participants for their participation. Participants received 30 CAD for full participation in the study and a prorated amount for partial participation as compensation for their time. However, all participants completed the study from start to finish.

### 5.2.3 Participants

I advertised our study using a flyer (see Appendix D), seeking university students to participate in our research. In total, 63 participants participated in this study (laptop condition = 31 and robot condition = 32). To ensure consistency in what participants experience in the sessions and the data quality, I had to drop 2 participants, one from each group, initially. In one case, I selected the incorrect recording device, and in the second case, the participant altered the volume settings of the laptop. After this filtering process, 61 participants remained (laptop condition = 30 and robot condition = 31). There were 16 females and 14 males in the laptop condition. Their age was in [18, 28] with $M_{\text{age}} = 22.27$ and SD = 2.53. In the robot condition, there were 16 females and 14 males, and 1 was in the 'other' category representing non-binary groups. The age range in this condition was in [18, 40] with $M_{\text{age}} = 23.97$ and SD = 4.51. We can see the distribution of age in both conditions in Figure 5.6.

For the remaining 61 participants, we had 9 rows (or data points) each—4 data points for the 4 tasks in Session 1 and 5 rows for the 5 tasks in Session 2. Columns from the other questionnaires were duplicated across all rows for each participant for ease of analysis. After initial filtering, we had $61 \times 9 = 549$ data points. I filtered out data points where the articulation rates, speaking durations, or number of syllables were outside the inter-quartile range for the speech samples of that task, and then manually verified these outliers. Since the study involved multiple tasks, with tasks 2 and 3 including subtasks, some participants did not complete certain tasks or subtasks as instructed. As a result, I excluded the corresponding data points. In total, six data points were dropped: one each from Task 2 and Task 3 in Session 1, one from Task 2 in Session 2, and three from Task 5 in Session 2. Two of these six data points that were dropped were from the same participant, and the other four were from separate participants. Finally, we had 543 data points.

### 5.2.4 Data Processing and Analysis

Before we discuss how I processed and analyzed the data, it is important to revisit the types of data we have. This study gathered data through questionnaires and audio recordings.

- **Questionnaires:**
  - **Pre-study questionnaire:** Collected demographic information such as age, gender, and TIPI.
  - **Session-based questionnaires:**

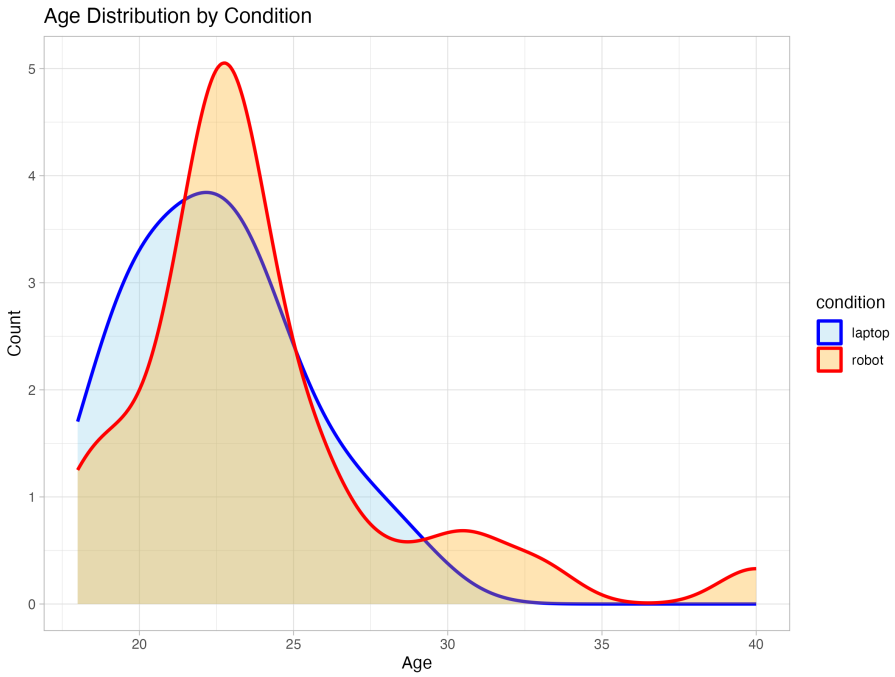



Figure 5.6: Age distribution of Participants by Condition

    - ∗ **Self-Assessment Manikin (SAM):** Collected before and after each session.
    - ∗ **Confidence assessment:** After the sessions, the participants rated their confidence in responding to Task 1, Task 4, and Task 5 (Task 5 was only in Session 2).
  - **Post-study questionnaire:** Collected at the end of the study, discussed in detail in the previous sections.

- **Audio Recordings:**
  - **Session Recordings:** Audio samples were collected during the sessions to capture participants' verbal responses for each task in both sessions. Entire sessions were recorded, and relevant samples were later extracted.
  - **Interview Recordings:** Interviews were audio recorded for transcription and qualitative analysis.

Complete audio recordings of each session were collected and analyzed. Since all tasks followed a predefined script, the expected speaking moments were identified. The timestamp when the voice system first spoke was marked, and a Python script extracted participant responses from the audio recordings. Finally, manual verification was conducted by me to ensure accuracy. After extracting the speech segments, I applied a high-pass Butterworth filter (80 Hz stopband, 100 Hz passband) to remove low-frequency noise and enhance speech clarity. Using SciPy[8] and pydub[9], the script extracted audio files and filtered out unwanted frequencies while preserving the speech signal. Each segment was loaded, filtered, and saved in a clean directory structure, ensuring consistency across all participants and sessions.

To extract speech metrics, I used the Praat Syllable Nuclei v3 Script [43]. Using Praat, I calculated the **articulation rate**, **pause ratio**, and **pause rate** for each audio file that was extracted for each task. Praat detects syllable nuclei and speech intervals. I recorded two audio clips in the lab—one while the robot was speaking and another when the room was silent—to analyze background noise using Welch's Power Spectral Density plots. I consulted with Professor Ewen MacDonald, a speech analysis expert, to determine the appropriate parameter values for the previously discussed high-pass Butterworth filter and also for the Praat script. I had passed the following parameters to the script: a silence threshold of $-15$ dB, a minimum dip of 2 dB, and a minimum pause duration of 0.20

[8] https://scipy.org/
[9] https://pypi.org/project/pydub/

seconds. I recorded an audio clip in the lab when the study was not in progress and the room was unoccupied, to capture the ambient background noise.

Using a Python script, each audio file was processed through the Praat script to extract the following metrics:

- **Syllable count** ($N_{syllables}$): The total number of detected syllables.
- **Speaking duration** ($T_{speaking}$): The total duration of all speech intervals (in seconds).
- **Pause duration** ($T_{pause}$): The total duration of silent gaps between speech intervals (in seconds).
- **Total utterance duration** ($T_{total}$): The time from the start of the first utterance to the end of the last utterance.
- **Number of pauses** ($N_{pauses}$): The count of silent pauses between speech intervals.

Using these values, I computed the following key metrics:

**Articulation Rate** ($AR$) measures the rate of syllables produced per second during speech-only segments and is defined as:

$$AR = \frac{N_{syllables}}{T_{speaking}} \tag{5.1}$$

**Pause Rate** ($P_{rate}$) indicates the number of pauses per second:

$$P_{rate} = \frac{N_{pauses}}{T_{total}} \tag{5.2}$$

**Pause Ratio** ($PR$) represents the proportion of time spent pausing relative to the total utterance duration:

$$PR = \frac{T_{pause}}{T_{total}} \tag{5.3}$$

These metrics provide insights into the speaking patterns across different participants and tasks. For Task 2, we had ten separate sentence repetition recordings per session, and for Task 3, we had four separate picture descriptions. For each subtask (in Tasks 2 and 3), I computed the average of these metrics for that specific task within the session. Finally,

I prepared a single dataset by merging the speech metrics dataset with the questionnaires dataset, ensuring all collected data was structured for statistical analysis.

To ensure consistency in categorizing participants' self-reported English accents, they were then grouped into broader categories. North American English (e.g., American, Canadian, Greater Toronto Area (GTA)), British English, Indian English, Asian English (e.g., Chinese, Korean), Persian/Iranian English, Other (for accents not fitting predefined categories), and Unspecified (for missing or unclear responses). This standardized classification helped maintain consistency in analyzing participants' language backgrounds.

Before discussing the various models used for statistical analysis, it is important to first list the different types of data I had collected. This will decide the transformations in data or selection of models. The dataset consists of **numerical**, **categorical**, and **ordinal** variables. An example of ordinal data would be for the question: *How often does the participant speak in English in their daily life?* with response options: *Never, Rarely, Sometimes, Often, All the time.*

We can treat categorical variables as factors in **Linear Mixed-effects Models (LMMs)** [15]. However, for ordinal predictors, we cannot treat them as categorical, as that would result in a loss of information since we have a ranking between various options: *Never* < *Rarely* < *Sometimes* < *Often* < *All the time.* We also cannot simply assign them numerical values and treat them as numeric variables (1, 2, 3, 4 and 5), as that would assume a linear relationship. To properly deal with this, we can treat ordinal data as **ordered factors**, allowing the LMM model to estimate relationships where it assumes linear, quadratic, cubic, or fourth-order trends with ordinal variables. These relationships in some cases will be listed in the reported tables of model outputs. To determine the simplest model that best fits the data, I first treated each ordinal variable as numeric and then as an ordered factor, compared model fits using ANOVA, and selected the model with the best fit.

In the case of SAM when analyzing post-session emotional states, the ordinal variable is a response variable instead of a predictor. Therefore, I used a **Cumulative Link Mixed-Effects Model (CLMM)** [34] instead of an LMM.

For predicting some post-study questionnaire responses, such as overall likability, how intelligent the participants found the Voice System, how much attention the participants paid to the Voice System, and how involved they were in the interactions, I used a regular Linear Regression Model since we had one observation per participant and no repeated measures.

In order to examine the data distribution and normality, I plotted histograms for continuous data and bar plots for categorical data that we plan to predict through our statistical models (see Figures 5.7, 5.8 and 5.9). We can also refer to the Q-Q plots for continuous

data in Figure 5.10. For most plots the data looks approximately normally distributed. In order to choose the simplest LMM model that satisfies my explorations towards the research questions, I will consider factors that would reasonably be expected to impact the speech metrics. The distribution for Accent Categories, as seen in the Figure 5.11, shows that the data is highly imbalanced among different classes. On the other hand, the distribution for native and non-native speakers (see Figure 5.12) is more balanced and will serve the same purpose of studying the impact of being a non-native or native speaker on the speech metrics of participants. I accounted for individual differences by including participant-specific random intercepts. To reduce the Akaike's Information Criterion (AIC) [21], only a subset of these factors were kept in the models. I validated the models by standard diagnostic tests, including the examination of residuals through histograms and Q-Q plots for assessing normality, and scatter plots of residuals versus fitted values to check for homoscedasticity, ensuring that the assumptions underlying linear regression were met.

## 5.3 Results

In this section, I report the results of statistical analyses using linear mixed-effects regression models, cumulative link mixed models, or standard linear regression models, depending on the nature of the response variable being predicted. Note that I have not reported or discussed the effects of Gender for the "Other" category, as this group included only one participant, compared to 32 females and 28 males. Although the models included this category, statistical significance based on a single data point is not interpretable and was therefore excluded from further discussion. I have also reported the results of the thematic analysis of interviews with participants at the end of the study.

In the analysis of this study, I ran multiple tests on similar metrics such as speech metrics (articulation rate, pause rate, and pause ratio), emotional states (valence, arousal, control), and experience ratings by participants (likability, perceived intelligence, and attention and involvement) with the Voice System that increases the risk of introducing Type I errors. In consultation with another researcher at SIRRL, I considered statistical results to be significant only when $p < 0.01$ rather than $p < 0.05$. This new threshold is comparatively stricter. I will report all the results but discuss only those factors with a p-value below this threshold as significant.

Additionally, for categorical predictors such as tasks and ordinal predictors with multiple levels (e.g., SAM ratings), I applied the Bonferroni correction [19] by adjusting p-values

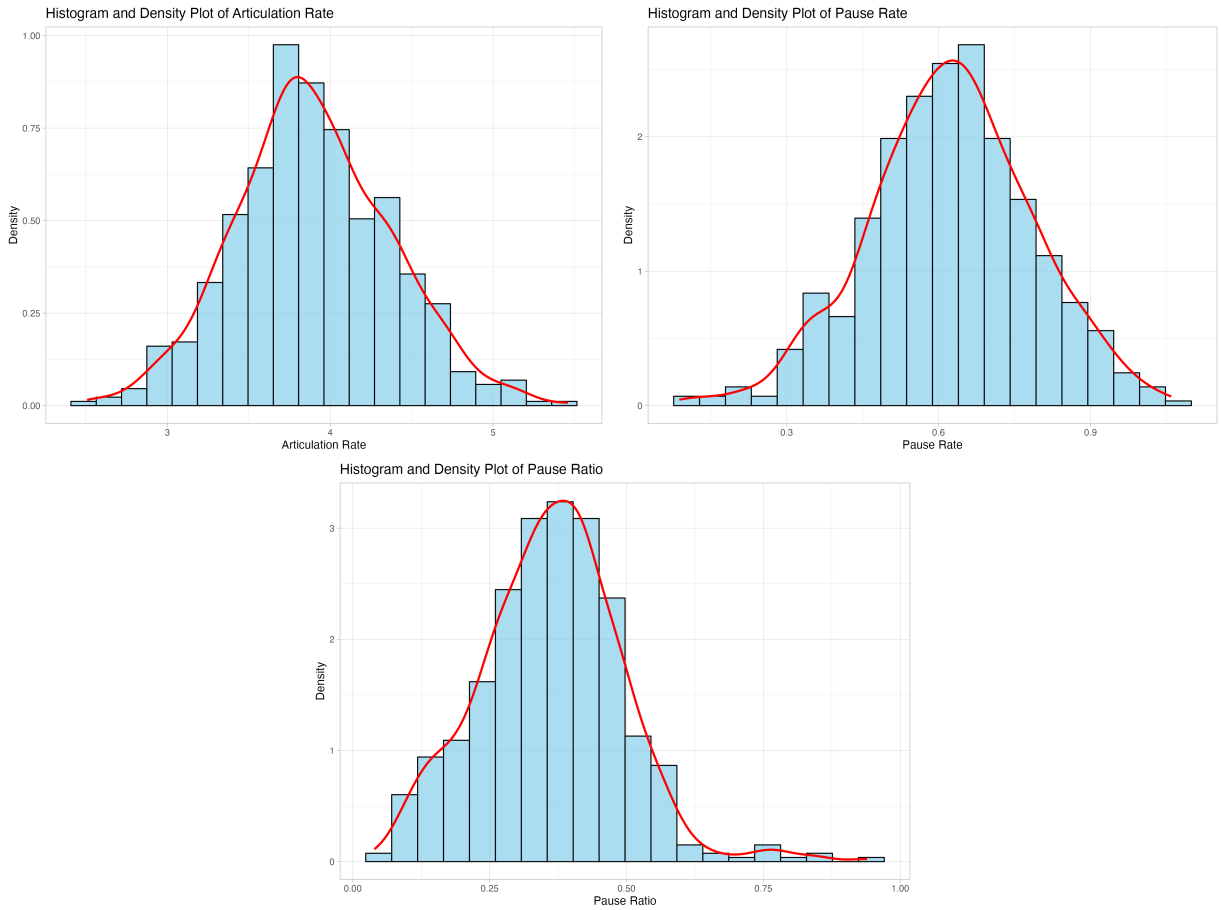



Figure 5.7: Histogram and Density Plot for Articulation Rates, Pause Rates and Pause Ratios of Participants

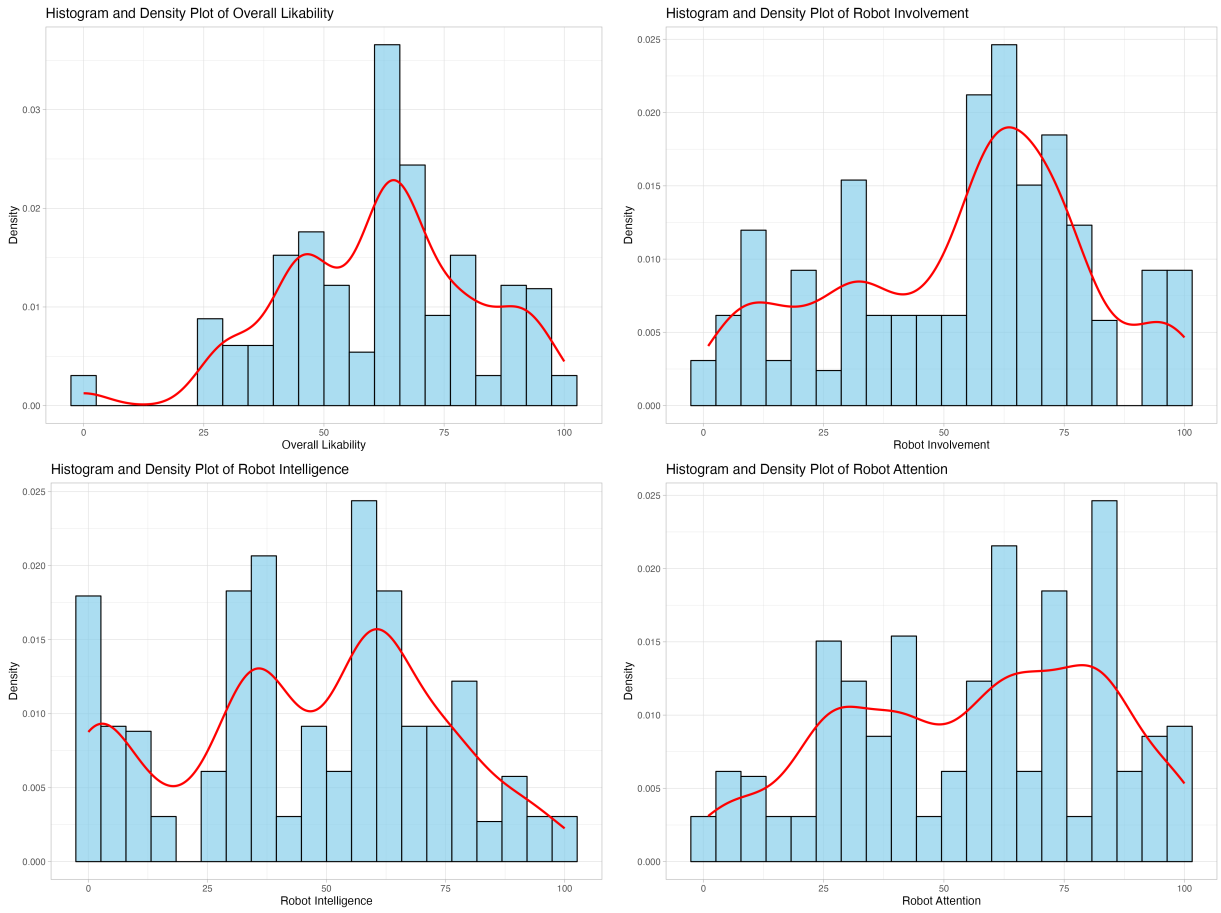



Figure 5.8: Histogram and Density Plot for Overall Likability, Involvement with the Voice System, Perceived Intelligence of the Voice System and Attention Towards the Voice System

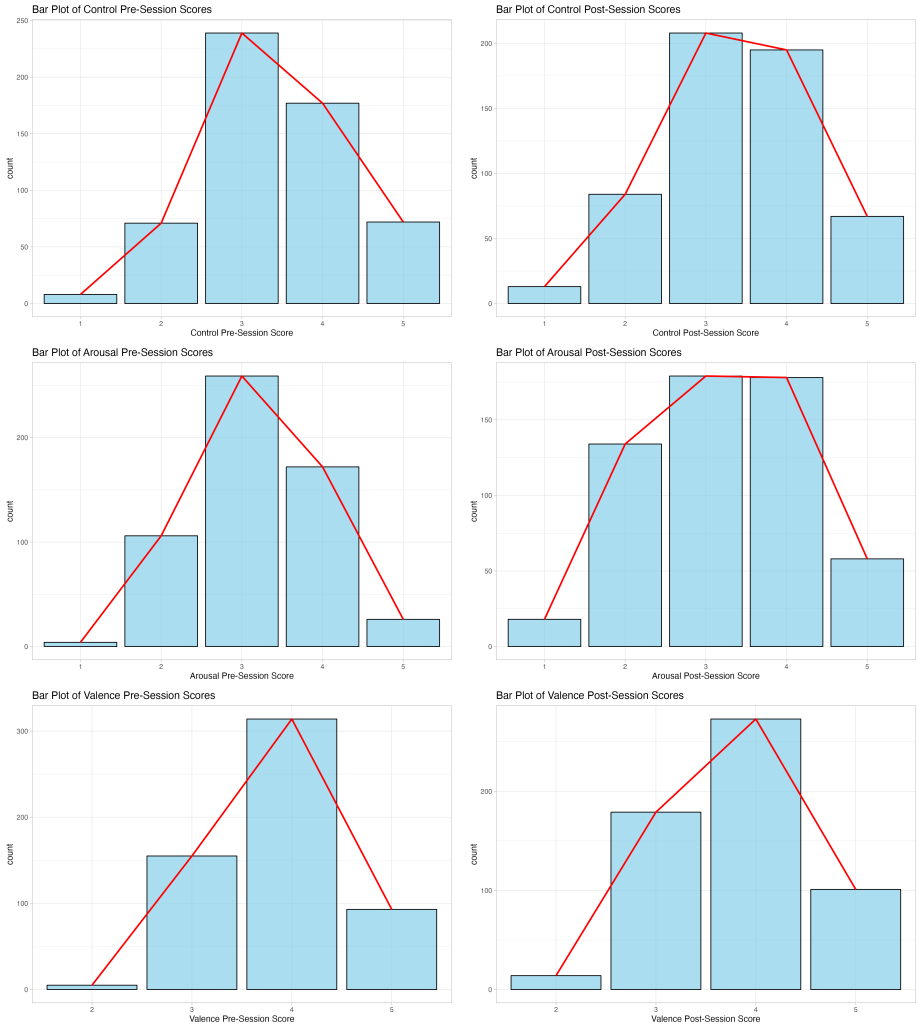



Figure 5.9: Bar Plots for Control, Arousal, and Valence Scores for both Pre- and Post-Session

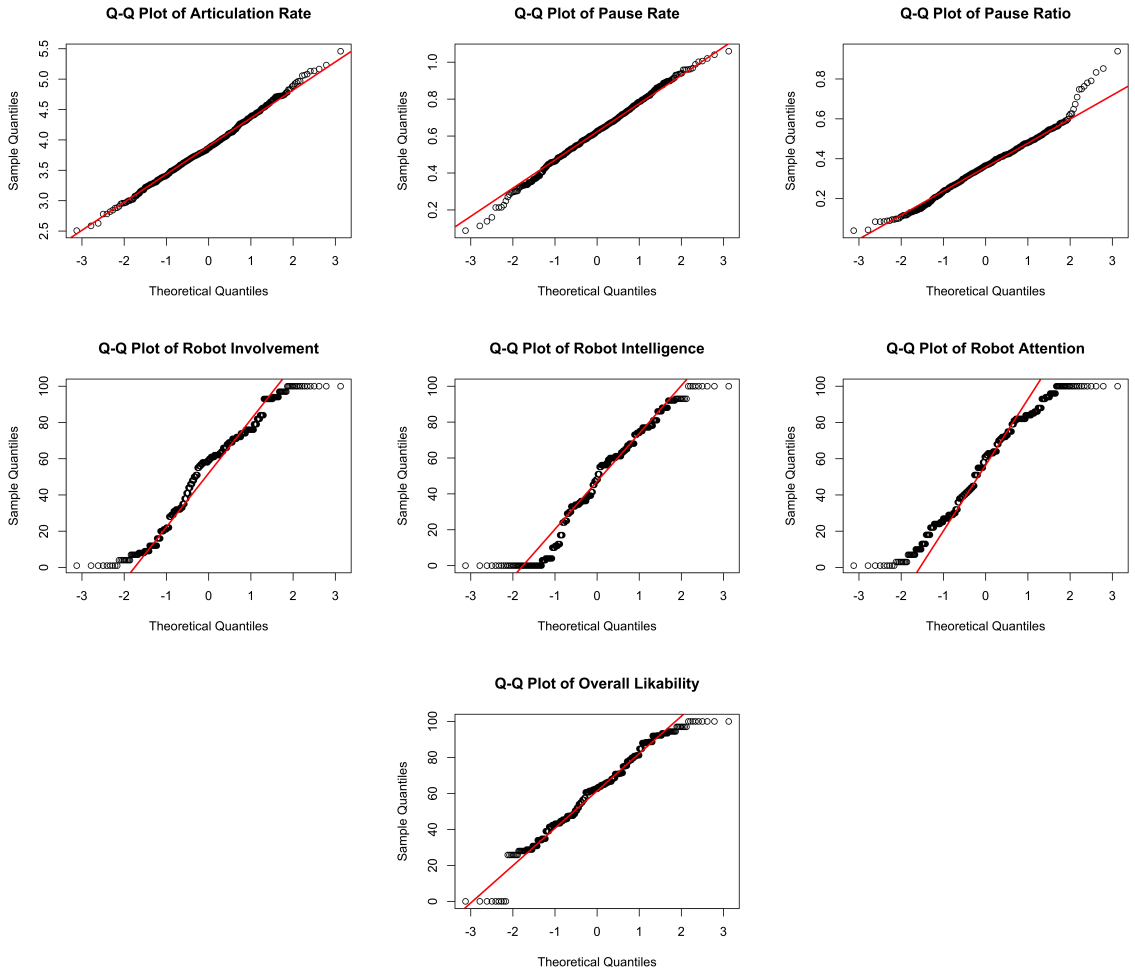



Figure 5.10: Q-Q Plots for Articulation Rate, Pause Rate, Pause Ratio, and Perceived Characteristics of the Voice System

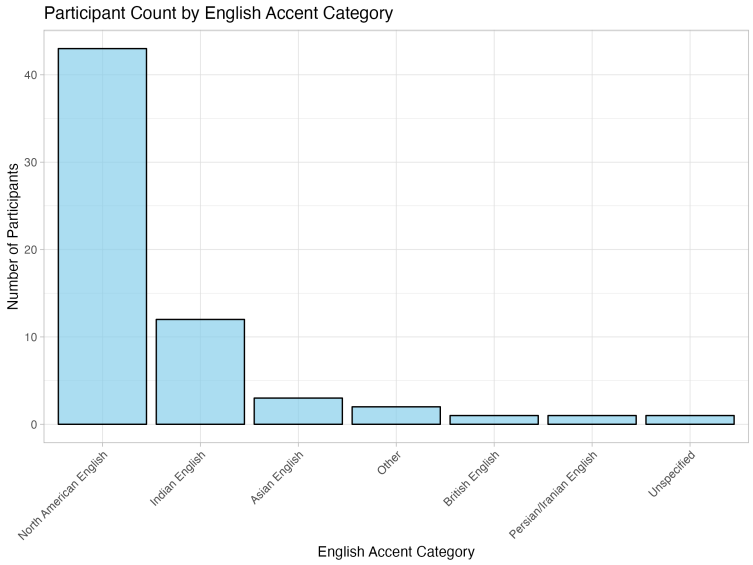



Figure 5.11: Participant count by English Accent Category

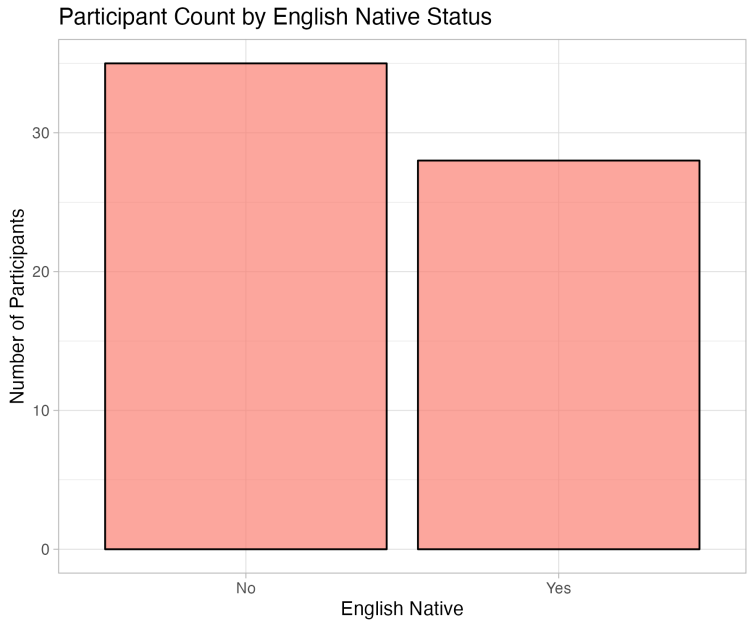



Figure 5.12: Participant count by English native language status

Table 5.2: LMMs predicting articulation rate in Session 1

| Covariate | Estimate | SE | $t$ | $p$ (Corrected) |
|---|---|---|---|---|
| Task 2 | -0.182 | 0.047 | -3.866 | *** |
| Task 3 | -0.180 | 0.047 | -3.808 | *** |
| Task 4 | 0.158 | 0.047 | 3.366 | *** |
| Voice System Speech Speed | 0.006 | 0.003 | 1.959 | . |
| Voice System Intelligence | -0.005 | 0.002 | -2.987 | ** |

.: $p < .1$; *: $p < .05$; **: $p < .01$; ***: $p < .001$. Reported p (Corrected) values are Bonferroni-adjusted.

according to the number of repeated tests for those factors. For example, if session 1 includes four tasks, then Tasks 2, 3, and 4 are compared against Task 1. Since this involves three comparisons, the raw p-values were multiplied by 3 (capped at 1) to control for the increased risk of Type I error. Following these correction measures makes the results more reliable.

### 5.3.1 Articulation Rate Models

We can see the effect on Articulation Rates of participants across tasks in both sessions in Figure 5.13.

**Session 1:**

As seen in Table 5.2, articulation rate significantly varied across tasks, with participants speaking more slowly in Task 2 ($SE = 0.047$, $t = -3.866$, $p < .001$) and Task 3 ($SE = 0.047$, $t = -3.808$, $p < .001$) compared to Task 1 which serves as the baseline sample. Articulation rate was significantly higher in Task 4 ($SE = 0.047$, $t = 3.366$, $p < .001$). There was a positive relationship between the Voice System's speech rate and participants' articulation rate, but without reaching significance (SE = 0.006, t = 1.959, p = .055) based on the new threshold of $p < 0.01$. Perceived intelligence of the Voice System negatively impacted articulation rate ($SE = 0.002$, $t = -2.987$, $p = .004$), suggesting that participants spoke more slowly when they perceived the Voice System as more intelligent.

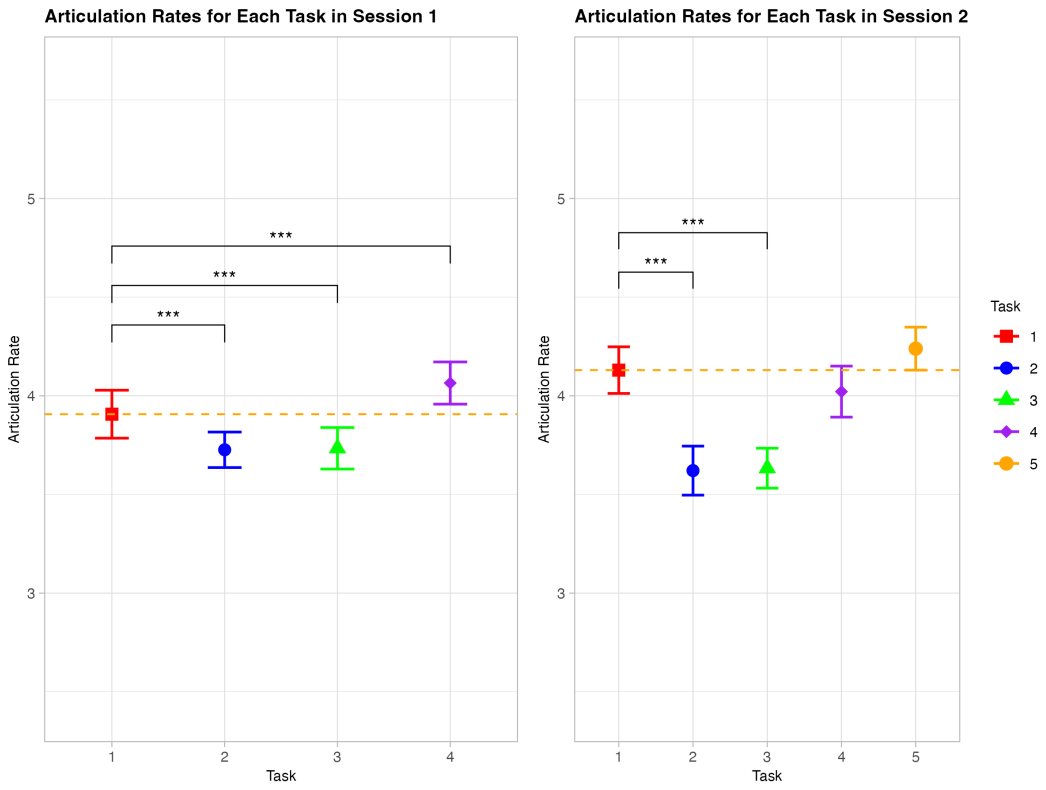



Figure 5.13: Participants' articulation rates across tasks for both sessions. Significance levels: *: $p < .05$, **: $p < .01$, ***: $p < .001$, .: $p < .1$.

Table 5.3: LMMs predicting articulation rate in Session 2

| Covariate | Estimate | SE | $t$ | $p$ (Corrected) |
|---|---|---|---|---|
| Task 2 | -0.517 | 0.044 | -11.637 | *** |
| Task 3 | -0.497 | 0.044 | -11.247 | *** |
| Task 4 | -0.109 | 0.044 | -2.464 | . |
| Task 5 | 0.097 | 0.045 | 2.164 | |
| Perceived Fast Speech | 0.130 | 0.045 | 2.896 | ** |
| English Comfort | 0.006 | 0.003 | 2.021 | * |

.: $p < .1$; *: $p < .05$; **: $p < .01$; ***: $p < .001$. Reported p (Corrected) values are Bonferroni-adjusted.

**Session 2:**

For the second session as seen in Table 5.3, articulation rates significantly varied across tasks, with participants speaking more slowly in Task 2 ($SE$ = 0.044, $t$ = -11.637, $p$ < .001) and Task 3 ($SE$ = 0.044, $t$ = -11.247, $p$ < .001) compared to Task 1. Task 4 led to a decrease in articulation rate, but it did not reach significance (SE = 0.044, t = -2.464, p = .058) based on the new threshold of p < 0.01 and Bonferroni corrections to the p-value.. Participants who self-reported that they are judged by others as fast speakers demonstrated significantly higher articulation rates ($SE$ = 0.045, $t$ = 2.896, $p$ = .003). People who reported more comfort in speaking English had a higher articulation rate with approaching significance ($SE$ = 0.03, $t$ = 2.021, $p$ = .005).

### 5.3.2 Pause Rate Models

We can see the effect on Pause Rates of participants across tasks in both sessions in Figure 5.14.

**Session 1:**

For the first session, as shown in Table 5.4, participants spoke with a significantly lower pause rate in Task 2 ($SE$ = 0.019, $t$ = -8.924, $p$ < .001), while pause rate did not significantly differ in Tasks 3 and 4. People in the Robot condition had a negative effect on pause rate with approaching significance ($SE$ = -0.056, $t$ = -2.076, $p$ < .042). Task 4 had

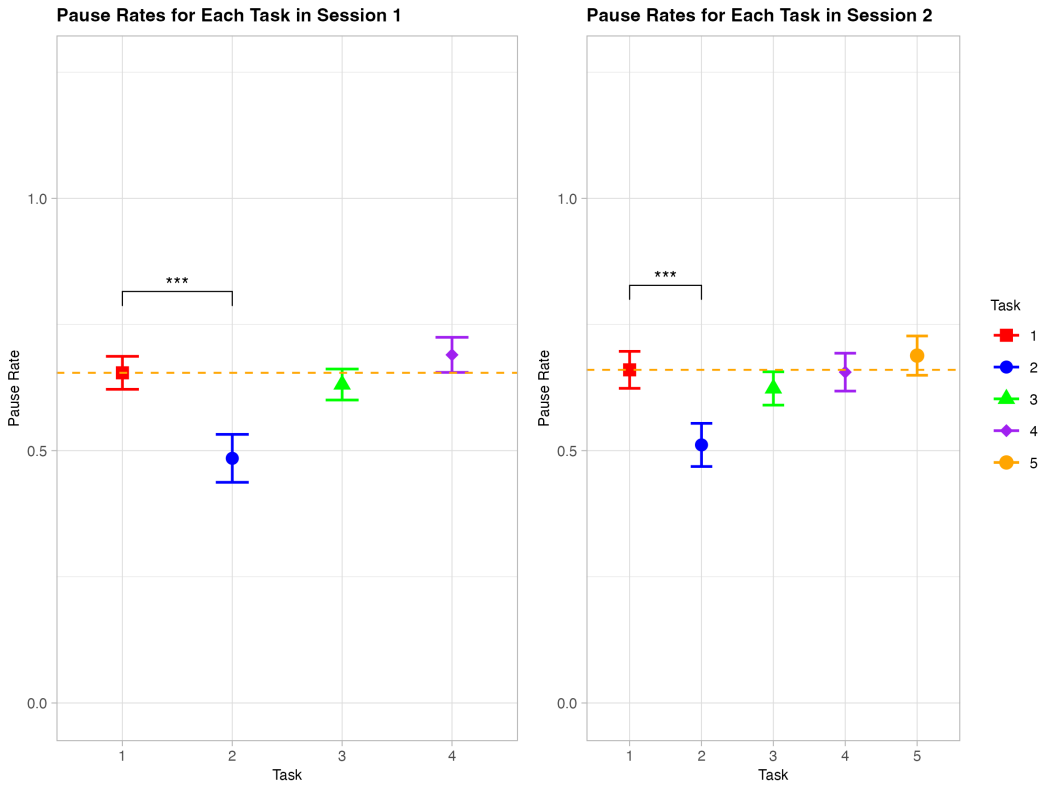



Figure 5.14: Participants' pause rates across tasks for both sessions. Significance levels: *: $p < .05$, **: $p < .01$, ***: $p < .001$, .: $p < .1$.

a positive effect on pause rate but did not reach significance ($SE = 0.019$, $t = 2.189$, $p < .09$) after applying Bonferroni correction and the stricter significance threshold.

**Session 2:**

For the second session, as seen in Table 5.5, participants exhibited significantly lower pause rates in Task 2 compared to Task 1 ($SE = 0.019$, $t =$ -7.636, $p < .001$). However, no significant differences were observed in pause rates for Tasks 3, 4 and 5. Pre-session Valence ($SE = 0.023$, $t =$ -2.283, $p < .026$) had a negative effect and pre-session Control ($SE = 0.020$, $t = 2.259$, $p < .028$) had a positive effect on the pause rate with approaching significance. Overall likability reported by the participants towards the Voice System had a negative effect with approaching significance ($SE = 0.001$, $t =$ -2.515, $p < .015$).

Table 5.4: LMMs predicting pause rate in Session 1

| Covariate | Estimate | SE | $t$ | $p$ (Corrected) |
|---|---|---|---|---|
| Condition (Robot) | -0.056 | 0.027 | -2.076 | * |
| Task 2 | -0.169 | 0.019 | -8.924 | *** |
| Task 3 | -0.021 | 0.019 | -1.090 | |
| Task 4 | 0.041 | 0.019 | 2.189 | . |
| Gender (Male) | 0.030 | 0.028 | 1.081 | |
| Agreeableness | 0.024 | 0.016 | 1.508 | |
| Emotional Stability | 0.013 | 0.012 | 1.084 | |

.: $p < .1$; *: $p < .05$; **: $p < .01$; ***: $p < .001$. Reported p (Corrected) values are Bonferroni-adjusted.

Table 5.5: LMMs predicting pause rate in Session 2

| Covariate | Estimate | SE | $t$ | $p$ (Corrected) |
|---|---|---|---|---|
| Task 2 | -0.147 | 0.019 | -7.636 | *** |
| Task 3 | -0.035 | 0.019 | -1.832 | |
| Task 4 | -0.002 | 0.019 | -0.098 | |
| Task 5 | 0.031 | 0.019 | 1.574 | |
| Valence (Pre-Session) | -0.052 | 0.023 | -2.283 | * |
| Control (Pre-Session) | 0.046 | 0.020 | 2.259 | * |
| Overall Likability | -0.002 | 0.001 | -2.515 | * |

.: $p < .1$; *: $p < .05$; **: $p < .01$; ***: $p < .001$. Reported p (Corrected) values are Bonferroni-adjusted.

### 5.3.3 Pause Ratio Models

We can see the effect on Pause Ratios of participants across tasks in both sessions in Figure 5.15.

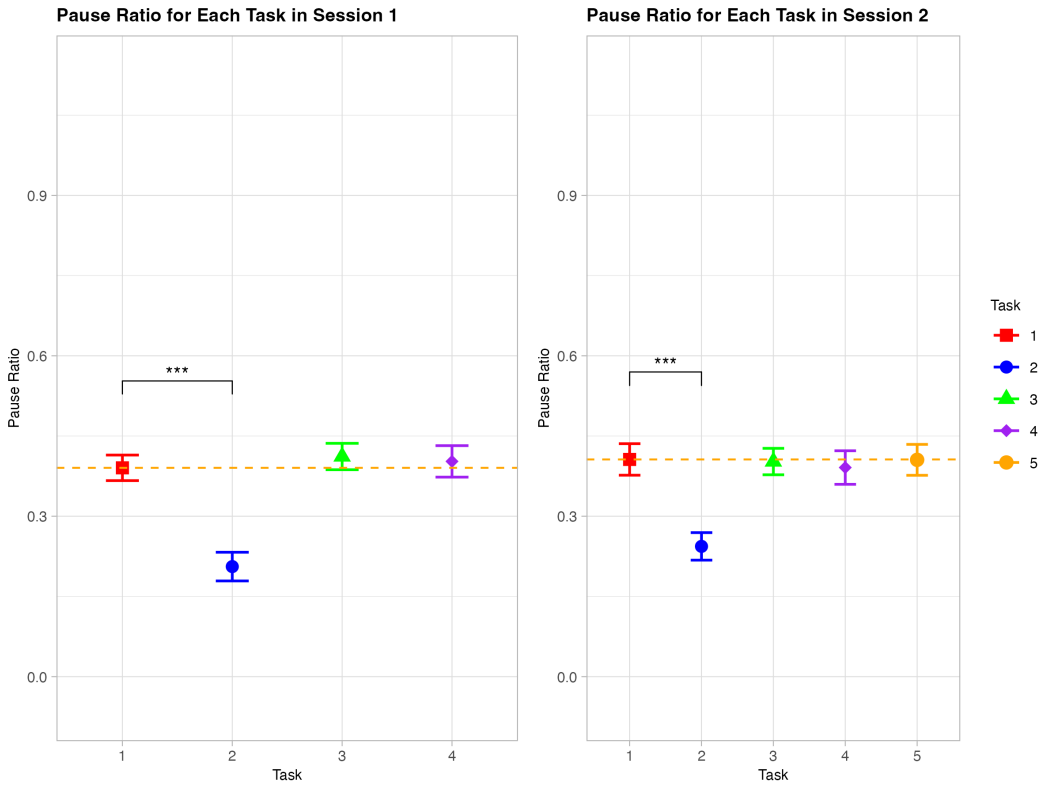



Figure 5.15: Participants' pause ratios across tasks for both sessions. Significance levels: *: $p < .05$, **: $p < .01$, ***: $p < .001$, .: $p < .1$.

**Session 1:**

As shown in Table 5.6, participants had significantly lower pause ratios in Task 2 compared to Task 1 ($SE$ = 0.013, $t$ = -14.216, $p < .001$). In contrast, neither Task 3 nor Task 4 showed significant differences in pause ratio compared to Task 1. Participants in the Robot condition had a lower pause ratio in comparison to the participants in the Laptop condition ($SE$ = 0.021, $t$ = -1.743, $p < .09$), but without a significant effect based on the threshold of p < 0.01. Male participants had higher pause ratios in comparison to the baseline female condition, without reaching significance ($SE$ = 0.021, $t$ = 1.851, $p < .07$).

Table 5.6: LMMs predicting pause ratio in Session 1

| Covariate | Estimate | SE | $t$ | $p$ (Corrected) |
|---|---|---|---|---|
| Condition (Robot) | -0.036 | 0.021 | -1.743 | . |
| Task 2 | -0.184 | 0.013 | -14.216 | *** |
| Task 3 | 0.020 | 0.013 | 1.507 | |
| Task 4 | 0.015 | 0.013 | 1.191 | |
| Gender (Male) | 0.038 | 0.021 | 1.851 | . |

.: $p < .1$; *: $p < .05$; **: $p < .01$; ***: $p < .001$. Reported p (Corrected) values are Bonferroni-adjusted.

**Session 2:**

As shown in Table 5.7, participants had significantly lower pause ratios in Task 2 compared to Task 1 ($SE$ = 0.014, $t$ = -12.143, $p < .001$). At the same time, Tasks 3, 4, and 5 did not show significant differences in pause ratios compared to Task 1.

Affective states reported before the session, as measured by the SAM questionnaire, also influenced pause behaviour. Higher pre-session Valence was associated with significantly lower pause ratios ($SE$ = 0.016, $t$ = -4.277, $p < .001$), while higher perceived Control was linked to significantly higher pause ratios as ($SE$ = 0.014, $t$ = 3.036, $p$ = .03). Participants who are perceived as slow speakers had higher pause ratio with approaching significance ($SE$ = 0.014, $t$ = 2.565, $p$ = .013).

### 5.3.4 Effect of Session on Speech Metrics

To understand the effect of different session on the speech metrics I further analyzed how articulation rate, pause rate, and pause ratio changed across sessions for different tasks in Table 5.8. I dropped the Task 5 data for all participants in this analysis as it was only used in Session 2. The baseline conditions were Session 1 and Task 1.

**Articulation Rate**

Overall, articulation rate was significantly higher in Session 2 compared to Session 1 ($SE$ = 0.053, $t$ = 4.189, $p < .001$). However, the effect of session differed across tasks. Compared to Task 1 for Session 1, articulation rate showed a greater decline in Session 2 for Task 2

Table 5.7: LMMs predicting pause ratio in Session 2

| Covariate | Estimate | SE | $t$ | $p$ (Corrected) |
|---|---|---|---|---|
| Condition (Robot) | -0.039 | 0.019 | -2.033 | |
| Task 2 | -0.165 | 0.014 | -12.143 | *** |
| Task 3 | -0.005 | 0.013 | -0.372 | |
| Task 4 | -0.016 | 0.013 | -1.213 | |
| Task 5 | 0.002 | 0.014 | 0.126 | |
| Gender (Male) | 0.013 | 0.019 | 0.675 | |
| Valence (Pre-Session) | -0.068 | 0.016 | -4.277 | *** |
| Control (Pre-Session) | 0.044 | 0.014 | 3.036 | ** |
| Perceived Slow Speech | 0.036 | 0.014 | 2.565 | * |

.: $p < .1$; *: $p < .05$; **: $p < .01$; ***: $p < .001$. Reported p (Corrected) values are Bonferroni-adjusted.

($SE = 0.076$, $t$ = -4.404, $p < .001$), Task 3 ($SE = 0.076$, $t$ = -4.193, $p < .001$), and Task 4 ($SE = 0.075$, $t$ = -3.539, $p < .001$).

### Pause Rate

Pause rate was significantly lower for Task 2 compared to Task 1 for Session 1 ($SE = 0.019$, $t$ = -8.691, $p < .001$). A main effect for Session 2 was not found, nor were there significant session-by-task interactions.

### Pause Ratio

Pause ratio was significantly lower for Task 2 compared to Task 1 for Session 1 ($SE = 0.015$, $t$ = -12.689, $p < .001$), but no significant session-by-task interactions were found. This indicates that while Task 2 consistently had lower pause ratios, the difference between sessions was not statistically significant.

Table 5.8: LMMs predicting articulation rate, pause rate, and pause ratio across all sessions and tasks

| Covariate | Estimate | SE | $t$ | $p$ (Corrected) |
|---|---|---|---|---|
| **Articulation Rate** | | | | |
| Session Number (Session 2) | 0.223 | 0.053 | 4.189 | *** |
| Task 2 | -0.183 | 0.054 | -3.419 | *** |
| Task 3 | -0.180 | 0.054 | -3.365 | *** |
| Task 4 | 0.158 | 0.053 | 2.962 | ** |
| Session 2 × Task 2 | -0.334 | 0.076 | -4.404 | *** |
| Session 2 × Task 3 | -0.317 | 0.076 | -4.193 | *** |
| Session 2 × Task 4 | -0.267 | 0.075 | -3.539 | *** |
| **Pause Rate** | | | | |
| Session Number (Session 2) | 0.009 | 0.019 | 0.443 | |
| Task 2 | -0.169 | 0.019 | -8.691 | *** |
| Task 3 | -0.020 | 0.019 | -1.008 | |
| Task 4 | 0.041 | 0.019 | 2.129 | |
| Session 2 × Task 2 | 0.021 | 0.027 | 0.771 | |
| Session 2 × Task 3 | -0.016 | 0.027 | -0.567 | |
| Session 2 × Task 4 | -0.043 | 0.027 | -1.574 | |
| **Pause Ratio** | | | | |
| Session Number (Session 2) | 0.018 | 0.014 | 1.245 | |
| Task 2 | -0.185 | 0.015 | -12.689 | *** |
| Task 3 | 0.020 | 0.015 | 1.390 | |
| Task 4 | 0.015 | 0.014 | 1.060 | |
| Session 2 × Task 2 | 0.020 | 0.021 | 0.978 | |
| Session 2 × Task 3 | -0.025 | 0.021 | -1.230 | |
| Session 2 × Task 4 | -0.032 | 0.020 | -1.548 | |

.: $p < .1$; *: $p < .05$; **: $p < .01$; ***: $p < .001$. Reported p (Corrected) values are Bonferroni-adjusted.

### 5.3.5 Self-Assessment Manikins (SAM)

Participants' emotional responses before and after the session were captured using SAM. The CLMM model for post-session Arousal is reported in Table 5.9. CLMM is used as the response variable here, the SAM ratings are ordinal in nature.

**Valence:**

The participants' overall likability of the session enhanced their post-session Valence ($SE$ = 0.021, $z$ = 4.631, $p < .001$) with a significant effect. A strong quadratic effect of pre-session Valence was observed, where higher initial Valence significantly increased post-session Valence ($SE$ = 1.003, $z$ = 5.514, $p < .001$). When we look at the cubic trend, it revealed that this relationship was not strictly linear, with pre-session Valence negatively influencing ratings ($SE$ = 0.486, $z$ = -2.559, $p$ = .010) with approaching significance. Participants in the second session reported significantly lower post-session Valence than those in the first session ($SE$ = 0.241, $z$ = -4.507, $p < .001$).

**Arousal:**

The data shows that participants' pre-session Arousal significantly predicted post-session Arousal ratings, where a strong linear effect was observed ($SE = 1.363$, $z = 4.870$, $p < .001$). Perceived intelligence of the Voice System was also found to positively influence Arousal, with higher perceptions of intelligence leading to significantly greater post-session Arousal ($SE$ = 0.024, $z$ = 2.047, $p$ = .041) with approaching significance. Also, the participants in the second session reported significantly lower post-session Arousal compared to the first session ($SE$ = 0.224, $z$ = -3.479, $p < .001$). This indicates that different sessions or repeated exposure had an impact on Arousal responses.

**Control:**

When asked about the perceived sense of Control over the current situation to the participants, their pre-session Control strongly predicted their post-session Control, with a significant linear effect ($SE$ = 1.123, $z$ = 10.031, $p < .001$) (see Table 5.9). A significant quadratic effect (($SE$ = 0.951, $z$ = -3.110, $p$ = .002) was also seen, suggesting a non-linear relationship between pre- and post-session Control. On top of that a cubic effect further suggests complexity in how pre-session control shaped post-session responses ($SE$ = 0.531,

$z = 4.531$, $p < .001$). Even for the ratings regarding participants' sense of Control in the current situation, the participants in the second session reported significantly lower post-session Control compared to the first session ($SE = 0.254$, $z =$ -4.323, $p < .001$). Session 2 appears to have a negative effect on all post-session SAM emotion ratings compared to Session 1.

### 5.3.6 Experience with the Voice System

Participants rated their experience with the Voice System based on likability, their level of involvement in interactions with the Voice System, the extent to which they felt they were interacting with an intelligent being, and how much attention they paid to the Voice System. The results are summarized in plots against Condition in Figure 5.16 and in Table 5.10.

#### Overall Likability:

Overall likability was significantly associated with involvement with the Voice System and perceived intelligence of the Voice System. Participants in the robot condition rated the voice system as more likable compared to those in the baseline laptop condition ($SE = 3.254$, $t = 2.532$, $p = .014$) with approaching significance. Additionally, participants who were more involved in the interactions reported higher likability ratings ($SE = 0.092$, $t = 3.223$, $p = .002$). Similarly, perceptions of higher Voice System's intelligence were also positively associated with overall likability ($SE = 0.089$, $t = 2.740$, $p = .008$).

#### Involvement:

Participants who rated the Voice System as more likable were also more involved in the interactions ($SE = 0.135$, $t = 2.432$, $p = .018$) with approaching significance. Perceptions of Voice System's intelligence had a similar positive effect ($SE = 0.100$, $t = 3.053$, $p = .003$), with people who perceived the Voice System as more intelligent were more involved in the interactions. The participants who paid more attention to the voice system reported higher robot involvement ($SE = 0.085$, $t = 4.881$, $p < .001$).

#### Intelligence:

Participants who rated the Voice System as more likable also perceived it as more intelligent ($SE = 0.149$, $t = 3.002$, $p = .004$). Similarly, those who were more involved in the

Table 5.9: CLMM predicting post-session SAM

| Covariate | Estimate | SE | $z$ | $p$ (Corrected) |
|---|---|---|---|---|
| **Valence** | | | | |
| Overall Likability | 0.097 | 0.021 | 4.631 | *** |
| Valence Pre-Session (Linear) | -1.062 | 1.258 | -0.844 | |
| Valence Pre-Session (Quadratic) | 5.533 | 1.003 | 5.514 | *** |
| Valence Pre-Session (Cubic) | -1.244 | 0.486 | -2.559 | * |
| Session Number (Session 2) | -1.086 | 0.241 | -4.507 | *** |
| **Arousal** | | | | |
| Arousal Pre-Session (Linear) | 6.641 | 1.363 | 4.870 | *** |
| Arousal Pre-Session (Quadratic) | 1.226 | 1.106 | 1.108 | |
| Arousal Pre-Session (Cubic) | -0.226 | 0.667 | -0.338 | |
| Arousal Pre-Session (4th Order) | -0.126 | 0.332 | -0.379 | |
| Voice System Intelligence | 0.050 | 0.024 | 2.047 | * |
| Voice System Attention | -0.033 | 0.024 | -1.349 | |
| Voice System Involvement | 0.055 | 0.031 | 1.765 | . |
| Session Number (Session 2) | -0.778 | 0.224 | -3.479 | *** |
| **Control** | | | | |
| Control Pre-Session (Linear) | 11.266 | 1.123 | 10.031 | *** |
| Control Pre-Session (Quadratic) | -2.957 | 0.951 | -3.110 | ** |
| Control Pre-Session (Cubic) | 2.406 | 0.531 | 4.531 | *** |
| Control Pre-Session (4th Order) | -0.531 | 0.322 | -1.647 | |
| Session Number (Session 2) | -1.098 | 0.254 | -4.323 | *** |

.: $p < .1$; *: $p < .05$; **: $p < .01$; ***: $p < .001$. Reported p (Corrected) values are Bonferroni-adjusted.

interactions also reported significantly higher intelligence ratings of the Voice System ($SE$ = 0.109, $t$ = 4.356, $p < .001$). A demographic effect was observed, where participants who identified as native English speakers rated the robot as significantly less intelligent than non-native speakers ($SE$ = 4.237, $t$ = -3.425, $p$ = .001). This suggests that non-

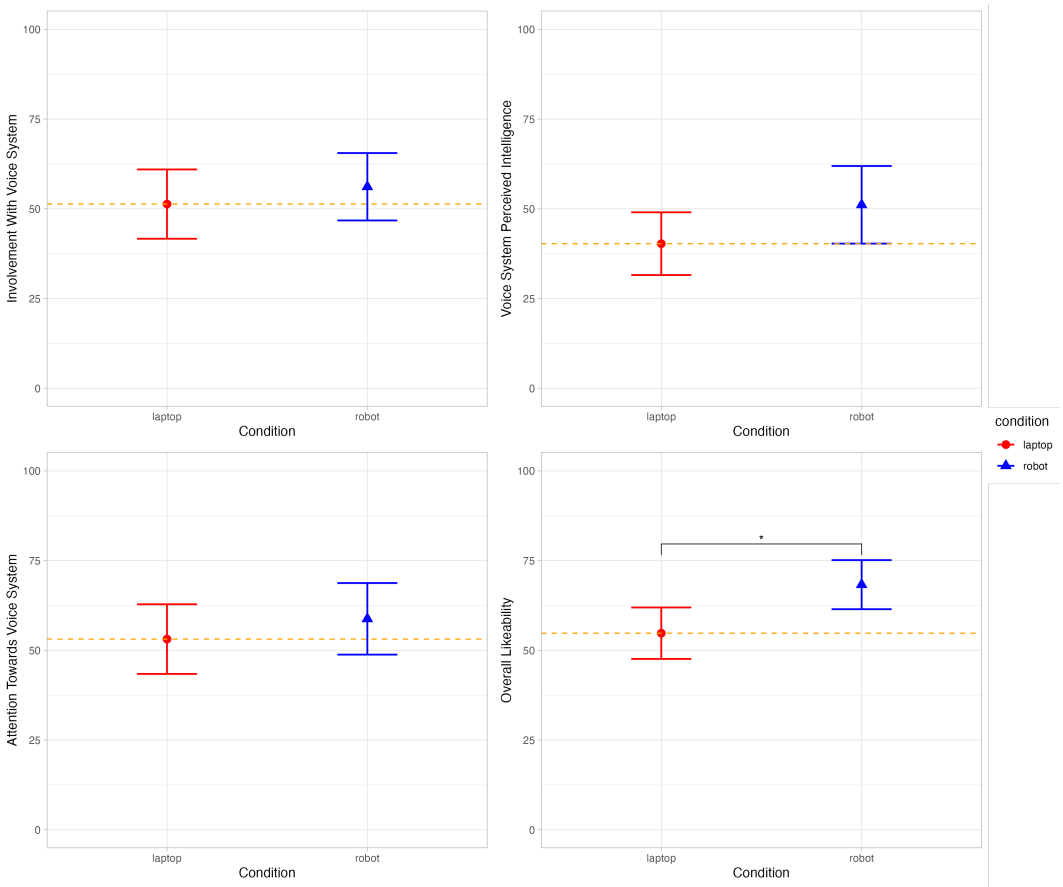



Figure 5.16: Participants experience with the Voice System in terms of Overall Likability, Involvement, Intelligence, and Attention. Significance levels: *: $p < .05$

native speakers may have had different expectations or experiences with the system that influenced their ratings.

**Attention:**

Overall likability was the only predictor of perceived attention paid to the Voice System that achieved significance, where participants who rated the system as more likable also paid significant attention to ($SE = 0.160$, $t = 4.380$, $p < .001$).

Table 5.10: LMs predicting overall likability, involvement, intelligence, and attention

| Covariate | Estimate | SE | $t$ | $p$ |
|---|---|---|---|---|
| **Overall Likability** | | | | |
| Condition (Robot) | 8.239 | 3.254 | 2.532 | * |
| Voice System Involvement | 0.297 | 0.092 | 3.223 | ** |
| Voice System Intelligence | 0.244 | 0.089 | 2.740 | ** |
| **Involvement** | | | | |
| Overall Likability | 0.329 | 0.135 | 2.432 | * |
| Voice System Intelligence | 0.304 | 0.100 | 3.053 | ** |
| Voice System Attention | 0.415 | 0.085 | 4.881 | *** |
| **Intelligence** | | | | |
| Overall Likability | 0.448 | 0.149 | 3.002 | ** |
| Voice System Involvement | 0.476 | 0.109 | 4.356 | *** |
| English Native (Yes) | -14.510 | 4.237 | -3.425 | ** |
| **Attention** | | | | |
| Overall Likability | 0.702 | 0.160 | 4.380 | *** |

.: $p < .1$; *: $p < .05$; **: $p < .01$; ***: $p < .001$

### 5.3.7 Thematic Analysis of Interview Questions

To understand how participants perceived the Voice System's speech, I explored their preferences regarding its speaking style. I also examined the perceived usefulness of slow and relaxed speech and whether they believed it could impact how they speak. To gather these insights, I conducted a short open-ended interview after the study. The audio responses from the recordings were transcribed using OpenAI's Whisper[10] [97], an open-source tool that was run locally on my laptop. I manually verified the transcriptions before analyzing them. Using a six-step Thematic Analysis [23]. I analyzed responses to the three questions for each participant. I conducted the Thematic Analysis of all the data, and another researcher at SIRRL independently reviewed 25% of the data. Later, I verified interrater reliability for the three analyses by using Krippendorff's alpha, which is a widely recom-

[10] https://github.com/openai/whisper

Table 5.11: Themes for the Voice System speech speed Preferences

| **Theme** | **Quote** |
|---|---|
| Faster Speech Preferred [A1T1] | *"...too slow for me personally because I feel like I'm fluent in English"* |
| Slow Speech Reduces Engagement [A1T2] | *"...more difficult to focus when the robot was speaking too slowly"* |
| Balance for Non-Native Speakers [A1T3] | *"...the pace was okay because for some listeners for whom English is not their native language, I feel that it's good because they might have some accent issues... "* |
| Add Expressiveness to Avoid Monotony [A1T4] | *"...increasing the speed a little bit or adding more emotion to it would help because it felt very monotone at some points"* |
| Should Not Change [A1T5] | *"...the voice system is fine as it is...So it was slow, so I can understand and follow the directions of the study"* |

mended reliability measure [73]. The alpha values were 0.968, 0.977 and 0.948 for the three questions, indicating a high level of agreement. I have assigned codes in the format AXTY, where A signifies 'answer', X represents the question number (1, 2, or 3), T stands for 'theme', and Y is the serial number assigned to the theme. In the discussion below, I have also included the percentage of responses in each condition, using L for the Laptop condition and R for the Robot condition.

### Preferences towards Voice System Speech

In the first question, I asked the participants if the Voice System should change the way it speaks, and if yes, then how and why. Based on their responses, I found the following themes:

- **Faster Speech is Preferred [A1T1] [L=93.55%, R= 78.12%]**: The majority of participants responded that they found the Voice System to be very slow. They preferred the natural human speed of talking. Some complained that the slower speech resulted in unnatural pauses and gaps that disrupted the flow.

- **Slow Speech Reduces Engagement [A1T2] [L=38.71%, R= 46.88%]**: Many participants found it difficult to maintain focus when the Voice System spoke too slowly. They reported getting bored and eventually getting distracted. The slow pace made them "zone out" or made them feel that the conversation was "dragging on".

- **Balance for Non-Native Speakers [A1T3] [L=3.23%, R= 12.50%]**: The speech should be slow enough for it to be understandable for non-native speakers. Even with slow speech, it should sound clear and natural without broken pronunciations. Too fast a speech could make non-native speakers struggle to process and understand the information well.

- **Add Expressiveness and Avoid Monotony [A1T4] [L=9.68%, R= 3.12%]**: Some participants suggested that we can make the speech more expressive to reduce a monotonous feeling. The monotonous tone made it harder for them to stay focused and also caused boredom.

- **Should Not Change [A1T5] [L=3.23%, R= 12.50%]**: Some participants felt that the Voice System should not change because the Voice System's slow pace made it easier to understand instructions for the different tasks.

**Perceived Usefulness of Slow Speech**

- **Slow Speech Helps Certain Listeners [A2T1] [L=45.16%, R= 50%]**: It was suggested that slow speech can be useful for non-native speakers by allowing time to process and understand language. Some people translate words to their native language in their minds. It might also be useful to increase comprehension for young children, older adults, and individuals with hearing impairments or cognitive difficulties.

- **Useful for Relaxation and Special Cases [A2T2] [L=25.81%, R= 18.75%]**: Some participants mentioned that slow speech can be helpful for relaxation when they are tired. Others pointed out that it could be beneficial in special cases, such as therapy or de-escalation situations, where a slower pace helps people feel more at ease.

- **Can Enhance Friendliness [A2T3] [L=12.90%, R= 6.25%]**: Some participants mentioned that slow speech can make a voice seem more friendly and approachable.

Table 5.12: Themes for perceived usefulness of slow speech

| Theme | Quote |
|---|---|
| Slow Speech Helps Certain Listeners [A2T1] | "...for people that might not speak English or that have some sort of speech or hearing disability" |
| Useful for Relaxation [A2T2] | "...useful in situations where a person is trying to relax" |
| Enhances Friendliness [A2T3] | "It'll seem more friendly, and it's good for customer service" |
| Excessive Slowness Reduces Engagement [A2T4] | "...sometimes it's too boring and the person loses interest" |
| Useful [A2T5] | "I do think it can be useful" |
| Not Useful [A2T6] | "...but this is slow, and I don't think this will be useful" |
| Helps Focus and Clarity [A2T7] | "...slow and relaxed speed is useful as it helps you to focus on the main information" |

Others noted that a slower pace and a gentle tone can create a sense of kindness and trust.

- **Excessive Slowness Reduces Engagement [A2T4] [L=6.45%, R= 9.38%]**: Some participants found excessive slowness frustrating or boring which made them them lose interest. Others reported that they felt disengaged, with one mentioning it caused yawning and another noting that the Voice System broke up words unnaturally. Participants indicated that a balanced speech rate was ideal.

- **Useful [A2T5] [L=90.32%, R= 90.62%]**: Most participants mentioned that it can be useful in general or some specific reasons covered in other themes.

- **Not useful [A2T6] [L=9.68%, R= 3.12%]**: A few participants indicated that the way in which slow and relaxed speech was implemented in this study might not be useful. They pointed out that when speech is too slow, it can break up words unnaturally and that reduces overall comprehension.

- **Helps Focus and Clear Understanding [A2T7] [L=9.68%, R= 21.88%]**: Some participants mentioned that slow speech helps them stay focused and prevents

Table 5.13: Themes for impact of the Voice System on user speech

| Theme | Quote |
|---|---|
| Mirror Speech Patterns [A3T1] | "I feel like while I was responding to the questions, I was talking a little bit slower than I would normally do." |
| Some Individuals Are Unaffected [A3T2] | "...if the robot speaks slow, then I will speak at my normal speed. Even if the robot speaks way fast, I'll still try to speak at my normal speed" |
| Long-Term Exposure Shapes Speech [A3T3] | "If I interacted with the voice system a lot, then maybe it would have more impact" |
| Some Users Counteract Slow Speech [A3T4] | "I was sort of really annoyed with the voice, I think I tended to speak faster. Like maybe when I was responding, I was like, Okay, that took so long, let me just get my response out of the way", |
| High Impact (70-100) [A3T5] | "I'd probably give it like a 90% impact" |
| Medium Impact (35-70) [A3T6] | "I think maybe like 60 or 70" |
| Low Impact (0-35) [A3T7] | "I think 10. I don't think it would impact my speaking as much" |
| Affects Content [A3T8] | "...when I was looking at the pictures and stuff, the robot was putting a lot of detail and stuff. So then it was making me think, oh maybe I should put more detail" |

them from losing track of the conversation. One of them suggested that it can be useful for making important announcements in railway stations or other crowded places.

### Impact of Voice System on User Speech

- **Speech Mirroring is Common [A3T1] [L=41.94%, R= 46.88%]**: Many participants indicated that they tend to match the speech patterns of the Voice System

in terms of speed and tone. They mentioned that it is natural for humans to match the communication style of their conversation partners.

- **Unaffected [A3T2] [L=25.81%, R= 31.25%]**: Several participants stated that the robot's speech did not impact how they spoke, as they naturally maintained their usual speaking pace.
- **Long-Term Exposure Can Shape Speaking Patterns [A3T3] [L=12.90%, R= 9.38%]**: Some participants mentioned that interacting with the Voice System in this study did not change how they spoke, but frequent exposure to slow speech over time could influence their speaking patterns. Over time, they will start adjusting their speed or catching specific phrases of the Voice System.
- **Opposite Behaviour [A3T4] [L=3.23%, R= 6.25%]**: Some participants mentioned that instead of mirroring the slow speech, they reacted by speaking faster out of frustration or impatience. They said that they felt agitated by the slow pace and deliberately spoke more quickly.
- **High Impact [A3T5] [L=19.35%, R= 31.25%]**: This theme corresponds to participants who responded that, on a scale of 0 to 100, they felt the robot's slow speech would impact their speaking style by more than 70.
- **Medium Impact [A3T6] [L=32.26%, R= 21.88%]**: This theme represents responses where the participant's indicated the impact on their speech would be between 35 to 70%.
- **Low Impact [A3T7] [L=19.35%, R= 15.62%]**: Low impact refers to participants who indicated that their speaking style would not be significantly affected, rating the impact on their speaking style between 0 and 34.
- **Affects Speech Content [A3T8] [L=12.9%, R= 9.38%]**: Some participants mentioned that the robot's speech not only influenced their speaking speed but also affected the content of their responses. One person mentioned that when they were describing pictures they tried to include similar phrases as the robot did.

## 5.4 Discussion

In this section, I will present my insights from the various analysis results of this study. I will discuss how the results, together, help us obtain answers to the research questions in Section 5.1.

### 5.4.1 Effect on Participants' Speech Metrics

The results of the speech metrics we studied—articulation rates, pause rates, and pause ratio—help us understand how the participant's speech changed across different tasks and between the two sessions. The responses for Task 1 in a given session served as the baseline against which all other tasks were compared.

#### Articulation Rate

The sentence repetition task (Task 2) is an extension of the word shadowing task commonly studied in speech entrainment research. It not only has a significant impact but also has the strongest effect on participants' articulation rates. In Task 3, where participants had to describe a picture in three sentences, it was slightly more open-ended than the sentence repetition task. This task also significantly reduced participants' articulation rates. We saw an increase in articulation rate for Task 4. In Task 4, participants had to respond to open-ended questions for 90 seconds. Overall, these findings are consistent with the well-established concept of speech entrainment. Speakers tend to unconsciously adjust their speech patterns to align with their conversation partners [54, 94]. For Session 2, we can see that Tasks 2 and 3 led to a greater relative drop in articulation rates when compared to Session 1. This suggests that with longer exposures, we can observe more speech entrainment. The plots in Figure 5.13 show an interesting trend. The relative drop in articulation rates for Tasks 2, 3, and 4 compared to Task 1 has increased. However, the values for these tasks in relation to their Session 1 values appear to be only slightly lower, and this is confirmed in the Table 5.8 that only Task 2 in Session 2 for articulation rate had a significant decrease in comparison to Task 1 in Session 1. There was no significant effect of the condition, Robot vs. Laptop. These observations help address the first three research questions (**RQ1**, **RQ2** and **RQ3**) regarding the effect of condition, task variations within a session, and articulation rate comparisons between sessions (see Research Questions in Section 5.1). These findings are also supported by the themes identified in the thematic analysis of the response to the three interview questions. More than half of the participants felt that the Voice System's speech will impact their speech rate either now [**A3T1**] or with long-term exposure [**A3T3**].

Regarding **RQ4**, which focused on the 'surprise' task, Task 5—where participants were asked to sell one of the items placed on a table to their right—this task was not mentioned in the information sheet. We observed that participants' articulation rates were higher to those recorded in Task 1 of that session but it did not reach significance after applying Bonferroni correction. This needs further investigation of scenarios when participants

can lose their acquired slow speech when faced with an unpredictable event during the session. For example, we should examine what happens if we introduce another structured task such as the sentence repetition exercise immediately after the surprise task. These scenarios would more closely resemble real-world gameplay situations used in HRI studies. Furthermore, this finding could inform session design decisions for future speech modelling using Social Robots if the researchers want to sustain some acquired speech patterns.

We also noticed that participants' experience with the Voice System's speech rate and perceived intelligence impacted their articulation rates. Participants who perceived the voice system as less intelligent tended to speak more slowly, with a significant effect. Even factors related to the participants themselves had an impact. People perceived as fast speakers by their peers exhibited a significant increase in their articulation rates. This indicates that the impact on articulation rate is highly subjective, and solutions need to be customized to align with individual attributes.

### Pause Rate

A lower pause rate indicates that an individual took fewer pauses during their total utterance duration. Task 2 led to a significant decrease in pause rate in both sessions. So I would interpret it that participants in a sentence repetition task do not need to pause and think as much which lead to a fewer pauses. This needs to be explored through further studies.

Likability towards the Voice System, Pre-session Valence and Control also had an effect on pause rates in the second session, with approaching significance. The significance threshold was made stricter by shifting to $p < 0.01$ from $p < 0.05$. Participants who reported more positive feelings on the SAM questionnaire prior to the session tended to have lower pause rates. This suggests that emotional state may be related to pause rate during verbal interactions with social robots. Similarly, those who felt a stronger sense of control over their situation also paused more frequently. This suggests that if we want the participants to pause more, then we should have study designs that promote a sense of control of the participants. In addition, participants who found the Voice System more likable took less frequent pauses. Having a more likable robot can possibly cause more involvement (see Table 5.10) of the participant, which can then decrease the pause rate. When designing sessions with social robots to increase the pause rate, we shall consider the effects of likability and involvement of participants in the interaction with the social robot.

#### Pause Ratio

The pause ratio measures the amount of time spent by the participant in pauses during speech. Similar to the pause rates, Task 2 again showed a significant decrease in pause ratio, likely due to its nature as a sentence repetition task. I would reason that participants did not have to think before repeating a sentence, they may have spent less time pausing. This will need further explorations to reach to a conclusion. No other task had significant effects in both sessions.

The pause ratios are significantly affected by the pre-session Valence and Control in the second session. Individuals who reported higher positive emotions before the session took shorter pauses. Participants with a higher sense of control took more extended pauses. In terms of pause rate and pause ratio, we aim to increase these metrics to achieve a slower and more relaxed speech style. We want to design sessions that are engaging to an extent but not overly involving, as that may reduce participants' pausing behaviour. Having structured sessions that promote a sense of control could possibly encourage more relaxed pause behaviour.

These findings suggest that pause durations are shaped by both external conditions, such as demands of the tasks and interaction context, and individual characteristics, like pre-session Valence and Control. Further explorations and design considerations need to be studied to develop applications that achieve the desired end goal of speech modelling. The same approach or task will not help us decrease articulation rate and, at the same time, increase pause rates and pause ratios.

### 5.4.2 Effect on Participants' Emotional States

Through SAM, we aimed to measure emotional responses before and after each session. The results indicate that the session number affected post-session Valence, Arousal, and Control. The condition (robot vs. laptop) did not have a significant effect. After the second session, participants reported lower Valence, Arousal, and Control compared to the first session. This means after Session 1, or what happened during the gap between the sessions, or other aspects of Session 2 (the theme or Task 5) could have played a role. The pre-session responses showed non-linear, complex effects, indicating no direct one-to-one relationship with post-session ratings. The overall likability had a significant positive impact on Valence. To answer **RQ5**, Condition did not play a significant role in impacting the affective states of participants. The decrease in Arousal and Control after the second session suggests that participants felt more relaxed. This decrease could also be linked to

disengagement. Both relaxation ([**A2T2**]) and disengagement ([**A1T2**], [**A2T4**]) that also emerged as central themes in the post-study interview responses. Although the comments were not frequent enough to form a theme in the responses to the third question, a few participants mentioned feeling relaxed, while others reported feeling disengaged.

### 5.4.3 Perceptions of the Voice System

Our results from Table 5.10 indicate that various measurements of participants' perceptions, such as their level of involvement in interactions with the Voice System such as their perception of its intelligence, the amount of attention they paid to it during interactions, and its overall likability (measured using a subscale of the Godspeed Likability questionnaire), these were all interconnected to some extent. These four metrics positively and significantly influence and reinforce one another. It shall be noted that involvement in interactions, perceptions of the Voice System's intelligence, and attention toward the system were all consistently and significantly influenced by overall likability, which acted as a positive driving factor. We reported that the participants rated the Voice System as significantly more likable when deployed on a social robot rather than in the baseline laptop condition. This indicates that physical embodiment affected user perception, and this could be due to increased anthropomorphism or a stronger sense of social presence.

Participants' background as native English speakers impacted how intelligent they felt the Voice System to be. They rated it to be less intelligent in comparison to the ratings by non-native speakers. This could indicate that non-native speakers may have different expectations, and they find the system's slow speech more aligned with their linguistic expectations.

## 5.5 Limitations and Future Work

I took note of several limitations during the course of the study, and some were highlighted through discussions with participants during the interviews. The limitations are listed as follows:

- The participants reported the conversations in the study to be less engaging. The primary reason behind this might be that the interactions during the sessions were scripted and hardcoded that lacked any dynamic elements. This design choice was planned to simplify data collection and analysis, as the primary focus was on audio data collection.

- Some participants found the instructions for different tasks in the study repetitive, and they expressed frustration as a result. This redundancy was considered during the study design phase, and we decided to retain it. Since there was no back-and-forth conversation, it was essential for participants to fully understand their tasks. They did not have the opportunity to ask for clarification or request the Voice System to repeat the instructions.

- The slow speech, which was the central theme of this study, led to decreased engagement and boredom among participants [**A1T2**], [**A2T4**]. The monotonous speech of the Voice System also contributed to the lack of engagement [**A1T4**]. As we focused exclusively on verbal speech sample collection, the study excluded interactive tasks involving physical actions. Engagement was not one of the metrics we aimed to maximise in our study. Additionally, this study involved partial disclosure, we did not inform the participants beforehand that this study was about reducing their articulation rate, increasing their number of pauses, or taking extended pauses. The participants were debriefed about the true purpose of the study at the end of the study. When exploring the applications of social robots in future slow and relaxed speech training for caregivers of stuttering preschoolers, understanding why the robot speaks slowly could improve acceptance and tolerance of this speech pattern. It was not unexpected that the University students who participated in our study would find the slow speech frustrating. Providing more background on its purpose could also potentially enhance speech entrainment, which can be further explored in future studies. For this study, we aimed to investigate whether speech entrainment occurs even when participants do not expect or intend to slow their speech.

- Some participants mentioned that they felt the Voice System would have impacted their speaking style if they had spoken for longer durations. Having just two sessions could have resulted in less speech entrainment in the case of some participants [**A3T3**].

- In some participants, Slow speech caused the reverse effect, they felt irritated and intentionally spoke faster to finish the tasks sooner [**A3T4**].

In future studies, HRI researchers can explore study designs and implementations that are more engaging. By carefully using real-time sensing through camera or audio input, the social robot can determine when a participant has stopped speaking. If we use such an approach then the participants are not forced to complete their responses within a given time while also avoid the risk of interrupting them mid-speech. Including physical props such as toys or board games can make interaction tasks more interesting, leading to an

increase in engagement. In our study, we observed a trend for Task 5 towards an increase in articulation rate. This indicates that the surprise element can lead participants to drift away from their acquired slow speech during the session. Further exploration shall be made in the direction of having more predictive tasks (such as Task 2 and Task 3) after the surprise task to see if participants can reacquire slow speech. Having more than one such unscripted task would be closer to real-life scenarios and could improve participant engagement. Conducting a similar study with children who stutter and their caregivers, where the social robot facilitates interactive games in a slow and relaxed speech, is the direction we intend to move forward in. Having a larger number of sessions in future studies could also improve speech entrainment in human speech. There is also scope to go beyond the challenges faced by stuttering preschoolers and consider voice system developments for other age groups while exploring varied fluency shaping techniques such as syllable-timed speech or the stretched syllable techniques [60, 114]. Combining speech entrainment with play-based sessions has the potential for subconscious modifications in speaking patterns.

A few novel application areas emerge from the themes of the interview responses. Slow speech can help people feel relaxed and can be used for meditation practices or with individuals with speech or hearing impairments [**A2T1**], [**A2T2**]. Slow speech can also be useful for non-native speakers [**A1T3**]. A slow speech, with a balance of factors that can improve engagement with a social robot, can be used in areas where people need to focus particularly on the content of speech or when making announcements in crowded areas [**A2T7**].

# Chapter 6

# Conclusion

## 6.1 Summary of Findings

In this section, I will sequentially list the research questions and objectives explored in this thesis. I will also present the answers to these questions and the outcomes of our studies.

### 6.1.1 Findings of Study 1

In this study, I examined the preferences for cultural greetings based on locational context (restaurant's theme) for two communities (HY and ASA) that were from different cultural backgrounds. Participants who identified with the greetings "Huān Yíng" or "As-salāmu alaykum" are referred to as **HY** and **ASA**, respectively.

**RQ1: Will individuals from HY and ASA groups prefer a Western greeting or their own ethnic greeting in a restaurant whose theme resonates with their ethnic greeting?**

When participants experienced greetings in a restaurant matching their ethnic background, results deviated from the initial hypothesis that all participants would prefer their own ethnic greetings. The results are interpreted with consideration of the limitation that the cultural greetings were implemented using the Pepper Robot's English language package. Additionally, the welcome message following the greeting was delivered in English to ensure comprehension for participants in both HY and ASA groups. The results suggest that the HY participants favored popular Western greetings and this was a result of their time spent in Western countries, concerns over accent and only the greetings and not the

entire communication being in aligned with their culture. Based on [110] we formulated a hypothesis that participants will prefer their ethnic greetings in a restaurant setting whose theme resonates with their own ethnic greeting. ASA participants, however, aligned with the hypothesis, preferring their ethnic greetings. Our findings suggest that familiarity of Western greetings and accuracy implementation can play a role on cultural greetings preferences.

**RQ2: In a restaurant whose theme does not resonate with their ethnic greeting or Western greeting, will people in HY and ASA groups prefer a Western greeting, one that matches the restaurant's theme, or their own ethnic greeting?**

In a restaurant whose theme did not align with their own or Western greetings neither group (HY and ASA) showed a strong preference for any particular greeting. This finding highlights the complex factors influencing greeting preferences. Thematic analysis revealed that both groups showed a lower preference for their ethnic greetings in this scenario. Some participants suggested that they value the authenticity of the experience of visiting a restaurant of a different background theme. They were also open towards learning about a new culture over the comfort of familiar greetings in unfamiliar cultural contexts.

### 6.1.2 Findings of Study 2

In this study, our research objective was to **find the applications of social robots in speech therapy clinics** to help preschoolers who stutter. Our intention was to analyze our data and come up with **design considerations for these applications** intended for use by SLPs that were grounded in real-world observations.

The findings of this study highlight three key aspects of social robots in the stuttering therapy of preschoolers: **practical applications**, **interaction design considerations**, and **communication design considerations**. 1) In terms of **practical applications**, social robots can serve as a toolbox that offers a variety of games, provide a consistent speech model, demonstrate stuttering examples for training parents, and act as storytellers. The social robot will complement rather than replace the role of SLPs. 2) Regarding **interaction design considerations**, the results suggest that play-based interactions can be effective. The interactions should promoted engagement through entertainment, interactivity, turn-taking, and reward-based gameplay experiences while also maintaining a structured, low-cognitive-load environment. Therapy can be made engaging through the usage of tangible objects and that would also help reducing screentime. The social robot can take on various roles, such as a peer, a teacher, or other supportive roles, depending

on the context around which activities are design and the therapeutic needs. 3) **communication design considerations** emphasise the importance of non-generic, personalised speech content with more descriptive comments around the activity, short sentences, repeated phrases, and gestures to support speech development. The speech of the robot should sound natural and shall be slow-paced with an encouraging tone. Cultural elements could also be used to integrate familiar greetings and gestures. The robot should also demonstrate active listening by using non-verbal gestures. The child should not be interrupted as it is important that the child feels heard and supported.

### 6.1.3 Findings of Study 3

I designed this study to understand how the slow and relaxed speech of a Voice System can impact participants' speech metrics and emotional states. The research questions and the findings are listed below.

**RQ1: How do human speech patterns evolve between sessions when interacting with a social robot or a laptop speaking at a slow speech rate?**

We can see in Figures 5.13, 5.14 and 5.15 the line passing through the mean value of the metric (articulation rate or pause rate or pause ratio) the baseline Task 1. We can notice that for articulation rate, the difference for each tasks 2, 3 and 4 from the baseline value Task 1 for that session has decreased more in Session 2. Even the estimates of statistical results in Tables 5.2 and 5.3 indicate that. Task 2 significantly decreased the pause rates and pause ratios in both sessions, as seen in Figure 5.14 and 5.15. Although the difference in estimates is not high also in tables related to pause rate: 5.4 and 5.5 and pause ratios: 5.6 and 5.7. For tasks 3 and 4, the effect of pause rates and pause ratios was not significant for both sessions to compare the impact of sessions. Table 5.8 compares all speech metrics (articulation rates, pause rates and pause ratios) across all tasks and sessions and shows that only Task 2 decreases in both sessions for all speech metrics.

**RQ2: Is there a difference in the extent of human speech alignment when interacting with a social robot versus a laptop speaking at a slow speech rate?**

Whether participants were in the Robot condition or the Laptop condition, it did not have any significant effect on their articulation rates. Participants in the Robot condition had lower pause rates in Session 1 with approaching significance, and no significant effect in pause ratio in both sessions based on the stricter threshold of $p < 0.01$. This indicates that the number of pauses could decrease in the Robot condition.

**RQ3: How does interacting with a social robot or a laptop speaking at a slow speech rate influence human speech patterns across different types of**

**speech tasks (sentence repetition, picture description, and open-ended responses) within a session?**

The articulation rates decreased for sentence repetition and picture description tasks within both sessions. The articulation rate increased for the open-ended task in Session 1 but then decreased in Session 2. Pause rates and pause ratios increase for the sentence repetition task. For a detailed discussion, please refer to Section 5.4.1

**RQ4: Does the participant maintain a consistent speech pattern during the task of 'surprise task', i.e. "selling" an object to the social robot or laptop at the end of the second session, given the subtle pressure of the task?**

We noticed that the participants' articulation rate had an increasing trend but did not reach significance when they responded to the surprise task, see Figure 5.13 and Table 5.3. We did not find any significant effects of the surprise task on pause rates and pause ratios.

**RQ5: How do the interaction context (social robot vs. laptop) and slow speech rate influence participants' emotional responses (valence, arousal, and dominance)**

In our data, the condition did not play a role, but the second session had a negative effect on post-session arousal and control ratings as seen in Table 5.9.

## 6.2 Contribution to Knowledge

This research contributes to the domain of HRI where we promote inclusivity in communication. I investigated how social robots can enhance inclusive communication from two perspectives: effective cultural integration and applications in stuttering therapy. The first user study explored the effects of locational context on cultural greeting preferences. In my second and third studies, I first proposed potential applications and design considerations for social robots in stuttering therapy, then examined whether speech entrainment could be used to induce slow and relaxed speech in adults. Slow and relaxed speech of caregivers may benefit stuttering therapy outcomes for preschoolers.

- Results of my first study showed the intricate relationship between a person's identity, reflected in their ethnic greeting. Their unique traits, and the percentage of life lived in Western countries had an effect on how they like social robot greetings. Participants in the HY and ASA groups indicated varying degrees of influence on their greeting preferences due to the voice and language used by the robot throughout

the interaction. It is important to understand the effects of these factors deeply to develop technologies that are welcoming and culturally sensitive. Integrating user-controlled adaptation (i.e., customization [89]) into the HRI design could ensure robots meet diverse needs and expectations that could enhance their acceptance and effectiveness. Customization can enhance perceived human control and promote a sense of identity by allowing users to make personally relevant feature choices. This actively involves users in shaping their interactions with products or systems [75]. HRI researchers must be mindful of the complexities around ethnicity and culture while ensuring that they do not reinforce stereotypes. Such efforts can help achieving the goal of improving the balance of autonomy between humans and systems.

- The second study explores applications of social robots in critical and sensitive domains such as therapy, particularly those that involve young children that presents unique challenges. Involving stakeholders in the design process is crucial if we want future implementations to find its use in the field. Given the novel nature of the applied research explorations needed for this study, we adopted the concept of C-GTM to root our findings on the data that we encountered. The subtle aspects identified through this study go beyond finding the applications of social robots in a clinical setting. They go deeper into the kind of interactions and the communication style that must be considered when designing the applications. The collective impact of the communication style and the way the social robot interacts with the clients and SLPs in the carefully designed game scenarios will guide us toward achieving meaningful and effective therapy outcomes.

- In my third study, I examined how the slow and relaxed speech of a social robot impacts various speech metrics of participants, such as articulation rate, pause rate, and pause ratio. Through my in-depth qualitative and quantitative analysis, I demonstrated that certain types of tasks are more effective in slowing down the articulation rate compared to others. We also observed a further decrease in articulation rates during the second session. Additionally, surprise elements in the session can possibly lead to an increase in articulation rate, but it will need further exploration. These observations can help HRI researchers guide the design of future studies and applications of social robots as models for slow and relaxed speech. This work could also inspire studies around speech entrainment and other fluency-shaping techniques.

The findings of this research work contribute to the HRI knowledge through improvements in cultural integration efforts and through conducting groundwork in applications of social robots in the stuttering therapy of preschoolers. It is important to have a broader

outlook towards inclusivity in communication, and through my work I address it from two directions.

## 6.3 Limitations and Future Work

The limitations of our work in terms of cultural integration, discussed in Chapter 3 *Study 1: Greeting Preferences in a Hospitality Context: A Cross-Cultural Study with a Social Robot*, were that it was an online study with participants mainly being university students from two communities. The interaction scenarios in that study incorporated cultural aspects limited only to greetings, which raised questions about the authenticity and quality of our implementation. There were also limitations related to the robot's pronunciation and accent, as the Pepper robot's English language package was used to implement both the cultural greetings and the welcome message. Additionally, there were fewer participants from the ASA group in comparison to the HY group. Future implementations should involve a more diverse set of populations in different contexts beyond a restaurant setting. We can get more insights from longer duration in-person studies with specific cultural accents and phonetics to create a more authentic interaction.

In our exploration of the applications of social robots in stuttering therapy for preschoolers in Chapter 4 *Study 2: Design Considerations for Applications of Social Robots in the Stuttering Clinic*, I was able to shadow only a limited number of stuttering preschoolers in therapy sessions due to challenges in obtaining parental and SLP consent within practical geographic proximity. The shadowing sessions I observed were conducted in English with preschoolers, the majority of whom were non-severe stutterers, which affected the generalizability of the results. The applications and design considerations suggested in this study need to be tested in practice to evaluate their effectiveness in helping SLPs and their clients improve therapy sessions. In future work, we plan to prototype games and applications based on the established design and conduct in-person studies involving preschoolers who stutter and SLPs integrating our solutions into their sessions.

In Chapter 5 *Study 3: Investigating Human Speech Rate Alignment in Response to Slow Robot Speech* we discussed our study on speech entrainment of the Voice System's slow speech. The study had certain limitations due to its scripted nature. Participants felt less engaged during interactions, and repetitive task instructions, along with monotonous communication, led to frustration and boredom for some. Future studies could use a larger sample size and could consider including factors that promote engagement and implementing real-time sensing to reduce the need for scripted sessions. Since we aim to use this approach for caregivers of children undergoing speech therapy, they would understand the

purpose of slow speech, which might help improve engagement. Conducting studies of longer duration with actual families where children are undergoing speech therapy can further validate our results. Additionally, we can use slow speech to benefit meditation practices, non-native speakers, and public announcements, where the focus is to make the content more accessible.

# References


[1] Choregraphe documentation. https://developer.softbankrobotics.com/nao6/choregraphe-documentation, 2024. Accessed: March 20, 2025.

[2] Softbank robotics - group: Global site. https://www.softbankrobotics.com/, 2024. Accessed: March 20, 2025.

[3] University of waterloo equity survey results. https://uwaterloo.ca/equity-diversity-inclusion-anti-racism/equity-survey/equity-survey-results, 2024. Accessed: March 20, 2025.

[4] Abeer Al-Nafjan, Noura Alhakbani, and Amal Alabdulkareem. Measuring engagement in robot-assisted therapy for Autistic children. *Behavioral Sciences*, 13(8):618, 2023.

[5] Amal Alabdulkareem, Noura Alhakbani, and Abeer Al-Nafjan. A systematic review of research on robot-assisted therapy for children with autism. *Sensors*, 22(3):944, 2022.

[6] Amazon. Amazon echo. https://www.amazon.com/echo, 2025. Accessed: March 17, 2025.

[7] Amazon. Amazon polly: Text-to-speech service. https://aws.amazon.com/polly/, 2025. Accessed: March 17, 2025.

[8] Sean Andrist, Micheline Ziadee, Halim Boukaram, Bilge Mutlu, and Majd Sakr. Effects of Culture on the Credibility of Robot Speech: A Comparison between English and Arabic. In *Proceedings of the tenth annual ACM/IEEE International Conference on Human-Robot Interaction*, pages 157–164, 2015.

[9] Apple. Apple siri. https://www.apple.com/siri, 2025. Accessed: March 17, 2025.

[10] Negin Azizi, Shruti Chandra, Mike Gray, Jennifer Fane, Melissa Sager, and Kerstin Dautenhahn. User evaluation of social robots as a tool in one-to-one instructional settings for students with learning disabilities. In *ICSR*, pages 146–159. Springer, 2022.

[11] Negin Azizi, Shruti Chandra, Mike Gray, Melissa Sager, Jennifer Fane, and Kerstin Dautenhahn. An Initial Investigation into the Use of Social Robots within an Existing Educational Program for Students with Learning Disabilities. In *IEEE RO-MAN*, pages 1490–1497, 2022.

[12] Alice Anne Gamble Bailey and Walter R Bailey. Managing the environment of the stutterer. *Journal of Childhool Communication Disorders*, 6(1):26–39, 1982.

[13] Christoph Bartneck, ME Foster, and M Giuliani. Godspeed Questionnaire Series. *Christoph Bartneck, Ph. D*, 2008.

[14] Christoph Bartneck, Tomohiro Suzuki, Takayuki Kanda, and Tatsuya Nomura. The influence of people's culture and prior experiences with Aibo on their attitude towards robots. *AI & Society*, 21:217–230, 2007.

[15] Douglas Bates, Martin Mächler, Ben Bolker, and Steve Walker. Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823*, 2014.

[16] Linda Liska Belgrave and Kapriski Seide. Grounded Theory Methodology: Principles and Practices. In Pranee Liamputtong, editor, *Handbook of Research Methods in Health Social Sciences*, pages 299–316. Springer Singapore, Singapore, 2019.

[17] Štefan Beňuš. Social aspects of entrainment in spoken interaction. *Cognitive Computation*, 6:802–813, 2014.

[18] Alina Betlej. Designing robots for elderly from the perspective of potential end-users: a sociological approach. *International Journal of Environmental Research and Public Health*, 19(6):3630, 2022.

[19] J Martin Bland and Douglas G Altman. Multiple significance tests: the bonferroni method. *Bmj*, 310(6973):170, 1995.

[20] Michael Blomgren. Behavioral treatments for children and adults who stutter: A review. *Psychology research and behavior management*, pages 9–19, 2013.

[21] Hamparsum Bozdogan. Model selection and Akaike's Information Criterion (AIC): The general theory and its analytical extensions. *Psychometrika*, 52(3):345–370, 1987.

[22] Margaret M Bradley and Peter J Lang. Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of behavior therapy and experimental psychiatry*, 25(1):49–59, 1994.

[23] Virginia Braun and Victoria Clarke. Using thematic analysis in psychology. *Qualitative research in psychology*, 3(2):77–101, 2006.

[24] Barbara Bruno, Aida Amirova, Anara Sandygulova, Birgit Lugrin, and Wafa Johal. Culture in Social Robots for Education. In *Cultural Robotics: Social Robots and Their Emergent Cultural Ecologies*, pages 127–145. Springer, 2023.

[25] Barbara Bruno, Nak Young Chong, Hiroko Kamide, Sanjeev Kanoria, Jaeryoung Lee, Yuto Lim, Amit Kumar Pandey, Chris Papadopoulos, Irena Papadopoulos, Federico Pecora, et al. Paving the way for culturally competent robots: A position paper. In *26th IEEE international symposium on robot and human interactive communication (RO-MAN)*, pages 553–560. IEEE, 2017.

[26] Laura Staum Casasanto, Kyle Jasmin, and Daniel Casasanto. Virtually accommodating: Speech rate accommodation to a virtual interlocutor. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, volume 32, 2010.

[27] José Carlos Castillo, Diego Alvarez-Fernandez, Fernando Alonso-Martin, Sara Marques-Villarroya, and Miguel A Salichs. Social robotics in therapy of apraxia of speech. *Journal of healthcare engineering*, 2018(1):7075290, 2018.

[28] Shruti Chandra, Garima Gupta, Torrey Loucks, and Kerstin Dautenhahn. Opportunities for social robots in the stuttering clinic: A review and proposed scenarios. *Paladyn, Journal of Behavioral Robotics*, 13(1):23–44, 2022.

[29] Shruti Chandra, Torrey Loucks, Gerardo Chavez Castaneda, and Kerstin Dautenhahn. Social Robots in the Stuttering Clinic: A Human-Centred Exploration with Speech Language Pathologists. In *International Conference on Social Robotics*, pages 299–313. Springer, 2022.

[30] Kathy Charmaz. Grounded theory: Objectivist and constructivist methods. *Handbook of qualitative research*, 2(1):509–535, 2000.

[31] Kathy Charmaz. *Constructing grounded theory: A practical guide through qualitative analysis*. sage, 2006.

[32] Tanya L Chartrand and John A Bargh. The chameleon effect: The perception–behavior link and social interaction. *Journal of personality and social psychology*, 76(6):893, 1999.

[33] Yen-Wei Chen, Kuan-Yi Li, Chu-Hsu Lin, Pei-Hsuan Hung, Hui-Tzu Lai, and Ching-Yi Wu. The effect of sequential combination of mirror therapy and robot-assisted therapy on motor function, daily function, and self-efficacy after stroke. *Scientific Reports*, 13(1):16841, 2023.

[34] Rune Haubo B Christensen. Analysis of ordinal data with cumulative link models—estimation with the r-package ordinal. *R-package version*, 28:406, 2015.

[35] Brett Christenson, Christine Ringler, and Nancy J Sirianni. Speaking fast and slow: How speech rate of digital assistants affects likelihood to use. *Journal of Business Research*, 163:113907, 2023.

[36] Michael Jae-Yoon Chung and Maya Cakmak. "How was Your Stay?": Exploring the Use of Robots for Gathering Customer Feedback in the Hospitality Industry. In *27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pages 947–954. IEEE, 2018.

[37] Michelle Cohn, Patrik Jonell, Taylor Kim, Jonas Beskow, and Georgia Zellou. Embodiment and gender interact in alignment to tts voices. In *CogSci*, 2020.

[38] Michelle Cohn, Ashley Keaton, Jonas Beskow, and Georgia Zellou. Vocal accommodation to technology: the role of physical form. *Language Sciences*, 99:101567, 2023.

[39] Tom Cole and Marco Gillies. More than a bit of coding:(un-) Grounded (non-) Theory in HCI. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts*, pages 1–11, 2022.

[40] Andreia P Costa, Georges Steffgen, Francisco Rodríguez Lera, Aida Nazarikhorram, and Pouyan Ziafati. Socially assistive robots for teaching emotional abilities to children with autism spectrum disorder. In *3rd Workshop on Child-Robot Interaction at HRI*, 2017.

[41] Matthew JA Craig, Xialing Lin, Chad Edwards, and Autumn Edwards. Can you repeat that again? investigating the mediating effects of perceived accommodation appropriateness for accommodative voice-based assistants. *Computers in Human Behavior: Artificial Humans*, 2(2):100102, 2024.

[42] Courtney Darves and Sharon Oviatt. Adaptation of users' spoken dialogue patterns in a conversational interface. In *Seventh International Conference on Spoken Language Processing*, 2002.

[43] Nivja H de Jong, Jos Pacilly, and Willemijn Heeren. Praat scripts to measure speed fluency and breakdown fluency in speech automatically. *Assessment in education: Principles, policy & practice*, 28(4):456–476, 2021.

[44] Jenny de Jong Gierveld, Suzan Van der Pas, and Norah Keating. Loneliness of older immigrant groups in canada: Effects of ethnic-cultural background. *Journal of cross-cultural gerontology*, 30:251–268, 2015.

[45] Ronan de Kervenoael, Rajibul Hasan, Alexandre Schwob, and Edwin Goh. Leveraging human-robot interaction in hospitality services: Incorporating the role of perceived value, empathy, and information sharing into visitors' intentions to use social robots. *Tourism Management*, 78:104042, 2020.

[46] Design Council. The Double Diamond Design Process Model. https://www.designcouncil.org.uk/our-resources/framework-for-innovation/. Accessed: 2024-09-12.

[47] Giada Devittori, Daria Dinacci, Davide Romiti, Antonella Califfi, Claudio Petrillo, Paolo Rossi, Raffaele Ranzani, Roger Gassert, and Olivier Lambercy. Unsupervised robot-assisted rehabilitation after stroke: feasibility, effect on therapy dose, and user experience. *Journal of NeuroEngineering and Rehabilitation*, 21(1):52, 2024.

[48] Saad Elbeleidy, Terran Mott, Dan Liu, and Tom Williams. Practical considerations for deploying robot teleoperation in therapy and telehealth. In *2022 31st IEEE RO-MAN*, pages 977–984. IEEE, 2022.

[49] Azita Emami, Sandra Torres, Juliene G Lipson, and Sirkka-Liisa Ekman. An ethnographic study of a day care center for iranian immigrant seniors. *Western journal of nursing research*, 22(2):169–188, 2000.

[50] Alireza Esfandbod, Zeynab Rokhi, Ali F Meghdari, Alireza Taheri, Minoo Alemi, and Mahdieh Karimi. Utilizing an emotional robot capable of lip-syncing in robot-assisted speech therapy sessions for children with language disorders. *International Journal of Social Robotics*, 15(2):165–183, 2023.

[51] David Estévez, María-José Terrón-López, Paloma J Velasco-Quintana, Rosa-María Rodríguez-Jiménez, and Valle Álvarez-Manzano. A case study of a robot-assisted speech therapy for children with language disorders. *Sustainability*, 13(5):2771, 2021.

[52] Vanessa Evers, Heidy C Maldonado, Talia L Brodecki, and Pamela J Hinds. Relational vs. group self-construal: untangling the role of national culture in HRI. In *Proceedings of the 3rd ACM/IEEE international conference on Human robot interaction*, pages 255–262, 2008.

[53] Galya Georgieva-Tsaneva, Anna Andreeva, Paulina Tsvetkova, Anna Lekova, Miglena Simonska, Vaska Stancheva-Popkostadinova, Georgi Dimitrov, Katia Rasheva-Yordanova, and Iva Kostadinova. Exploring the potential of social robots for speech and language therapy: a review and analysis of interactive scenarios. *Machines*, 11(7):693, 2023.

[54] Howard Giles, Nikolas Coupland, and Justine Coupland. Accommodation theory: Communication, context, and consequence. *Contexts of accommodation: Developments in applied sociolinguistics*, 1:1–68, 1991.

[55] Howard Giles, Anthony Mulac, J Bradac, and Patricia Johnson. Speech accommodation theory: The next decade and beyond. communication yearbook 10, 13-48, 1987.

[56] Glaser, Barney G. and Strauss, Anselm. *The discovery of grounded theory: strategies for qualitative research*. Aldine, Chicago, IL, 1967.

[57] Samuel D Gosling, Peter J Rentfrow, and William B Swann Jr. A very brief measure of the Big-Five personality domains. *Journal of Research in personality*, 37(6):504–528, 2003.

[58] Sheryl Ridener Gottwald and C Woodruff Starkweather. Fluency intervention for preschoolers and their families in the public schools. *Language, Speech, and Hearing Services in Schools*, 26(2):117–126, 1995.

[59] Lucrezia Grassi, Danilo Canepa, Amy Bellitto, Maura Casadio, Antonino Massone, Carmine Tommaso Recchiuto, and Antonio Sgorbissa. Diversity-Aware Verbal Interaction Between a Robot and People With Spinal Cord Injury. In *32nd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, pages 2512–2519. IEEE, 2023.

[60] Stephen Groner. The looooong list of 27 incredible stuttering treatment techniques. https://slpstephen.com/blogs/news/the-complete-list-of-stuttering-treatment-techniques, 2025. Accessed: March 20, 2025.

[61] Whoqol Group et al. The world health organization quality of life assessment (whoqol): position paper from the world health organization. *Social science & medicine*, 41(10):1403–1409, 1995.

[62] Barry Guitar. *Stuttering: An integrated approach to its nature and treatment*. Lippincott Williams & Wilkins, 6th edition, 2013.

[63] Barry Guitar and Lisa Marchinkoski. Influence of mothers' slower speech on their children's speech rate. 2001.

[64] Garima Gupta, Shruti Chandra, Kerstin Dautenhahn, and Torrey Loucks. Stuttering treatment approaches from the past two decades: comprehensive survey and review. *Journal of Student Research*, 11(2), 2022.

[65] Rashina Hoda. Socio-technical grounded theory for software engineering. *IEEE Transactions on Software Engineering*, 48(10):3808–3832, 2021.

[66] Andrew Gary Darwin Holmes. Researcher Positionality–A Consideration of Its Influence and Place in Qualitative Research–A New Researcher Guide. *Shanlax International Journal of Education*, 8(4):1–10, 2020.

[67] Takamasa Iio, Masahiro Shiomi, Kazuhiko Shinozawa, Takahiro Miyashita, Takaaki Akimoto, and Norihiro Hagita. Lexical entrainment in human-robot interaction: Can robots entrain human vocabulary? In *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 3727–3734. IEEE, 2009.

[68] Interaction Design Foundation - IxDF. Shadowing in User Research - Do You See What They See? https://www.interaction-design.org/literature/article/shadowing-in-user-research-do-you-see-what-they-see. Accessed: 2024-06-23.

[69] Stanislav Hristov Ivanov, Craig Webster, and Katerina Berezina. Adoption of Robots and Service Automation by Tourism and Hospitality Companies. *Revista Turismo & Desenvolvimento*, 27(28):1501–1517, 2017.

[70] Seyed Mohammad Sadegh Khaksar, Rajiv Khosla, and Mei-Tai Chu. Socially assistive robots in service innovation context to improve aged-care quality: A grounded theory approach. In *2015 7th IEEE CIS and IEEE RAM*, pages 161–166. IEEE, 2015.

[71] Oliver Korn, Neziha Akalin, and Ruben Gouveia. Understanding Cultural Preferences for Social Robots: A Study in German and Arab Communities. *ACM Transactions on Human-Robot Interaction (THRI)*, 10(2):1–19, 2021.

[72] Bettina-Johanna Krings and Nora Weinberger. Assistant without master? Some conceptual implications of assistive robotics in health care. *Technologies*, 6(1):13, 2018.

[73] Klaus Krippendorff. *Content analysis: An introduction to its methodology*. Sage publications, 2018.

[74] Łukasz Kwaśniewicz, Wiesława Kuniszyk-Jóźkowiak, Grzegorz M Wójcik, and Jolanta Masiak. Adaptation of the humanoid robot to speech disfluency therapy. *Bio-Algorithms and Med-Systems*, 12(4):169–177, 2016.

[75] Dimitri Lacroix, Ricarda Wullenkord, and Friederike Eyssel. Who's in Charge?: Using Personalization vs. Customization Distinction to Inform HRI Research on Adaptation to Users. In *Companion of the ACM/IEEE International Conference on Human-Robot Interaction*, pages 580–586, 2023.

[76] Gabriella Lakatos, Marina Sarda Gou, Patrick Holthaus, Luke Wood, Sílvia Moros, Vicky Litchfield, Ben Robins, and Farshid Amirabdollahian. A feasibility study of using Kaspar, a humanoid robot for speech and language therapy for children with learning disabilities. In *2023 32nd IEEE RO-MAN*, pages 1233–1238. IEEE, 2023.

[77] J Richard Landis and Gary G Koch. The Measurement of Observer Agreement for Categorical Data. *biometrics*, pages 159–174, 1977.

[78] Hee Rin Lee, Haodan Tan, and Selma Šabanović. That robot is not for me: Addressing stereotypes of aging in assistive robot design. In *IEEE RO-MAN*, pages 312–317. IEEE, 2016.

[79] Kwan Min Lee, Wei Peng, Seung-A Jin, and Chang Yan. Can robots manifest personality?: An empirical test of personality recognition, social responses, and social presence in human–robot interaction. *Journal of communication*, 56(4):754–772, 2006.

[80] Deirdre E Logan, Cynthia Breazeal, Matthew S Goodwin, Sooyeon Jeong, Brianna O'Connell, Duncan Smith-Freedman, James Heathers, and Peter Weinstock. Social robots for hospitalized children. *Pediatrics*, 144(1), 2019.

[81] Nichola Lubold, Erin Walker, and Heather Pon-Barry. Effects of adapting to user pitch on rapport perception, behavior, and state with a social robotic learning companion. *User Modeling and User-Adapted Interaction*, 31:35–73, 2021.

[82] Maxim Makatchev, Reid Simmons, Majd Sakr, and Micheline Ziadee. Expressing ethnicity through behaviors of a robot character. In *8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 357–364. IEEE, 2013.

[83] Nikolaos Mavridis. A review of verbal and non-verbal human–robot interactive communication. *Robotics and Autonomous Systems*, 63:22–35, 2015.

[84] George A Miller. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological review*, 63(2):81, 1956.

[85] Iain R Murray and John L Arnott. Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion. *The Journal of the Acoustical Society of America*, 93(2):1097–1108, 1993.

[86] J Nakanishi, I Kuramoto, J Baba, K Ogawa, Y Yoshikawa, and H Ishiguro. Continuous hospitality with social robots at a hotel. SN Appl Sci 2 (3): 1–13, 2020.

[87] Junya Nakanishi, Itaru Kuramoto, Jun Baba, Kohei Ogawa, Yuichiro Yoshikawa, and Hiroshi Ishiguro. Continuous Hospitality with Social Robots at a hotel. *SN Applied Sciences*, 2, 2020.

[88] Diego Nieto, Alexis Quesada-Arencibia, Carmelo R García, and Roberto Moreno-Díaz. A Social Robot in a Tourist Environment. In *Ubiquitous Computing and Ambient Intelligence. Personalisation and User Adapted Services: 8th International Conference, 2014. Proceedings 8*, pages 21–24. Springer, 2014.

[89] Rita Orji, Kiemute Oyibo, and Gustavo F Tondello. A Comparison of System-Controlled and User-Controlled Personalization Approaches. In *Adjunct publication*

*of the 25th conference on user modeling, adaptation and personalization*, pages 413–418, 2017.

[90] Mark L Ornelas, Gary B Smith, and Masoumeh Mansouri. Redefining culture in cultural robotics. *AI & SOCIETY*, 38(2):777–788, 2023.

[91] Yadong Pan, Haruka Okada, Toshiaki Uchiyama, and Kenji Suzuki. Direct and indirect social robot interactions in a hotel public space. In *International Conference on Robotics and Biomimetics (ROBIO)*, pages 1881–1886. IEEE, 2013.

[92] Yadong Pan, Haruka Okada, Toshiaki Uchiyama, and Kenji Suzuki. On the Reaction to Robot's Speech in a Hotel Public Space. *International Journal of Social Robotics*, 7:911–920, 2015.

[93] Chris Papadopoulos, Nina Castro, Abiha Nigath, Rosemary Davidson, Nicholas Faulkes, Roberto Menicatti, Ali Abdul Khaliq, Carmine Recchiuto, Linda Battistuzzi, Gurch Randhawa, et al. The CARESSES Randomised Controlled Trial: Exploring the Health-Related Impact of Culturally Competent Artificial Intelligence Embedded Into Socially Assistive Robots and Tested in Older Adult Care Homes. *International Journal of Social Robotics*, 14(1):245–256, 2022.

[94] Jennifer S Pardo. On phonetic convergence during conversational interaction. *The Journal of the Acoustical Society of America*, 119(4):2382–2393, 2006.

[95] Roberto Pinillos, Samuel Marcos, Raul Feliz, Eduardo Zalama, and Jaime Gómez-García-Bermejo. Long-term assessment of a service robot in a hotel environment. *Robotics and Autonomous Systems*, 79:40–57, 2016.

[96] Qualtrics. Qualtrics survey software. https://www.qualtrics.com, 2025. Accessed: March 17, 2025.

[97] Alec Radford, Jong Wook Kim, Tao Xu, Greg Brockman, Christine McLeavey, and Ilya Sutskever. Robust speech recognition via large-scale weak supervision. In *International conference on machine learning*, pages 28492–28518. PMLR, 2023.

[98] Samira Rasouli, Garima Gupta, Elizabeth Nilsen, and Kerstin Dautenhahn. Potential applications of social robots in robot-assisted interventions for social anxiety. *International Journal of Social Robotics*, 14(5):1–32, 2022.

[99] Giacomo Rizzolatti and Corrado Sinigaglia. The mirror mechanism: a basic principle of brain function. *Nature Reviews Neuroscience*, 17(12):757–765, 2016.

[100] Vladimir Robles-Bykbaev, Mario Ochoa-Guaraca, Marco Carpio-Moreta, Daniel Pulla-Sánchez, Luis Serpa-Andrade, Martín López-Nores, and Jorge García-Duque. Robotic assistant for support in speech therapy for children with cerebral palsy. In *2016 IEEE ROPEC*, pages 1–6. IEEE, 2016.

[101] Furhat Robotics. Furhat robot. https://furhatrobotics.com, 2025. Accessed: March 17, 2025.

[102] SoftBank Robotics. Nao robot. https://www.softbankrobotics.com, 2025. Accessed: March 17, 2025.

[103] Eduardo Rodriguez-Lizundia, Samuel Marcos, Eduardo Zalama, Jaime Gómez-García-Bermejo, and Alfonso Gordaliza. A bellboy robot: Study of the effects of robot behaviour on user engagement and comfort. *International Journal of Human-Computer Studies*, 82:83–95, 2015.

[104] Maha Salem, Micheline Ziadee, and Majd Sakr. Marhaba, how may i help you?: effects of politeness and culture on robot acceptance and anthropomorphization. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction*, pages 74–81, 2014.

[105] Maggi Savin-Baden and Claire Howell Major. *Qualitative Research: The Essential Guide to Theory and Practice*. Routledge, Oxford, 2013.

[106] Jean Sawyer, Colleen Matteson, Hua Ou, and Takahisa Nagase. The effects of parent-focused slow relaxed speech intervention on articulation rate, response time latency, and fluency in preschool children who stutter. *Journal of Speech, Language, and Hearing Research*, 60(4):794–809, 2017.

[107] Antonio Sgorbissa, Alessandro Saffiotti, Nak Young Chong, Linda Battistuzzi, Roberto Menicatti, Federico Pecora, Irena Papadopoulos, Amit Kumar Pandey, Hiroko Kamide, Christina Koulouglioti, Sanjeev Kanoria, Raffaele Mastrolonardo, Chris Papadopoulos, Len Merton, Jaeryoung Lee, Gurch Randhawa, and Yuto Lim. CARESSES:The Flower that Taught Robots about Culture. In *14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 371–371, 2019.

[108] Suleman Shahid, Emiel Krahmer, and Marc Swerts. Child–robot interaction across cultures: How does playing a game with a social robot compare to playing a game alone or with a friend? *Computers in Human Behavior*, 40:86–100, 2014.

[109] Murray Shanahan. Talking about large language models. *Communications of the ACM*, 67(2):68–79, 2024.

[110] Mohammad Shidujaman and Haipeng Mi. "Which Country Are You from?" A Cross-Cultural Study on Greeting Interaction Design for Social Robots. In *Cross-Cultural Design. Methods, Tools, and Users: 10th International Conference, Proceedings, Part I 10*, pages 362–374. Springer, 2018.

[111] Supriya Singh, Kylie Bartolo, et al. Grounded Theory and User Requirements: a challenge for qualitative research. *Australasian Journal of Information Systems*, 12(2), 2005.

[112] Fran Soljacic, Theresa Law, Meia Chita-Tegmark, and Matthias Scheutz. Robots in healthcare as envisioned by care professionals. *Intelligent Service Robotics*, pages 1–17, 2024.

[113] Micol Spitale, Silvia Silleresi, Franca Garzotto, and Maja J Matarić. Using socially assistive robots in speech-language therapy for children with language impairments. *International Journal of Social Robotics*, 15(9):1525–1542, 2023.

[114] Stamurai Team. Stuttering: 6 fluency-shaping techniques. https://www.stamurai.com/blog/stuttering-fluency-shaping-techniques/, 2021. Accessed: March 20, 2025.

[115] Anselm Leonard Strauss and Juliet M. Corbin. *Basics of qualitative research*, volume 15. Sage, Newbury Park, CA, 1990.

[116] Alexandra Tanner, Sarah Bestgen, and Hartmut Schulze. Social robots to improve addiction treatment: development and assessment of a digital intervention. *Gruppe. Interaktion. Organisation. Zeitschrift für Angewandte Organisationspsychologie (GIO)*, pages 1–12, 2025.

[117] Gabriele Trovato, Martin Do, Masuko Kuramochi, Massimiliano Zecca, Ömer Terlemez, Tamim Asfour, and Atsuo Takanishi. A Novel Culture-Dependent Gesture Selection System for a Humanoid Robot Performing Greeting Interaction. In *Social Robotics: 6th International Conference*, pages 340–349. Springer, 2014.

[118] Gabriele Trovato, Massimiliano Zecca, Salvatore Sessa, Lorenzo Jamone, Jaap Ham, Kenji Hashimoto, and Atsuo Takanishi. Cross-cultural study on human-robot greeting interaction: acceptance and discomfort by Egyptians and Japanese. *Journal of Behavioral Robotics*, 4(2):83–93, 2013.

[119] Vincent Wing Sun Tung and Norman Au. Exploring customer experiences with robotics in hospitality. *International Journal of Contemporary Hospitality Management*, 30(7):2680–2697, 2018.

[120] Tom Williams. The Eye of the Robot Beholder: Ethical Risks of Representation, Recognition, and Reasoning over Identity Characteristics in Human-Robot Interaction. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction*, 2023.

[121] Olga A Wudarczyk, Murat Kirtay, Doris Pischedda, Verena V Hafner, John-Dylan Haynes, Anna K Kuhlen, and Rasha Abdel Rahman. Robots facilitate human language production. *Scientific Reports*, 11(1):16737, 2021.

[122] J Scott Yaruss, Craig Coleman, and David Hammer. Treating preschool children who stutter: Description and preliminary evaluation of a family-focused treatment approach. 2006.

[123] Yuichiro Yoshikawa, Hiroaki Kobayashi, Naomi Sakai, Hiroshi Ishiguro, and Hirokazu Kumazaki. Therapeutic potential of robots for people who stutter: a preliminary study. *Frontiers in Psychiatry*, 15:1298626, 2024.

[124] Fengpei Yuan, Elizabeth Klavon, Ziming Liu, Ruth Palan Lopez, and Xiaopeng Zhao. A systematic review of robotic rehabilitation for cognitive training. *Frontiers in Robotics and AI*, 8:605715, 2021.

[125] Eduardo Zalama, Jaime Gómez García-Bermejo, Samuel Marcos, Salvador Domínguez, Raúl Feliz, Roberto Pinillos, and Joaquín López. Sacarino, a Service Robot in a Hotel Environment. In *ROBOT2013: First Iberian Robotics Conference: Advances in Robotics, Vol. 2*, pages 3–14. Springer, 2014.

[126] Georgia Zellou, Michelle Cohn, and Bruno Ferenc Segedin. Age-and gender-related differences in speech alignment toward humans and voice-ai. *Frontiers in Communication*, 5:600361, 2021.

# APPENDICES

# Appendix A

# Ethics Clearance Certificates

# Study 1 Ethics Clearance Certificate

## UNIVERSITY OF WATERLOO

**Notification of Ethics Clearance to Conduct Research with Human Participants**

---

Principal Investigator: Kerstin Dautenhahn

Student investigator: Priyank Avijeet

Collaborator: Pourya Aliasghari

File #: 45545

Title: Social Robots and Cross-cultural Greetings in the hospitality industry

---

The Human Research Ethics Board is pleased to inform you this study has been reviewed and given ethics clearance.

**Initial Approval Date: 09/25/23 (m/d/y)**

University of Waterloo Research Ethics Boards are composed in accordance with, and carry out their functions and operate in a manner consistent with, the institution's guidelines for research with human participants, the Tri-Council Policy Statement for the Ethical Conduct for Research Involving Humans (TCPS2 2022), the Ontario Personal Health Information Protection Act (PHIPA), and all laws and regulations of the province of Ontario (as applicable). Additionally, CREB operates in a manner consistent with the International Conference for Harmonization of Technical Requirements for Pharmaceuticals for Human Use (ICH) Guidance E6(R2): Good Clinical Practice , the International Organization for Standardization of Good Clinical Practices (GCP) as set out by ISO 14155 - Clinical investigation of medical devices for human subjects, Part C, Division 5 of the Food and Drug Regulations, Part 4 of the Natural Health Products Regulations, Part 3 of the Medical Devices Regulations. Both Boards are registered with the U.S. Department of Health and Human Services under the Federal Wide Assurance, FWA00021410, and IRB registration number IRB00002419 (HREB) and IRB00007409 (CREB).

**Expiry Date: 09/26/24 (m/d/y)**

Multi-year research must be renewed at least once every 12 months unless a more frequent review has otherwise been specified. Studies will only be renewed if the renewal report is received and approved before the expiry date. Failure to submit renewal reports will result in the investigators being notified ethics clearance has been suspended and Research Finance being notified the ethics clearance is no longer valid.

Level of review: Delegated Review

Signed on behalf of the Human Research Ethics Board



Heather Dekker, Ethics Advisor, hdekker@uwaterloo.ca, 519-888-4567, ext. 41506

This above named study is to be conducted in accordance with the submitted application and the most recently approved versions of all supporting materials.

Documents reviewed and received ethics clearance for use in the study and/or received for information:

file: other_attachments_email_thread_V2_202309.pdf

file: other_attachments_email_thread_2_V1_202309.pdf

file: feedback_letter_uni_V3_202309.pdf

file: InformationAndConsentForm_V4_202309.pdf

file: Poster_V2_202308.pdf

file: email_V1_202307.pdf

file: Email-Link_V1_202307.pdf

file: Email-Clarification_V1_202307.pdf

file: Procedue_V3_202309.pdf

file: Sample_video_1_v1_202308.mp4

file: Sample_video_2_202308.mp4

file: Sample_video_3_202308.mp4

Approved Protocol Version 4 in Research Ethics System

**This is an official document. Retain for your files.**

**You are responsible for obtaining any additional institutional approvals that might be required to complete this study.**

# Study 3 Ethics Clearance Certificate

## UNIVERSITY OF WATERLOO

**Notification of Ethics Clearance to Conduct Research with Human Participants**

---

Principal Investigator: Kerstin Dautenhahn

Student investigator: Priyank Avijeet

File #: 46793

Title: Interactions with a Voice System

---

The Human Research Ethics Board is pleased to inform you this study has been reviewed and given ethics clearance.

**Initial Approval Date: 11/18/24 (m/d/y)**

University of Waterloo Research Ethics Boards are composed in accordance with, and carry out their functions and operate in a manner consistent with, the institution's guidelines for research with human participants, the Tri-Council Policy Statement for the Ethical Conduct for Research Involving Humans (TCPS2 2022), the Ontario Personal Health Information Protection Act (PHIPA), and all laws and regulations of the province of Ontario (as applicable). Additionally, CREB operates in a manner consistent with the International Conference for Harmonization of Technical Requirements for Pharmaceuticals for Human Use (ICH) Guidance E6(R2): Good Clinical Practice , the International Organization for Standardization of Good Clinical Practices (GCP) as set out by ISO 14155 - Clinical investigation of medical devices for human subjects, Part C, Division 5 of the Food and Drug Regulations, Part 4 of the Natural Health Products Regulations, Part 3 of the Medical Devices Regulations. Both Boards are registered with the U.S. Department of Health and Human Services under the Federal Wide Assurance, FWA00021410, and IRB registration number IRB00002419 (HREB) and IRB00007409 (CREB).

**Expiry Date: 11/19/25 (m/d/y)**

Multi-year research must be renewed at least once every 12 months unless a more frequent review has otherwise been specified. Studies will only be renewed if the renewal report is received and approved before the expiry date. Failure to submit renewal reports will result in the investigators being notified ethics clearance has been suspended and Research Finance being notified the ethics clearance is no longer valid.

Level of review: Delegated Review

Signed on behalf of the Human Research Ethics Board



Heather Dekker, Ethics Advisor, hdekker@uwaterloo.ca, 519-888-4567, ext. 41506

This above-named study is to be conducted in accordance with the submitted application and the most recently approved versions of all supporting materials.

Documents reviewed and received ethics clearance for use in the study and/or received for information:

file: other_attachments_email_thread_2_V1_20241003.pdf

file: other_attachments_email_thread_V1_20241003.pdf

file: QT_docs_V1_20241001.zip

file: feedback_letter_uni_V2_20241107.pdf

file: information_sheet_V2_20241107.pdf

file: Online_ConsentForm_V2_20241107.pdf

file: Poster_V2_20241107.pdf

file: email_V2_20241107.pdf

file: tipi.pdf

file: The-Self-Assessment-Manikin-SAM-SAM-scale-used-in-the-experiment-to-capture.png

file: SAM.pdf

file: email-link_V2_20241107.pdf

file: Procedue_V2_20241107.pdf

file: bartneckGodspeedChapter2023.pdf

file: flow_chart_V3_20241107.pdf

file: Sample_Script_V1_20241109.pdf

file: debriefing_letter_V3_20241107.pdf

file: in-lab_debriefing_verbal_script_V3_20241107.pdf

Approved Protocol Version 2 in Research Ethics System

**This is an official document. Retain for your files.**

**You are responsible for obtaining any additional institutional approvals that might be required to complete this study.**

# Appendix B

# Study 1 - Study Materials (Flyer, Web Interface and Questionnaires)

# Study 1 - Flyer

Receive $3 in WatCard for 10 mins!!!

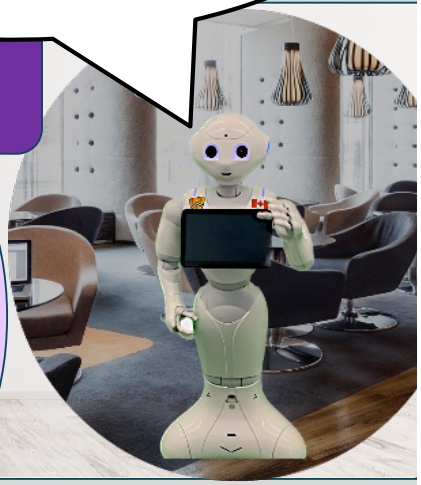
Huān yíng

As-salamu alaykum

Watch the Pepper Robot greet you



If one of these is your ethnic greeting, then please participate in our study!



- All participants aged 18+ from Chinese and Muslim communities are welcome
- 10 minutes
- $ 3
- Questionnaire data will be collected
- Online

We recruit from two communities that are among the University of Waterloo's major demographic groups. Participate to help us understand your cultural expectations from service robots and move one step closer to culturally inclusive social robots.

To participate please contact Priyank Avijeet:

SIRRL Social and Intelligent Robotics Research Laboratory

UNIVERSITY OF WATERLOO

pavijeet@uwaterloo.ca

pavijeet@uwaterloo.ca

# Study 1 - Web Interface and Questionnaires

UNIVERSITY OF WATERLOO

## Information and Consent Form

**Research Team:**

Principal Investigator: Professor Kerstin Dautenhahn, Department of Electrical and Computer Engineering, University of Waterloo, Canada
Collaborator: Pourya Aliasghari, Department of Electrical and Computer Engineering, University of Waterloo, Canada
Student Investigator: Priyank Avijeet, Department of Electrical and Computer Engineering, University of Waterloo, Canada

Overview: In this study, we would like to ask your opinion about a service robot that welcomes you using cross-cultural greetings in different hospitality industry settings. Researchers are working on understanding people's likeability towards culturally competent robots across different settings. We recruit from two communities that are among the University of Waterloo's major demographic groups. In this online study, we will ask you to respond to short questionnaires each time you observe videos of a service robot performing a greeting in a particular setting.

Procedure: The study will start after you give consent by checking the box and pressing the button at the end of this page.

First, you will be asked to complete a short questionnaire about your demographic information (only your gender, age, ethnicity, occupation, and

educational level). You may choose not to answer certain questions. Next, we will ask you some questions related to your personality and a few other aspects about yourself.

Next, we will show you a short video of the robot where it introduces itself so that you get a bit accustomed to its voice and appearance. Then, we will proceed to the main experiment to watch videos in which you are welcomed in a Chinese restaurant and an Arabic restaurant.

In each restaurant, you will experience two or three greetings. You will be then asked to rate your experience through short questionnaires. In total, you will see 5 videos each lasting 25-30 seconds. You will be completing the study by an online survey operated by Qualtrics. Qualtrics has implemented technical, administrative, and physical safeguards to protect the information provided via the Services from loss, misuse, and unauthorized access, disclosure, alteration, or destruction. However, no Internet transmission is ever fully secure or error-free. Please Note: We do not collect or use Internet Protocol (IP) addresses or other information which could link your participation to your computer or electronic device. In the post-study questionnaire, we ask two open-ended questions to get a better understanding of your experience.

Finally, you will reach the end of the experiment and submit your responses. You will be asked to click a link that redirects you to a remuneration details page. You will provide your Full Name and University of Waterloo Student/Employee ID as per your WatCard along with the Remuneration Code which you will get in the end message of this study. We will send you your remuneration after the study has been completed. Personal information, Name and Student/Employee ID information collected in the remuneration details page will be stored separately from your responses and will be deleted once the entire remuneration step has been completed.

**I. Your responsibilities as a participant**

Participation in this study will approximately take 15-20 minutes. Your answers to the questions will be logged. This does not require you to download anything onto your computer. We ask you not to switch to other tasks during the experiment.

You can complete this study only once. Your responses are valuable information that can help us in our research. You need to complete this task on a desktop computer or laptop, as phone or tablet screens would be too small for this study. You need to be able to load videos for this study. Please pay attention and provide the answers as accurately as you can.

The remuneration amount received is taxable. It is your responsibility to report this amount for income tax purposes.

**II. Your rights as a participant**

Participation in this study is entirely voluntary, and you have the right to choose not to answer any question(s) within the demographic information form. At any point, you are free to discontinue your participation.

Remuneration: You will receive $6 loaded to your WatCard if you complete all the steps of the experiment. If you do not wish to continue at any point, your participation in the study will end. In the event of partial completion, should you wish to claim remuneration, please feel free to reach out to us at pavijeet@uwaterloo.ca. In case you have completed less than half of the questions in the study, we will credit $3 to your WatCard. Otherwise, if you have completed more than half, we will credit $6 to your WatCard.

Benefits of the study: Participation in this study may not provide any personal benefit to you. However, the study will benefit the academic community and the research team. We plan to publish the results of this study, and researchers in the field of social robotics can potentially benefit from that. We hope the outcomes of this experiment pave the way for designing more intelligent and culturally competent robotic systems.

Confidentiality of identities and data: You will not be identified individually in any written reports of this research. Note that when information is transmitted over the internet, privacy cannot be entirely guaranteed. There is always a risk that your responses may be intercepted by a third party (e.g., government agencies or hackers). Any confidential information (namely your name) will be removed from our records after 1 month of your participation. You may withdraw your consent and ask us to delete your data within this period by contacting pavijeet@uwaterloo.ca. The data captured will be securely stored on a password-protected lab server with access only for the researchers. Anonymized data may be published publicly in the future and will be stored for a minimum of 7 years. Data may be deposited in an online public repository/database. This process is integral to the research process as it allows other researchers to verify results and avoid duplicating research.

**III. Questions, comments, or concerns**

Ethics clearance: This study has been reviewed and received ethics clearance through the University of Waterloo Human Research Ethics Board (REB #45545). If you have questions for the Board, contact the Office of Research Ethics, toll-free at 1-833-643-2379 (Canada and USA), 1-519-888 4440, or reb@uwaterloo.ca.

Contact Information: If you have any questions regarding this study, or would like additional information about your participation, please contact us by email at pavijeet@uwaterloo.ca.

**IV. Consent**

Thank you for your interest in our research and your assistance with this project. You must be 18 years of age or older to consent to take part in this study. By accepting this consent form, you are not waiving your legal rights or releasing the

investigator(s) or involved institution(s) from their legal and professional responsibilities.

You may save this page for your future reference if you prefer so.

| | I agree | I disagree |
|---|---|---|
| With full knowledge of all foregoing, I agree, of my own free will, to participate in this study. | ○ | ○ |

**Please note:** The data from this study may be shared in an online public repository/database. This process is integral to the research process as it allows other researchers to verify results and avoid duplicating research.

| | I agree | I disagree |
|---|---|---|
| I allow my data to be shared in a repository or archive for other researchers to verify results and avoid duplicating research. | ○ | ○ |

Please enter your name:

→

Powered by Qualtrics




Please select your ethnic greeting that matches your cultural background:

- (●) Huan Ying
- ( ) As-salamu Alaykum
- ( ) Other

→

Powered by Qualtrics

UNIVERSITY OF WATERLOO

Before we can start, we need a few information about you (You may skip any question that you are not comfortable answering):

Gender

Age

Ethnicity

Occupation

Education

In which country or countries have you resided for more than 6 months? Please provide the names of the countries along with the approximate number of years you spent in each, using commas to separate multiple countries. For instance: Canada (4 years), China (18 years).

→

Powered by Qualtrics

UNIVERSITY OF WATERLOO

How much do you identify with the Western culture?

A little A lot

How much do you identify with Chinese culture?

A little A lot

→

Powered by Qualtrics

UNIVERSITY OF WATERLOO

I see myself as:

| | Disagree strongly | Disagree moderately | Disagree a little | Neither agree nor disagree | Agree a little | Agree moderately | Agr stro |
|---|---|---|---|---|---|---|---|
| Extraverted, Enthusiastic | ● | ○ | ○ | ○ | ○ | ○ | ○ |
| Critical, Quarrelsome | ○ | ○ | ● | ○ | ○ | ○ | ○ |
| Dependable, Self-disciplined | ○ | ○ | ○ | ○ | ○ | ○ | ○ |
| Anxious, Easily upset | ○ | ○ | ○ | ● | ○ | ○ | ○ |
| Open to new experiences, Complex | ○ | ○ | ○ | ● | ○ | ○ | ○ |
| Reserved, Quiet | ○ | ○ | ○ | ● | ○ | ○ | ○ |
| Sympathetic, Warm | ○ | ○ | ○ | ○ | ● | ○ | ○ |
| Disorganized, Careless | ○ | ○ | ○ | ○ | ○ | ● | ○ |
| Calm, Emotionally stable | ○ | ○ | ○ | ○ | ○ | ● | ○ |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Conventional, Uncreative | ○ | ○ | ○ | ○ | ○ | ○ | ◐ |

→

Powered by Qualtrics



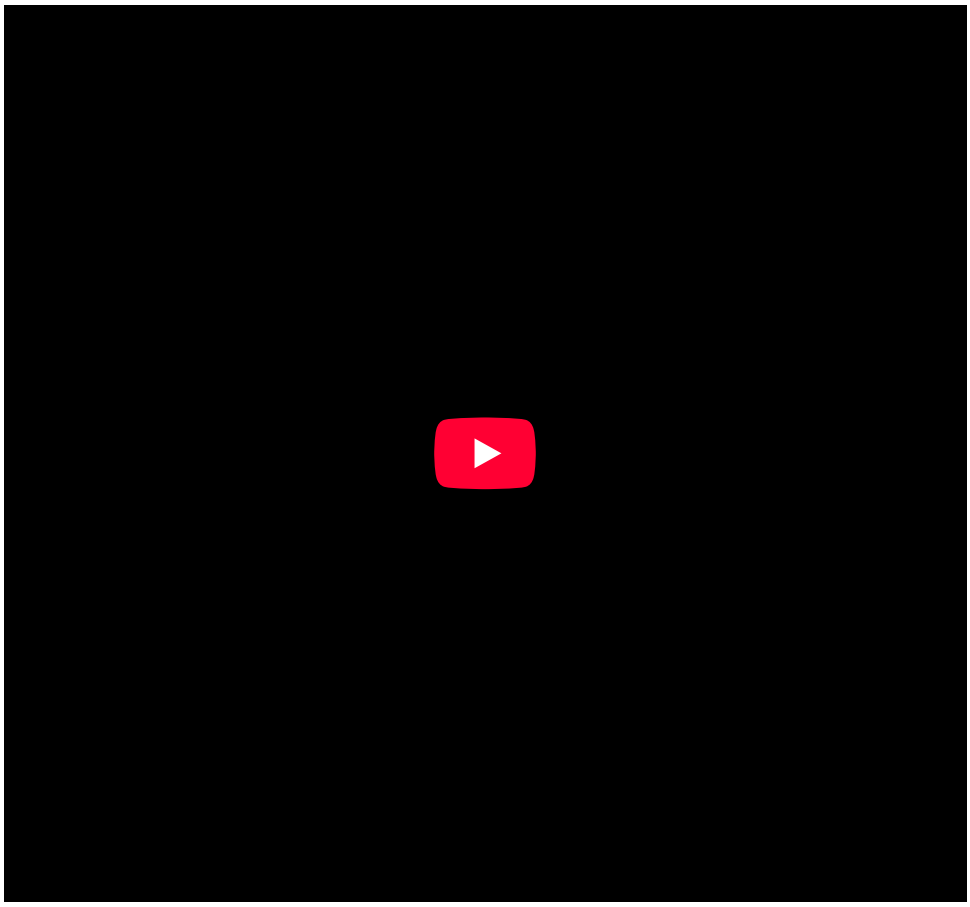

Please unmute and listen carefully while watching the videos.

→

Powered by Qualtrics

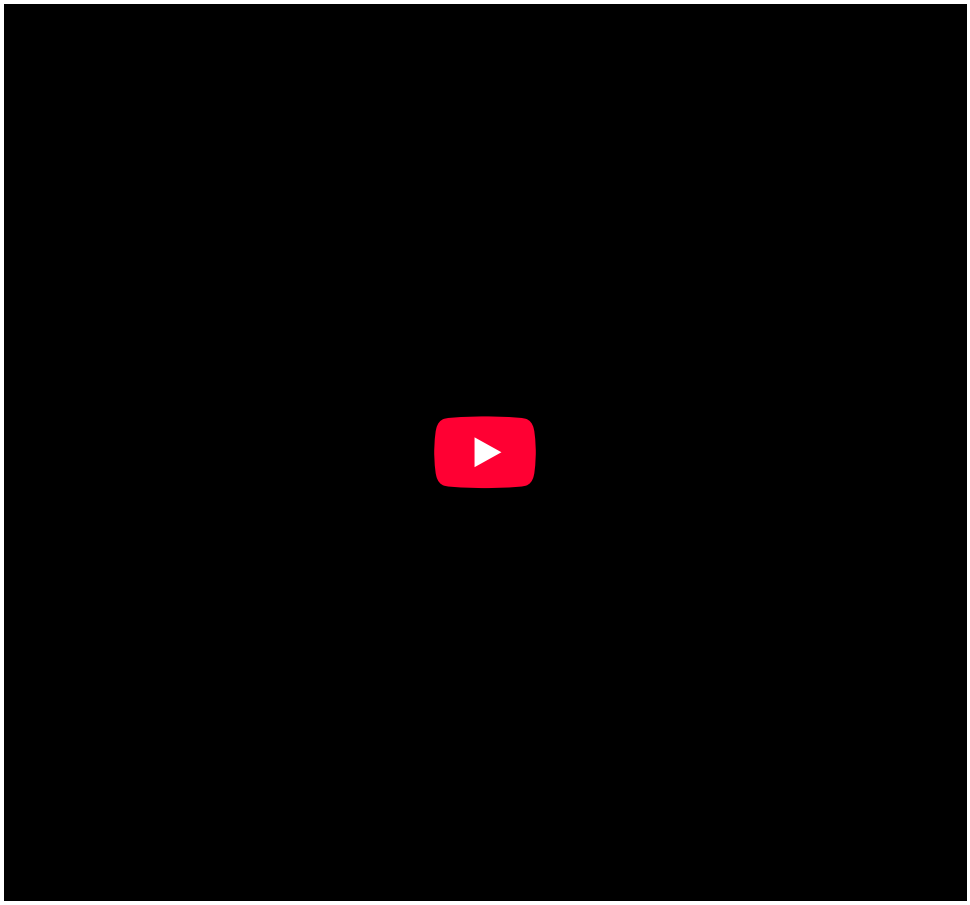
UNIVERSITY OF WATERLOO

First, please carefully watch this video:



UNIVERSITY OF WATERLOO

Please carefully watch this video:



Please rate your experience with this particular greeting:

Dislike Like



Unfriendly Friendly



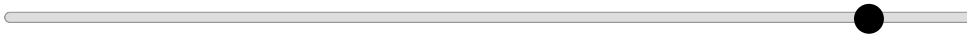
Unkind Kind



Unpleasant Pleasant



Awful Nice



In which type of restaurant did the Pepper Robot welcome you in the above

In which type of restaurant did the Pepper Robot welcome you in the above video?

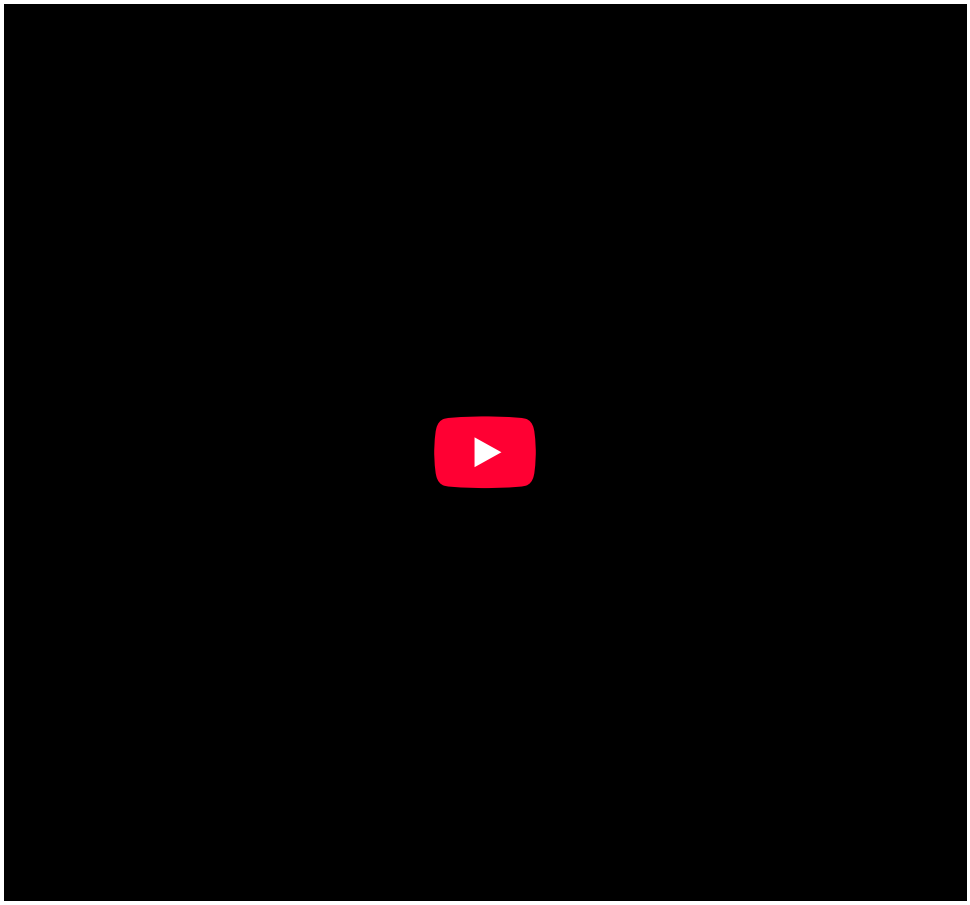
- ○ Spanish
- ● Arabic
- ○ Chinese
- ○ French

→

Powered by Qualtrics

UNIVERSITY OF WATERLOO

Please carefully watch this video:



Please rate your experience with this particular greeting:

Dislike Like



Unfriendly Friendly



Unkind Kind



Unpleasant Pleasant



Awful Nice



In which type of restaurant did the Pepper Robot welcome you in the above

In which type of restaurant did the Pepper Robot welcome you in the above video?

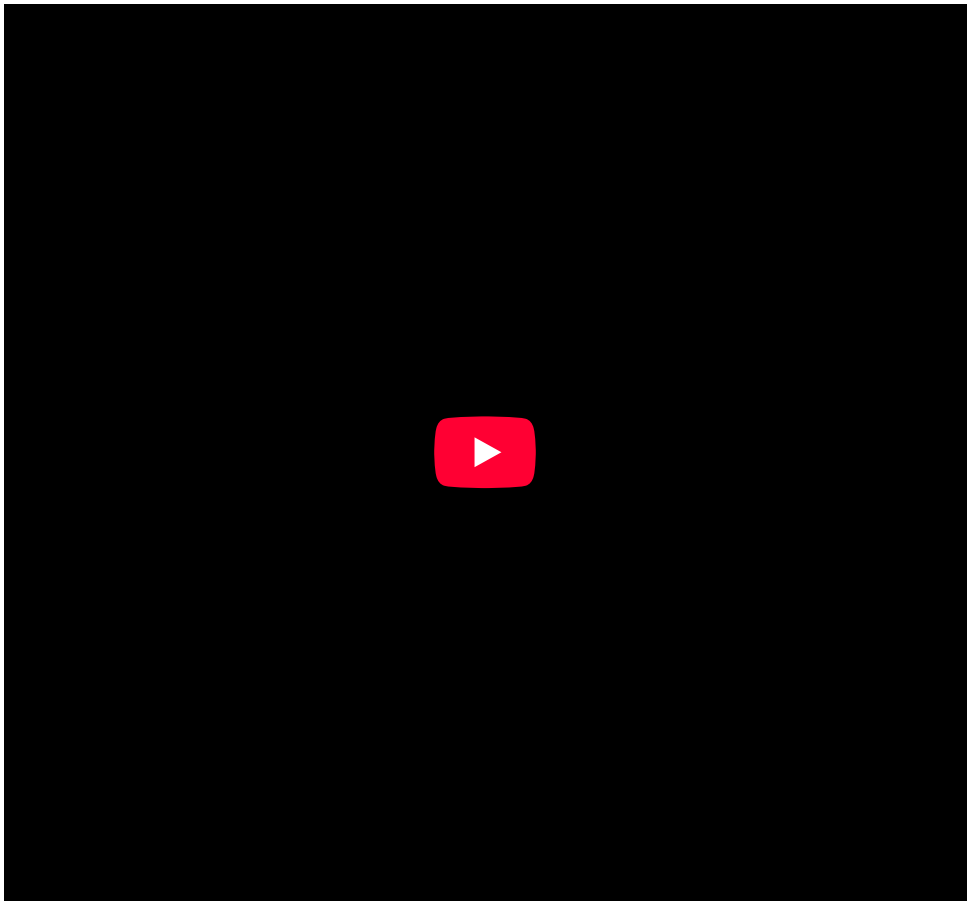
- (●) Arabic
- ( ) American
- ( ) Chinese
- ( ) Mexican

→

Powered by Qualtrics

UNIVERSITY OF WATERLOO

Please carefully watch this video:



Please rate your experience with this particular greeting:

Dislike Like



Unfriendly Friendly



Unkind Kind



Unpleasant Pleasant



Awful Nice



In which type of restaurant did the Pepper Robot welcome you in the above

In which type of restaurant did the Pepper Robot welcome you in the above video?

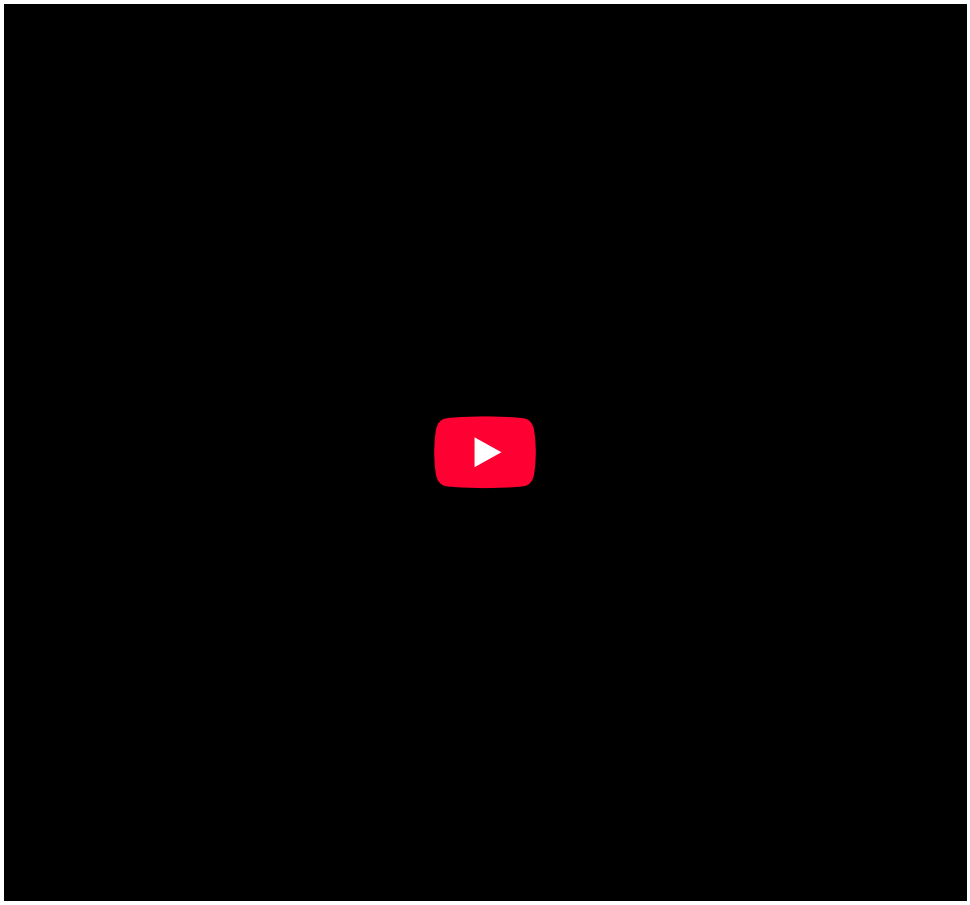
- ○ Thai
- ○ Greek
- ● Arabic
- ○ Chinese

→

Powered by Qualtrics

UNIVERSITY OF WATERLOO

Please carefully watch this video:



Please rate your experience with this particular greeting:

Dislike Like



Unfriendly Friendly



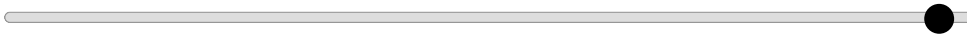
Unkind Kind



Unpleasant Pleasant



Awful Nice



In which type of restaurant did the Pepper Robot welcome you in the above

In which type of restaurant did the Pepper Robot welcome you in the above video?

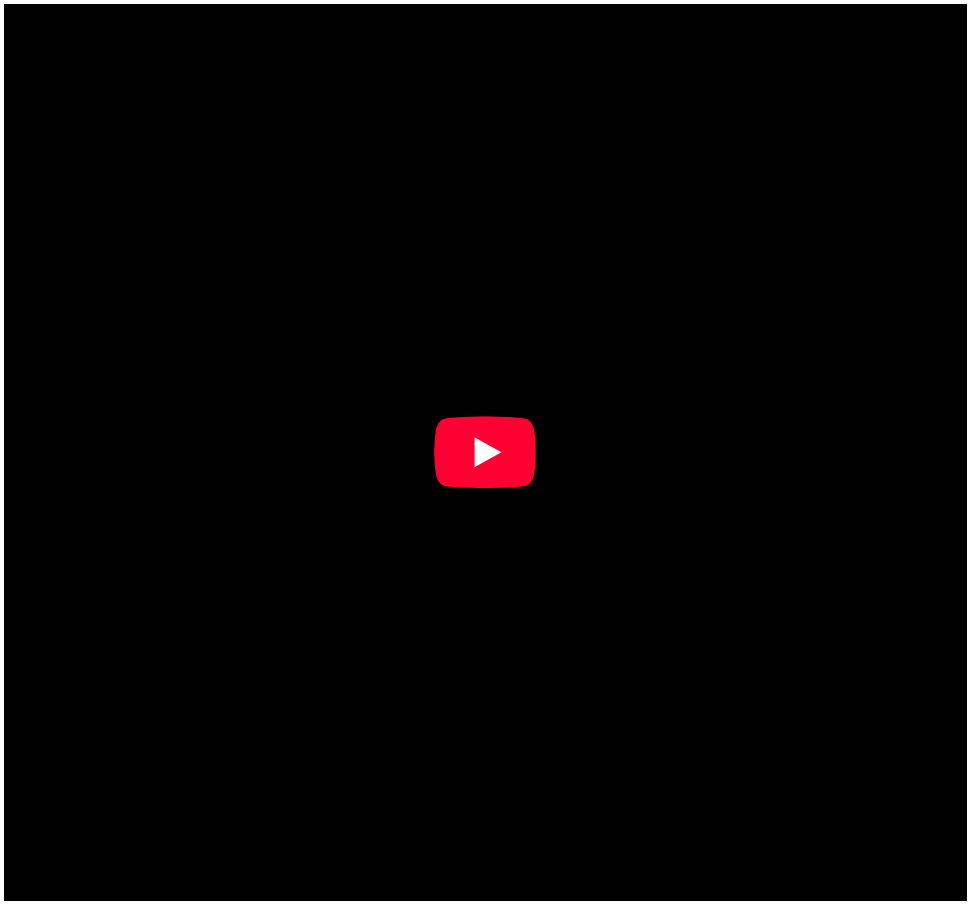
- (●) Chinese
- ( ) Arabic
- ( ) Mexican
- ( ) Thai

→

Powered by Qualtrics

UNIVERSITY OF WATERLOO

Please carefully watch this video:



Please rate your experience with this particular greeting:

Dislike Like



Unfriendly Friendly



Unkind Kind



Unpleasant Pleasant



Awful Nice



In which type of restaurant did the Pepper Robot welcome you in the above

In which type of restaurant did the Pepper Robot welcome you in the above video?

- ○ Indian
- ● Chinese
- ○ Greek
- ○ Arabic

→

Powered by Qualtrics

UNIVERSITY OF WATERLOO

In a restaurant setting that **matches** your cultural background (for example, a Chinese-themed restaurant), would you prefer to be welcomed by a social robot that aligns with your cultural background or follows popular Western culture customs, or do you have any other preferences? Please provide your choice and reasons to support your preference.

In a restaurant setting that **does not match** your cultural background (for example, an Italian or Japanese-themed restaurant), would you prefer to be welcomed by a social robot that aligns with your cultural background or follows popular Western culture customs, or do you have any other preferences? Please provide your choice and reasons to support your preference.

→

Powered by Qualtrics

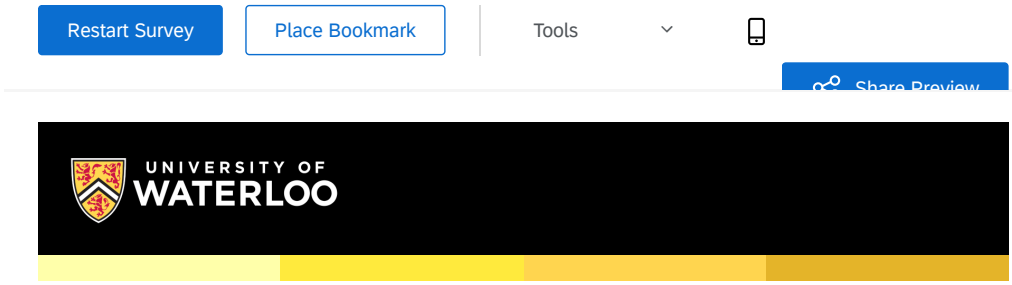
UNIVERSITY OF WATERLOO

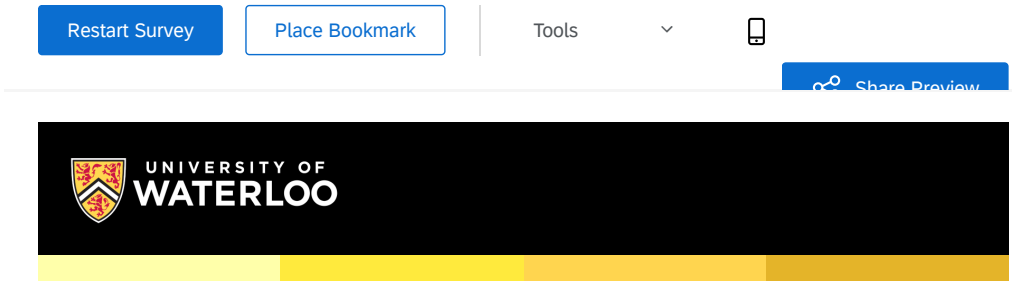
**Project Title:** ***Social Robots and Cross-cultural greetings in the hospitality industry***

**Student Investigator:** ***Priyank Avijeet, pavijeet@uwaterloo.ca***
**Collaborator:** ***Pourya Aliasghari, pourya.aliasghari@uwaterloo.ca***
**Faculty Advisor:** ***Kerstin Dautenhahn, kerstin.dautenhahn@uwaterloo.ca***

We appreciate your participation in our study and thank you for spending the time helping us with our research!
You will provide your Full Name and University of Waterloo Student/Employee ID as per your WatCard along with the Remuneration Code for remuneration. **Your remuneration code is R5S21322AR0I2L, please copy and save it somewhere to use in the next step. Please click this link Remuneration Details Link and provide information to get the remuneration.** Personal information, Name and Student/Employee ID information collected in the remuneration details page will be stored separately from your responses and will be deleted once the entire remuneration step has been completed.
You will receive the remuneration amount of $6 loaded to your WatCard after the study has completed.
In this study, we recorded your experience with different cross-cultural greetings in different scenarios by Pepper as a hospitality service robot. This will help us study the impact of location on people's preferences of cross-cultural greetings and the differences cross-cultural greetings could make in comparison to popular Western greetings.
Electronic data and video recordings will be kept for at least 7 years on a secure computer in a locked room in E7. All identifying information will be removed from the records prior to storage. Your name will not be included or in any other way associated with the data collected in the study.
This study has been reviewed and received ethics clearance through a University of Waterloo Research Ethics Board (REB #45545). If you have questions for the Board, contact the Office of Research Ethics, toll-free at 1-833-643-2379 (Canada and USA), 1-519-888-4440, or reb@uwaterloo.ca.

If you have any questions, contact Priyank Avijeet at pavijeet@uwaterloo.ca.
We really appreciate your participation, and hope that this has been an interesting experience for you.


Restart Survey
Place Bookmark
Tools
Share Preview
UNIVERSITY OF
WATERLOO
Please enter your Name:
Please enter your University of Waterloo Student ID:
Please enter your unique Remuneration Code:
→
Powered by Qualtrics




We thank you for your time spent taking this survey.
Your response has been recorded.

Powered by Qualtrics

# Appendix C

# Study 1 - Extra Plots and Figures

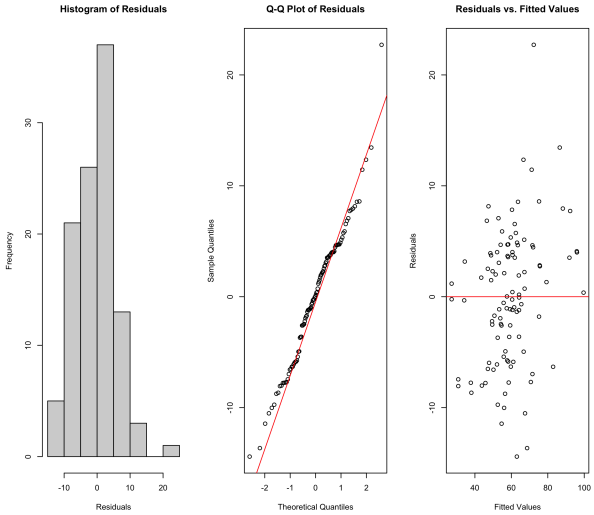



Figure C.1: Residuals Assumption Check for HY Scenario 1

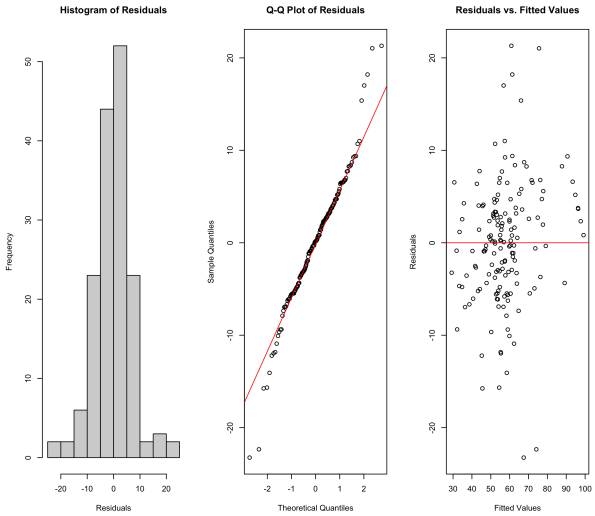



Figure C.2: Residuals Assumption Check for HY Scenario 2

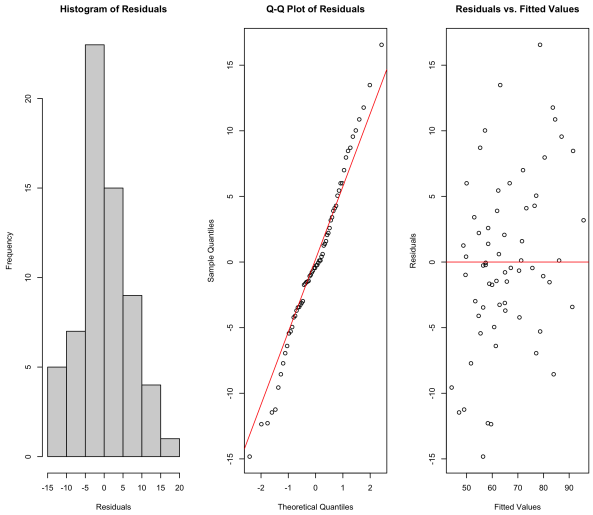



Figure C.3: Residuals Assumption Check for ASA Scenario 1

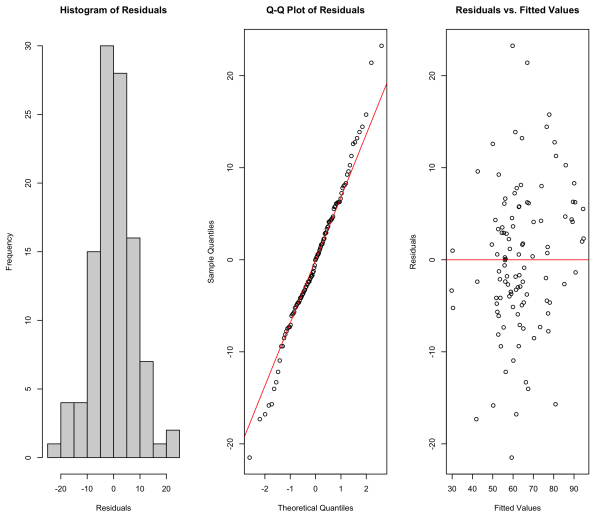



Figure C.4: Residuals Assumption Check for ASA Scenario 2

# Appendix D

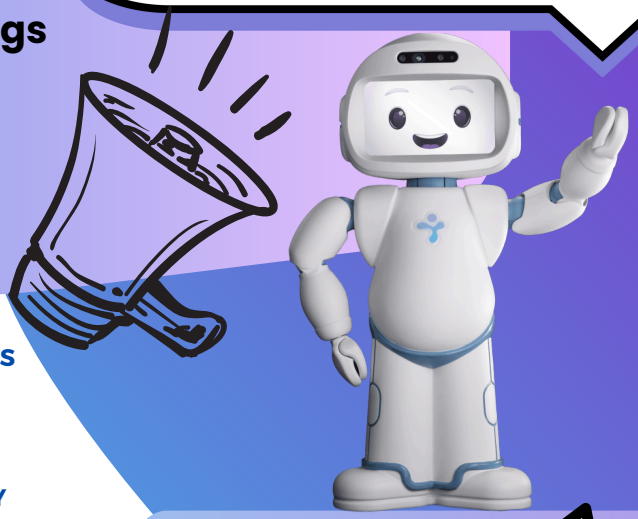
# Study 3 - Study Materials (Flyer, Scripts, Web Interface and Questionnaires)

# Study 3 - Flyer

Hear recommendations from our Voice System and let it know about yourself and the things you like—join our exciting study now!

# HEAR ME OUT!

**WE ARE LOOKING FOR PARTICIPANTS**



- **UNIVERSITY OF WATERLOO STUDENTS ARE ELIGIBLE**
- **TWO 30-MINUTE SESSIONS IN ONE DAY (4-6 HOURS APART)**
- **UP TO $30 IN WATCARD**
- **QUESTIONNAIRE, AUDIO AND VIDEO DATA WILL BE COLLECTED**
- **E5 (IN-PERSON)**

**Participate and receive up to $30!!!**

To participate please contact Priyank :

SIRRL Social and Intelligent Robotics Research Laboratory — UNIVERSITY OF WATERLOO

pavijeet@uwaterloo.ca

This study has been reviewed and received ethics clearance through the University of Waterloo Human Research Ethics Board

pavijeet@uwaterloo.ca

# Study 3 - Session 1 Task 1
# Instruction Sheet

## Interactions with a Voice System

## Session 1:

**Task 1:**

Hello! Please try to answer the questions about yourself that are asked below. **Robin**, our Voice System, will let you know when to begin and when to finish. On your screen, you'll see a color-coded indicator to help you manage your time: yellow for preparation, green for speaking, and red to wrap up your response. Please aim to **speak** for about **90 seconds**.

To assist you, brief hints related to the questions will also appear on the screen. You'll have **60 seconds** to **prepare**, after which Robin will prompt you to begin speaking.

Please try to answer the following questions:

What's your name, and where did you grow up? Could you tell me a bit about the family members or friends you spent time with as a child? I'd love to know about the schools you went to and any favorite places you enjoyed exploring. Were there any special dishes or snacks you loved that remind you of home? Take a moment to think about these memories, from the people who shaped your world to the places and flavors that made it special.

# Study 3 - Session 2 Task 1
# Instruction Sheet

## Interactions with a Voice System

# Session 2:

**Task 1:**

Hello! Please try to answer the questions about yourself that are asked below. **Robin**, our Voice System, will let you know when to begin and when to finish. On your screen, you'll see a color-coded indicator to help you manage your time: yellow for preparation, green for speaking, and red to wrap up your response. Please aim to **speak** for about **90 seconds**.

To assist you, brief hints related to the questions will also appear on the screen. You'll have **60 seconds** to **prepare**, after which Robin will prompt you to begin speaking.

Please try to answer the following questions:

What program are you enrolled in, and which term are you in? Could you tell me a bit about the courses you're taking this term and whether any of them are particularly interesting or challenging? Now, here's something a bit more playful—if you could step into the life of any celebrity, who would you choose to become? Take a moment to think about what draws you to them, from their personality or accomplishments to the values they represent or the life they lead.

# Study 3 - Debriefing Script

**RESEARCHER'S SCRIPT FOR VERBALLY DEBRIEFING PARTICIPANTS INVOLVED IN A PARTIAL DISCLOSURE STUDY**

**Study Title**: Interactions with a Voice System

Thank you for spending the time helping us with our research. We greatly appreciate your participation in this study.

Do you have any guesses about what the study is about? We would be interested in hearing any ideas you might have. Did anything seem strange or odd to you?

I would now like to tell you a bit more about the study. The debriefing letter that I gave you describes the details of the study, and you can keep this copy. I will go over the main points with you now.

When you began the study, you were told that we were interested in exploring human-robot interaction in general with our Voice System. We had also informed that audio and video data would be recorded. However, we left out a few specific details about the true focus of the study. Some studies involve partial disclosure, meaning that participants are not informed of the full purpose of the study at the start to avoid influencing their behaviour. This study followed that approach. Do you have any questions?

The actual purpose of this study was to examine whether participants subconsciously adjust their speech rate to align with the speech rate of the Voice System, when the system speaks at a slow pace. We also intended to study the changes in your speech pattern in the mildly stressful task that involved you selling one of the objects on the table to the Voice System. We are interested in if people naturally adapt their speech during interaction with a Voice System without consciously thinking about it. To do this, we needed to ensure that participants weren't aware we were specifically measuring speech rate, as knowing this could have influenced your behaviour. The tasks you completed were designed to observe how your speech naturally aligned with the robot's speech patterns.

So, the reason we used partial disclosure in this study was to ensure that your speech behaviour remained as natural as possible. If we had told you beforehand that we were specifically investigating speech rate alignment, you might have consciously monitored or altered your speech, which would have affected the accuracy of the data we were trying to collect.

To summarize:

1. The purpose of this study was to investigate whether people naturally adjust their speech rate to align with the speech rate of a Voice System during interaction. We also intend to compare the speech alignment in the two conditions.
2. You were assigned to either a robot or a laptop group, where the Voice System spoke at a predetermined slow pace.
3. Your speech data, including recordings, were collected to measure any adjustments in response to the robot's/laptop's speech rate.

We apologize for not fully explaining the true purpose of the study earlier. We hope that you understand why it was necessary to withhold that information to avoid influencing your natural responses. Do you have any questions about the partial disclosure and why it was used?

I'd also like to assure you that not all human-robot interaction studies involve partial disclosure like this one. In many cases, participants are fully informed about the study's objectives upfront.
If you have any further questions or concerns about the study or feel any discomfort about your participation, please feel free to contact the Student Investigator, me (Priyank Avijeet) or the Office of Research Ethics. The contact information is provided in the debriefing letter I gave you.

Because this study involved partial disclosure, it's important that you do not discuss the details of your experience with others who might participate in the study in the future. Sharing information could affect the results, as future participants might alter their behaviour if they know the study's true purpose in advance. The accuracy of the research depends on everyone coming into the study without prior knowledge.

Even though this study involved partial disclosure, the information we gave you about confidentiality, data storage, and security still applies. All data collected is confidential and securely stored. Only the researchers will have access to your data. These details are also outlined in the debriefing letter.

We really appreciate your participation, and we hope that this experience was interesting for you.

# Study 3 - Experiment Script

## Session 1: Theme – Food

### Task 1:

"Hello! I am Robin. It is a pleasure to meet you. Could you please respond to the questions you read on the sheet? You may start speaking now!"

(90 seconds for the response)

*What's your name, and where did you grow up? Could you tell me a bit about the family members or friends you spent time with as a child? I'd love to know about the schools you went to and any favorite places you enjoyed exploring. Were there any special dishes or snacks you loved that remind you of home? Take a moment to think about these memories, from the people who shaped your world to the places and flavors that made it special."*

*End of Speaking Phase*

"You may stop speaking now! Those were some interesting facts! Let us move on to the second task!

---

### Task 2:

"In this task, please repeat each sentence soon after I have finished speaking the sentence. I will not verbally instruct you to start and stop after each sentence! Instead, I will move on to the next sentence after 10 seconds. You are expected to repeat the sentence within these 10 seconds. The screen will display color-coded sections to guide you: Red means it's time to listen. Green signals when you should be speaking. There will be ten sentences in total. I will start with the first sentence now!"

---

*Speaking Phase (Sentence Repetition Begins)*

Sentences for Session 1:

1. Fresh apples are perfect for quick snacks.
2. Eggs help us build strong muscles.
3. Chicken curry tastes great with rice.
4. Bananas and milk make healthy breakfasts.
5. Tea helps us feel relaxed and calm.
6. Pasta with sauce is always comforting food.
7. Sweet oranges are packed with vitamin C.
8. A warm sandwich satisfies any quick hunger.
9. Ice cream melts fast on sunny days.
10. Crispy fries taste amazing with ketchup.

(Each sentence is followed by a 10-second gap for repetition.)

---

*End of Speaking Phase*

*"Great job! Now, we will move on to the third task!"*

Task 3:

"Well, this is going to be a fun activity! Pictures will appear on the screen, and we will take turns describing them. Each time, I will talk about one picture, and you will talk about the next one. I will not verbally instruct you to start and stop after each picture! Instead, I will move on to the next picture after 25 seconds. You are expected to say 3 sentences about the picture within these 25 seconds. The screen will display a color-coded section to help you know when to speak: Green signals when you should be speaking. Red means it's time to listen. There will be four pictures each for both of us in total. I will start with the first picture now!"Voice System:



A warm bowl of chicken soup sits on the table. Fresh herbs and a piece of bread accompany it. The soup appears both tasty and comforting.

Participant:



Voice System:



Friends are happily sharing a sushi platter. Soy sauce bowls are placed on the table. The cozy restaurant has a warm atmosphere.

Participant:



Voice System:



A person sits on a park bench. They hold an ice cream cone in hand. Tall trees and buildings surround the park.

4

Participant:



Voice System:



A person is relaxing in a cozy chair while reading a book. A table in front of them holds a cup of coffee and a plate with toast. The setting is calm and simple, perfect for relaxation.

5

Participant:



Task 4:

“Now, in this task, I am excited to share some of my favorite foods around the University of Waterloo. Some are easy to make at home, and others I like to get from nearby places.

Let’s start with breakfast. In the morning, I love making pancakes with a bit of butter and syrup on top. Sometimes, I change things up with toast or a bowl of cereal. And when I want something savory, I make an omelette filled with veggies and cheese. It’s tasty and keeps me full for a while!

Around campus, I enjoy trying out different places to eat. For pizza, my top pick is The Score Pizza near Waterloo Public Square. They have a thin, crispy, vegetarian pizza with goat cheese and fresh veggies. It’s so good! When I am craving shawarma, I go to Tuh-hee-nee's. I think they have the best shawarma in Waterloo. It is located at 250 Philip Street. Its large chicken shawarma wrap is on my list of favorites. It is served with delicious fire-roasted chicken cooked to perfection.

Next, for a quick bite and coffee, I go to Williams Fresh Cafe on campus. They have sandwiches, wraps, and plenty of drink options. It’s cozy and close by, making it perfect for studying or relaxing between classes. And for dessert, I love the milk cake at D-Spot! This sponge cake is soaked in three types of milk, making it super creamy and sweet. It’s a perfect treat!

These are some of my favorite places around the University of Waterloo. I hope you get a chance to try all of them!

For this task, now it is your turn to tell me about yourself in the context of food. You are expected to speak for around 90 seconds. I will guide you on when to start and stop speaking.

Additionally, the screen will display a color-coded section to help you know when to speak: Yellow indicates preparation time. Green signals when you should be speaking. Red means it's time to stop speaking

Let us start now. Please try to answer these questions:

What kind of food do you like eating? What are some of your favorite dishes? Do you prefer eating at home or dining out? Can you tell me about a restaurant or café you like? Is there a dish you would like to learn how to prepare?

Short hints related to these questions will also appear on the screen to assist you in recalling them. You have 60 seconds to prepare your answer. I will instruct you to start speaking at the end of 60 seconds. Your preparation time starts now!"

*End of Speaking Phase*

"You may stop speaking now! That was an excellent response! That was it for this session. See you in the next session for another fun conversation! Please inform Priyank that you have completed this session."

## Session 2: Theme – Music

### Task 1:

"Hello! Nice to see you again. Could you please respond to the questions you read on the sheet? You may start speaking now!"

(90 seconds for the response)

---

*End of Speaking Phase*

"You may stop speaking now! Those were some interesting insights! Let us move on to the second task!

*What program are you enrolled in, and which term are you in? Could you tell me a bit about the courses you're taking this term and whether any of them are particularly interesting or challenging? Now, here's something a bit more playful—if you could step into the life of any celebrity, who would you choose to become? Take a moment to think about what draws you to them, from their personality or accomplishments to the values they represent or the life they lead."*

### Task 2:

"In this task, please repeat each sentence soon after I have finished speaking the sentence. I will not verbally instruct you to start and stop after each sentence! Instead, I will move on to the next sentence after 10 seconds. You are expected to repeat the sentence within these 10 seconds. The screen will display color-coded sections to guide you: Red means it's time to listen. Green signals when you should be speaking. There will be ten sentences in total. I will start with the first sentence now!"

---

*Speaking Phase (Sentence Repetition Begins)*

Sentences for Session 2:

1. Soft melodies help clear your thoughts.
2. Guitar strings make beautiful, pleasant sounds.
3. Drums beat loudly at every concert.
4. Jazz music fills rooms with calm.
5. A piano plays notes of grace.
6. Singing together brings people great joy.
7. Classical music helps you feel peace.
8. Phones let you play songs anywhere.
9. Violin strings create peaceful, flowing tunes.

10. Music connects hearts across all boundaries.

(Each sentence is followed by a 10-second gap for repetition.)

---

*End of Speaking Phase*

"Great job! Now, we will move on to the third task!"

## Task 3:

"Well, this is going to be a fun activity! Pictures will appear on the screen, and we will take turns describing them. Each time, I will talk about one picture, and you will talk about the next one. I will not verbally instruct you to start and stop after each picture! Instead, I will move on to the next picture after 25 seconds. You are expected to say 3 sentences about the picture within these 25 seconds. The screen will display a color-coded section to help you know when to speak: Green signals when you should be speaking. Red means it's time to listen. There will be four pictures each for both of us in total. I will start with the first picture now!"

Voice System:



A boy sits on a bench, focused on playing his guitar. Tall buildings and green trees surround him. He seems deep in concentration, fully absorbed in creating music.

Participant:



Voice System:



A young man sits at a piano, focused on the keys. Warm light from a lamp fills the room. The soft lighting and quiet make the scene calm and thoughtful.

Participant:



Voice System:



A musician stands on stage playing the saxophone as the audience watches. Warm lights create a cozy feel in the dim jazz club. Other musicians play in the background, adding to the live music experience.

Participant:



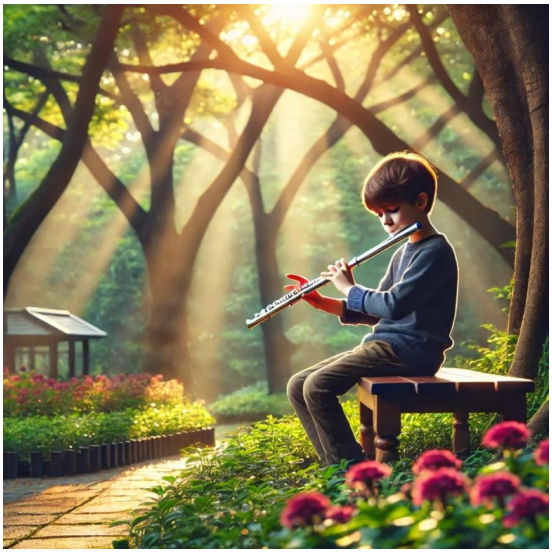
Voice System:



A pianist sits at a grand piano on a large stage. Bright lights shine down, creating a dramatic effect. The concert hall is silent as everyone watches and listens.

12

Participant:



Task 4:

“Now, in this task, I am excited to share some of my favorite music with you! Music comes in many styles, and exploring different genres can be a fun way to experience something new. Some music I like to listen to on my own, while other kinds are more enjoyable with a group of friends.

Let’s start with a classic! One genre I really enjoy is classic rock. Bands like The Beatles and The Rolling Stones helped shape modern music with their memorable songs and unique sounds. The Beatles, with songs like *Hey Jude*, are known for their harmonies and thoughtful messages. The Rolling Stones, on the other hand, bring a cheerful energy with hits like *Paint It Black* and *Satisfaction*.

When I'm in the mood for something more lively and upbeat, I switch to pop music. This genre is full of catchy and fun tunes that make you want to move! Taylor Swift and Dua Lipa are two of my favorites because of their energetic and engaging music. Taylor Swift tells stories with her songs, like *Shake It Off* and *Blank Space*, which are easy to sing along to. Dua Lipa, with songs like *Don’t Start Now* and *Levitating*, creates beats that are perfect for dancing.

Another genre I enjoy is indie and alternative rock. Bands like Arctic Monkeys and Tame Impala have a unique sound that feels relaxed yet interesting. Arctic Monkeys’ album *AM* has cool lyrics, while Tame Impala’s

*Currents* adds a modern touch to rock. Indie and alternative rock are great for chilling out or hanging with friends.

These are some of my favorite artists! I hope you get a chance to listen to some of these!

For this task, now it is your turn to talk about yourself in the context of music. You are expected to speak for around 90 seconds. I will guide you on when to start and stop speaking. Additionally, the screen will display a color-coded section to help you know when to speak: Yellow indicates preparation time. Green signals when you should be speaking. Red means it's time to stop speaking

Let us start now. Please try to answer these questions:What kind of music do you like listening to? Who are some of your favorite artists? What do you enjoy about their music? If you could see any artist perform live, who would it be? Is there a musical instrument or style you'd like to learn? Can you tell me about a few songs you like? What do you think they are about?

Short hints related to these questions will also appear on the screen to assist you in recalling them. You have 60 seconds to prepare your answer. I will instruct you to start speaking at the end of 60 seconds. Your preparation time starts now!"

*End of Speaking Phase*

"You may stop speaking now! That was an amazing response! Let us move on to the last task of this study!"

Task 5:

"For this task, your challenge is to prepare a sales pitch to persuade me to purchase one or more items placed on the table to your right. You will have: 60 seconds to prepare. 90 seconds to deliver your pitch. I will guide you on when to start and stop speaking. Additionally, the screen will display a color-coded section to help you know when to speak: Yellow indicates preparation time. Green signals when you should be speaking. Red means it's time to stop speaking Let us start now. You have 60 seconds to prepare your pitch. I will instruct you to start speaking at the end of the 60 seconds. Your preparation time starts now! You may start speaking now!"

(90 seconds for the sales pitch)

*End of Speaking Phase*

"You may stop speaking now! That was a convincing pitch! That was it for this study. Thank you for your time! Please inform Priyank that you have completed this session."

# Study 3 - Web Interface

UNIVERSITY OF WATERLOO

Please enter your Participant ID:

Powered by Qualtrics

UNIVERSITY OF WATERLOO

**Title of the study:** Interactions with a Voice System

**Research Team:**

**Principal Investigator:** Prof. Kerstin Dautenhahn
Departments of Electrical and Computer Engineering/Systems Design Engineering,
University of Waterloo, Canada
Phone: +1-519-888-4567 x31455
Email: kerstin.dautenhahn@uwaterloo.ca

**Student Investigator:** Priyank Avijeet
Department of Electrical and Computer Engineering,
University of Waterloo, Canada
Office: E7 5412
Email: pavijeet@uwaterloa.ca

To help you make an informed decision regarding your participation, this letter will explain what the study is about, the possible risks and benefits, and your rights as a research participant. If you do not understand something in the letter,

please ask one of the investigators prior to consenting to the study. You will be provided with a copy of the information and consent form if you choose to participate in the study.

**Research purposes and objectives:**

This study aims to explore how you interact with a Voice System application running on either a QT Robot or a laptop. You will be assigned to a group that interacts with one of these two devices. There will be two sessions, and each session will be based on a general theme. The verbal tasks in the sessions will mostly be around the theme of the session. Before and after each session, you'll be asked to answer a question about yourself or your preferences relating to the theme of the session. This study aims to gather natural speech samples; therefore, the accuracy of the content in your responses will not be assessed. During each session, you may be asked to repeat phrases spoken by the Voice System, describe images to the Voice System, or respond to questions. Additionally, you'll complete a brief questionnaire before and after each interaction with the Voice System. At the end of the study, we will conduct a short post-study questionnaire and a semi-structured interview to gain deeper insights into your experience.

This study will be a part of a master's thesis that will be submitted for the degree completion of the student investigator, Priyank

Avijeet.

**Procedure:**

This study includes two in-person sessions, each lasting approximately 30 minutes, conducted on the same day with a 4-6 hour interval between sessions. Prior to your first session, you will receive an Information Sheet, Consent Form, and a link to complete a demographic questionnaire.

Each session will follow a similar format. Upon arrival at your scheduled time, you'll begin with a brief questionnaire to assess your emotional state. Some basic questions about yourself written on a sheet will be provided to you at the start of the session. After you have read the question, you will then interact with the Voice System by entering your Participant ID on a screen to start the session. The Voice System will introduce itself and will allow you to prepare for 60 seconds before you respond verbally to the question on the sheet. You should try to keep your responses around 90 seconds long. You are encouraged to answer as many questions as you feel comfortable with, but it is not mandatory to respond to all.

In each task across both sessions, the Voice System will prompt you when to start and stop. Additionally, visual signals on the screen, like colour changes, will guide you. For instance, a green

section on the screen will indicate that you should begin and continue speaking, while a red section will signal you to stop speaking.

During each session, you will interact with the Voice System by repeating 10 short sentences one by one. The system will say a sentence and pause for 10 seconds before moving on to the next sentence, during which you should repeat the sentence. If, for some reason, you cannot repeat a sentence, then you should just say that you do not remember the sentence, and hence, you cannot repeat it. Following this, the Voice System will display a picture on the screen and describe it in three sentences. Next, a new picture will appear, and you should describe or discuss the picture in three sentences. You can say anything that you feel is related to the picture. The system will allow 25 seconds for each response. This process will repeat three more times for both you and the system, with each of you taking turns using different pictures each time.

In the final task, the system will speak on a topic for around four minutes, related to the session's theme. It will then ask you questions, to which you'll respond for 90 seconds. You will be allotted 60 seconds of preparation time. As mentioned, the system will indicate when to start and stop. You are encouraged to answer as many questions as you feel comfortable with, but it

is not mandatory to respond to all. Note the goal of this study is to collect natural speech samples, so the content accuracy of your responses will not be evaluated.

At the end of each session, you will complete a short questionnaire to reassess your emotional state and report your confidence level in responding to the questions provided on the sheet at the beginning of the session, as well as the questions asked by the system at the session's end.

This process will be repeated for both sessions, with the Voice System's questions and session themes changing in each session to maintain variety and reduce boredom. After the two sessions, the Student Investigator, Priyank Avijeet, will conduct a brief post-study interview and administer a final questionnaire with additional basic questions relevant to the study.

Please note that all sessions and the interview will be audio and video recorded for analysis purposes.

**Participation criteria:**

The study is accessible to the University of Waterloo PhD, graduate and undergraduate students. You do not need to have any special knowledge or background in robotics.

## I. Your responsibilities as a participant:

Participation in this study will require two in-person sessions, each lasting approximately 30 minutes. It will be conducted in the Social and Intelligent Robotic Research Laboratory, E5, fifth floor. During each session, you will interact with the Voice System, answer brief questionnaires, and provide verbal responses, which will be logged for analysis. You are not required to download any software or bring any additional equipment.

We ask that you commit to attending both sessions on the same day with a 4-6 hour interval between the sessions and avoid any distractions during your interactions with the Voice System. Your responses, including summaries and questionnaires, will provide valuable data to support our research. Your interactions with the voice system will be audio and video recorded, which will help us with further data analysis related to the study. Please note that while your verbal responses are important, we are not evaluating the accuracy of the content, so there's no need to worry about correctness.

## II. Your rights as a participant:

Your participation in this study is voluntary. You may decide to

leave the study at any time by communicating this to the researcher. You may decline to answer any question(s) you prefer not to answer. At the beginning of the study, you will be given a participant ID. Your participant ID will be removed from our records after 1 month of your participation. Within this period, you may ask for deleting your data by contacting us by email and sending us the code you were given. Please make sure that you will keep this code, as we would like to ensure that the data is de-identified, we will not keep any identifying information. If you have any questions or wish to withdraw from the study, please contact: pavijeet@uwaterloo.ca.

**Remuneration:**

To thank you for your participation, you will receive remuneration of up to $30, which will be credited to your WatCard. After completing the study, you will be asked to provide your name and student ID via a digital form. This information will be stored in a password-protected file and shared with the WatCard office to process your remuneration. The amount received is taxable. It is your responsibility to report this amount for income tax purposes. If you choose to withdraw your participation from the study prior to study completion, you will be remunerated $5 plus the per-session remuneration depending on the number of sessions you attended based on the following information:

| Session | Remuneration |
|---|---|
| Pre-study Questionnaires and post-study interview | $5 |
| Session 1 | $12.5 |
| Session 2 | $12.5 |
| **Expected Total Remuneration** | **up to $30** |

**Benefits of the study:**

Participation in this study may not provide any personal benefit to you. However, the study will benefit the academic community and the research team. We plan to publish the results of this study, and researchers in the field of social robotics can potentially benefit from that. We hope the outcomes of this experiment pave the way for novel applications of social robots.

**Confidentiality of identities and data:**

You will not be identified individually in any written reports of this research. The information you share will be kept confidential by

assigning an ID code so that individual names are not associated with the data. Therefore, the connection between your identity and data is confidential and will be kept in a separate document on the password-protected computers of the investigators. The file linking names and participant numbers will be kept for a month, and only the investigators named above will have access to it. You may withdraw your consent and ask us to delete your data within this period by contacting pavijeet@uwaterloo.ca. Video footage will not be shared outside the research team without your consent, and the collected data will be de-identified. The data captured will be securely stored on a secure password-protected lab server with access only for the investigators. The lab servers are secure and created by the University of Waterloo. Immediately after the session, the data files will be uploaded to that server and will be removed from the local computer. If you decide to withdraw before the end of the study session, we will delete all the files and data corresponding to you. The de-identified data may be shared publicly. Your identity will remain confidential even in that case. Participation in this study, including consent to data sharing, is entirely voluntary, and participants are free to choose their level of involvement. However, once the data is placed in a repository and if it is de-identified or anonymized, it may not be possible to withdraw the data due to the lack of identifiable information. The de-identified data will be stored for a minimum of 7 years on a password-protected server located in

the principal investigators' laboratory.

Please note that the session will be video, and audio recorded, and we may quote you anonymously in papers and publications with permission. Also, with your permission de-identified data collected in pre-study may be used by the investigators in future studies on related topics that have received clearance from a research ethics board. Your consent to future use can be withdrawn after 1 month of your participation.

The data collected in this study may be used for future research, particularly in the fields of speech and audio analysis. While the actual video or audio files will not be shared, anonymized data derived from labeling and processing participants' speech may be made available. If shared, this data might include non-identifying details such as age and gender for each sample. While the scope and specifics of these future research projects are not yet determined, there is a possibility that some research could be conducted internationally. This means that privacy protections could vary based on the location of the research institution using the data.

Finally, while third-party researchers may have access to the data, we currently do not foresee access by researchers who are not governed by TCPS2 (the Canadian standard for ethical

conduct in research involving humans). Additionally, there are no current plans to link the collected data with other data sources, such as public records.

## III. Questions, comments, or concerns Ethics clearance:

This study has been reviewed and received ethics clearance through the University of Waterloo Research Ethics Board (REB #46793). If you have questions for the Board, contact the Office of Research Ethics, toll-free at 1-833-643-2379 (Canada and USA), 1-519-888-4440, or reb@uwaterloo.ca.

## Contact Information:

If you have any questions regarding this study or would like additional information about your participation, please contact Priyank Avijeet by email at pavijeet@uwaterloo.ca.

Powered by Qualtrics

UNIVERSITY OF WATERLOO

**Title of the study:** Interactions with a Voice System

**Research Team:**

**Principal Investigator:** Prof. Kerstin Dautenhahn
Departments of Electrical and Computer Engineering/Systems Design Engineering,
University of Waterloo, Canada
Phone: +1-519-888-4567 x31455
Email: kerstin.dautenhahn@uwaterloo.ca

**Student Investigator:** Priyank Avijeet
Department of Electrical and Computer Engineering,
University of Waterloo, Canada
Office: E7 5412
Email: pavijeet@uwaterloa.ca

By providing your consent, you are not waiving your legal rights or releasing the investigator(s) or involved institution(s) from their

legal and professional responsibilities.

I have read the information presented in the information letter about a study led by Priyank Avijeet as part of his master's thesis supervised by Professor Kerstin Dautenhahn, Department of Electrical and Computer Engineering, University of Waterloo. I have had the opportunity to ask questions related to the study and have received satisfactory answers to my questions and any requests for additional details.

I was informed that participation in the study is voluntary and that I can withdraw this consent at any time during the study session and within one month of participation in the study by informing the researcher. I am aware that I may withdraw my consent for any of the above statements or withdraw my study participation during the study session and within one month of participation in the study without any penalty by advising the researcher.

Please note that the session will be recorded, and we may quote you anonymously in papers and publications.

This study has been reviewed and received ethics clearance through the University of Waterloo Research Ethics Board (REB #46793). If you have questions for the Board, contact the Office

of Research Ethics, toll-free at 1-833-643-2379 (Canada and USA), 1-519-888-4440, or reb@uwaterloo.ca. For all other questions contact Priyank Avijeet by email at pavijeet@uwaterloo.ca.

The session will be videotaped. The purpose of videotaping is for post-study data analysis and the participant's identity will not be revealed if used in publications or presentations.

Please fill in details and check boxes:

- [x] **I agree to participate in this study, and I confirm that I am a student at the University of Waterloo. [NOTE: Checking this box is required to take part in this study.]**
- [ ] I agree that anonymous quotations in any thesis or publication that comes from this research can be used.
- [ ] I also agree that any audio/video clips, digital images, or photographs in which I appear may be used in teaching, scientific presentations and/or publications with the understanding that I will not be identified by name and my face will be blurred.
- [ ] I also consent for de-identified data collected in this study to be used by the investigators in future studies on related topics that have received clearance from a research ethics board. My consent to future use can be withdrawn after 1 month of my participation.
- [ ] I also agree that pictures and videos taken during the experiment with my face blurred can be used on social media by the University of Waterloo research team.

Powered by Qualtrics

UNIVERSITY OF WATERLOO

Thank you for taking the time to review our Information sheet and Consent form. We have successfully recorded your consent. As the next step, please click the link below to provide us with some basic information about yourself.

Click here for the Demographic Survey

Powered by Qualtrics

UNIVERSITY OF WATERLOO

Please enter your Participant ID:

Powered by Qualtrics

UNIVERSITY OF WATERLOO

Before we can start, we need a few pieces of information about you (you can skip the question you are not comfortable answering):

Age

Gender

Powered by Qualtrics

UNIVERSITY OF WATERLOO

## I see myself as:

| | Disagree strongly | Disagree moderately | Disagree a little | Neither agree nor disagree | Agree a little | Agree moderately | Agree strongly |
|---|---|---|---|---|---|---|---|
| Extraverted, Enthusiastic | ○ | ○ | ○ | ○ | ○ | ○ | ○ |
| Critical, Quarrelsome | ○ | ○ | ○ | ○ | ○ | ○ | ○ |
| Dependable, Self-disciplined | ○ | ○ | ○ | ○ | ○ | ○ | ○ |
| Anxious, Easily upset | ○ | ○ | ○ | ○ | ○ | ○ | ○ |
| Open to new experiences, Complex | ○ | ○ | ○ | ○ | ○ | ○ | ○ |
| Reserved, Quiet | ○ | ○ | ○ | ○ | ○ | ○ | ○ |
| Sympathetic, Warm | ○ | ○ | ○ | ○ | ○ | ○ | ○ |
| Disorganized, Careless | ○ | ○ | ○ | ○ | ○ | ○ | ○ |
| Calm, Emotionally stable | ○ | ○ | ○ | ○ | ○ | ○ | ○ |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Conventional, Uncreative | ○ | ○ | ○ | ○ | ○ | ○ | ○ |

Powered by Qualtrics

UNIVERSITY OF WATERLOO

We thank you for your time spent taking this survey.
Your response has been recorded.

Powered by Qualtrics

UNIVERSITY OF WATERLOO

Please enter your Participant ID:

Please enter the session number:

- ◯ 1
- ◯ 2

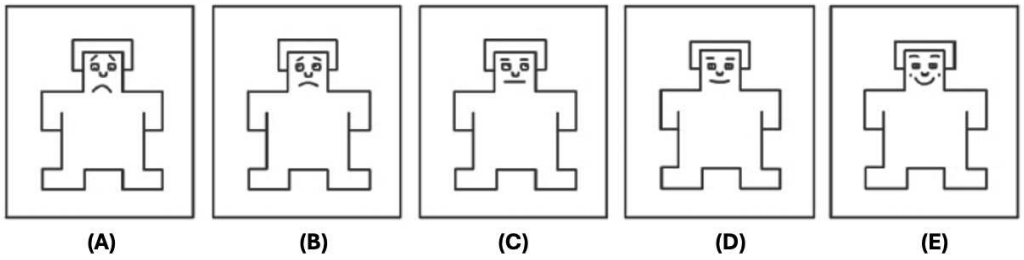
Is this pre-session or post-session survey

- ◯ Pre-session
- ◯ Post-session

Powered by Qualtrics

UNIVERSITY OF WATERLOO

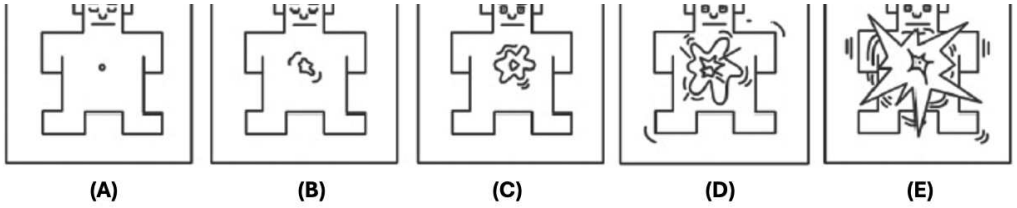
## How do you feel in general right now?




- ○ A
- ○ B
- ○ C
- ○ D
- ○ E

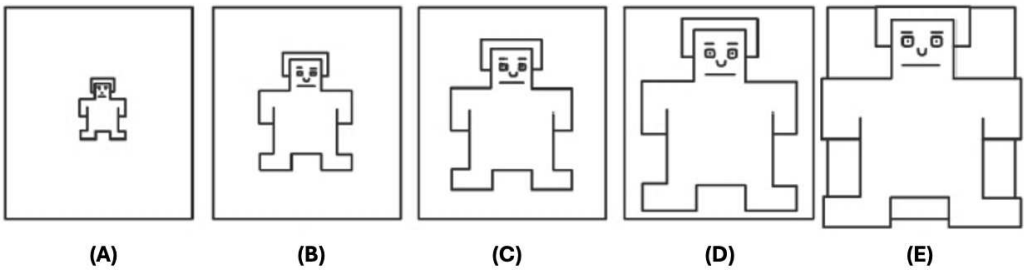
## How energized do you feel?






- ○ A
- ○ B
- ○ C
- ○ D
- ○ E

How in control do you feel of your current situation?




- ○ A
- ○ B
- ○ C

- ○ D
- ○ E

Powered by Qualtrics

UNIVERSITY OF WATERLOO

How confident were you in answering the questions about your personal background and early memories the start of this session?

Not at all confident Very Confident

How confident were you in answering  the questions about your food preferences and dining habit at the end of this session?

Not at all confident Very Confident

Powered by Qualtrics




How confident were you in answering the questions about your academic journey and celebrity inspiration at the start of this session?

Not at all confident Very Confident

How confident were you in answering the questions about your music tastes and aspirations at the end of this session?

Not at all confident Very Confident

How confident were you in the sales pitch preparation challenge at the end of this session?




Powered by Qualtrics

UNIVERSITY OF WATERLOO

We thank you for your time spent taking this survey.
Your response has been recorded.

Powered by Qualtrics

UNIVERSITY OF WATERLOO

Please enter your Participant ID:

Powered by Qualtrics

UNIVERSITY OF WATERLOO

As per your experience how fast or slow did the robot speak?

Very slow Very fast

Powered by Qualtrics

UNIVERSITY OF WATERLOO

Please rate your experience with the robot's speech based on its speech rate:

Dislike Like

Unfriendly Friendly

Unkind Kind

Unpleasant Pleasant

Awful Nice



Powered by Qualtrics

UNIVERSITY OF WATERLOO

How much did you feel as if you were interacting with an intelligent being?

Not at all Very much

How much attention did you pay to it?

Not at all Very much

How much did you feel involved in the interaction with it?

Not at all Very much



Powered by Qualtrics

UNIVERSITY OF WATERLOO

Please provide information about the languages you speak and your fluency levels:

Is English your first (native) language?

- (●) **Yes**
- ( ) No

How fluent are you in speaking English?

- (●) **Basic Knowledge: Basic words and phrases; simple, short conversations.**
- ( ) Intermediate: Converses on everyday topics; some hesitations.
- ( ) Advanced: Comfortable in complex conversations; minor limitations.
- ( ) Completely fluent: Near-native proficiency; fluency in any context.

How long have you been speaking English?

- ○ Less than 1 year
- ○ 1-2 years
- ○ 3-5 years
- ○ 6-10 years
- ◉ **more than 10 years**

Powered by Qualtrics

UNIVERSITY OF WATERLOO

Please answer the following questions based on feedback you have received from friends, family members, peers, coaches, or instructors about how you speak English:

Have you ever been told that when you speak English, you speak too fast?

- ○ Never
- ○ Rarely
- ○ Sometimes
- ○ Often
- ◉ **All the time**

Have you ever been told that when you speak English, you speak too slowly?

- ○ Never
- ○ Rarely
- ○ Sometimes

○ Often

◉ **All the time**

Powered by Qualtrics

UNIVERSITY OF WATERLOO

How often do you speak English in your daily life?

- ○ Never
- ○ Rarely
- ○ Sometimes
- ○ Often
- ◉ **All the time**

In which contexts do you frequently speak English? Check all that apply.

- ☑ **Formal (e.g., At work, In academic settings, In professional meetings)**
- ☑ **Informal (e.g., Socially with friends, At home, Casual conversations)**

How often do you think in English when speaking the language?

- ○ Never
- ○ Rarely
- ○ Sometimes

- ○ Often
- ◉ **All the time**

Do you currently have, or have you ever had, any speech disfluency or impairment, such as stuttering, cluttering, lisping, apraxia of speech, etc.?

- ○ Yes, I currently have speech disfluency
- ○ Yes, I had speech disfluency
- ○ No
- ◉ **Prefer not to answer**

Powered by Qualtrics

UNIVERSITY OF WATERLOO

How comfortable do you feel when speaking English?

Not comfortable at all — Extremely comfortable

Do you often feel nervous or anxious when speaking in English?

- ○ Never
- ○ Rarely
- ○ Sometimes
- ◉ **Often**
- ○ All the time

What is the main accent or dialect of English you have been exposed to?

Indian

Have you lived in an English-speaking country?

- (●) **Yes**
- ( ) No

Powered by Qualtrics

UNIVERSITY OF WATERLOO

How long have you lived in an English-speaking country?

- ○ Less than 1 year
- ◉ **1-2 years**
- ○ 3-5 years
- ○ more than 5 years

Powered by Qualtrics

UNIVERSITY OF WATERLOO

We thank you for your time spent taking this survey.
Your response has been recorded.

Powered by Qualtrics

UNIVERSITY OF WATERLOO

Please enter your Participant ID:

Please enter your Name (as on WatCard):

Please enter your Student ID (as on WatCard):

Powered by Qualtrics

UNIVERSITY OF WATERLOO

We thank you for your time spent taking this survey.
Your response has been recorded.

Powered by Qualtrics

# Appendix E

# Study 3 - Extra Plots and Figures

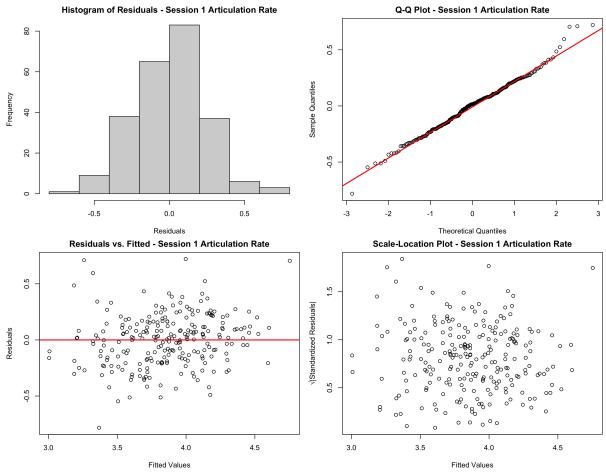



Figure E.1: Residuals Assumption Check for Articulation Rate in Session 1

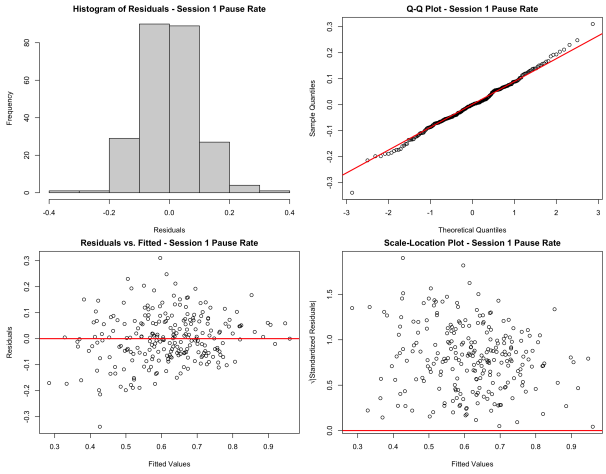



Figure E.2: Residuals Assumption Check for Pause Rate in Session 1

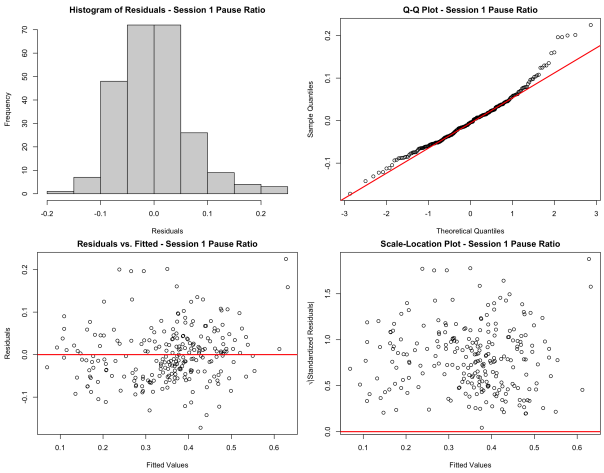



Figure E.3: Residuals Assumption Check for Pause Ratio in Session 1

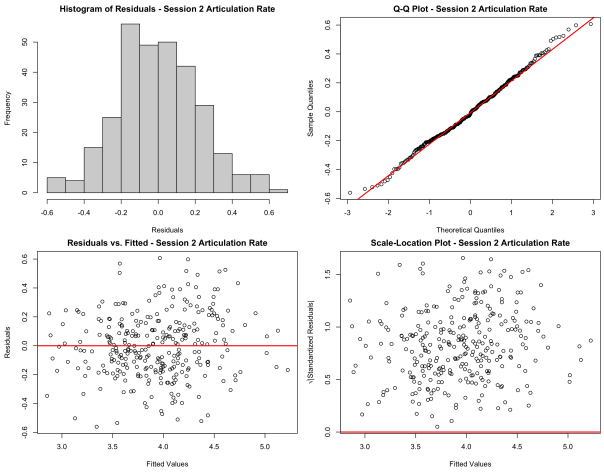



Figure E.4: Residuals Assumption Check for Articulation Rate in Session 2

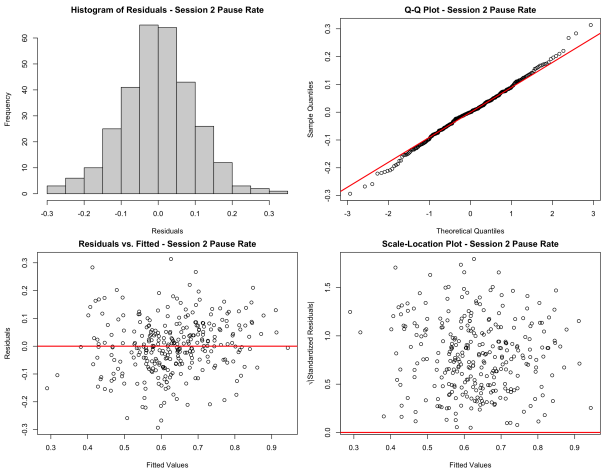



Figure E.5: Residuals Assumption Check for Pause Rate in Session 2

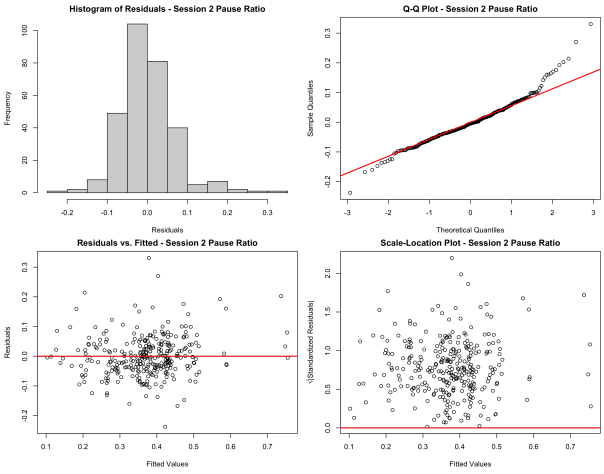



Figure E.6: Residuals Assumption Check for Pause Ratio in Session 2

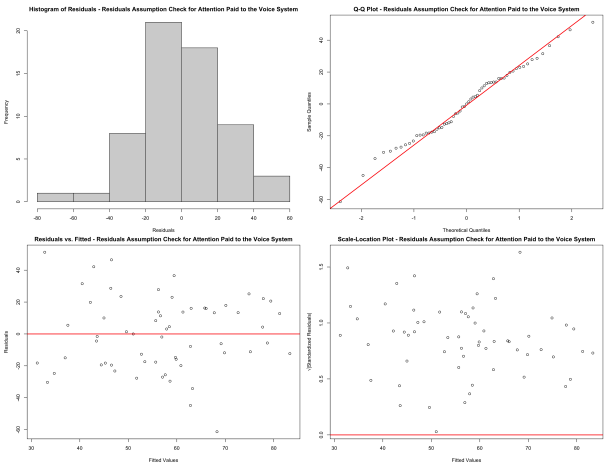



Figure E.7: Residuals Assumption Check for Attention Paid to the Voice System

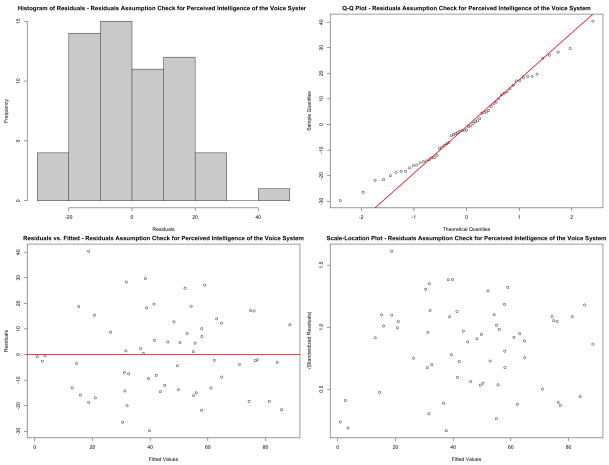



Figure E.8: Residuals Assumption Check for Perceived Intelligence of the Voice System

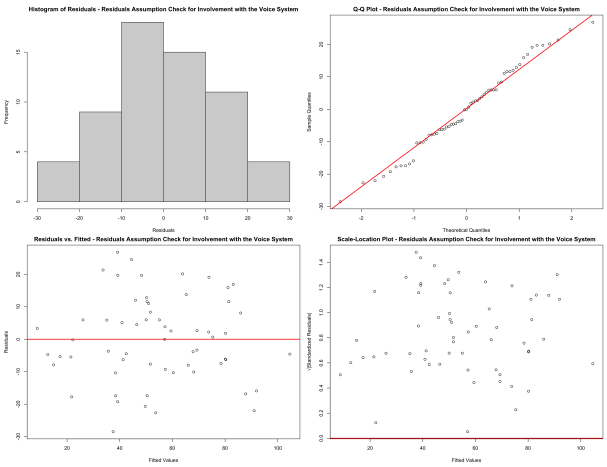



Figure E.9: Residuals Assumption Check for Involvement with the Voice System

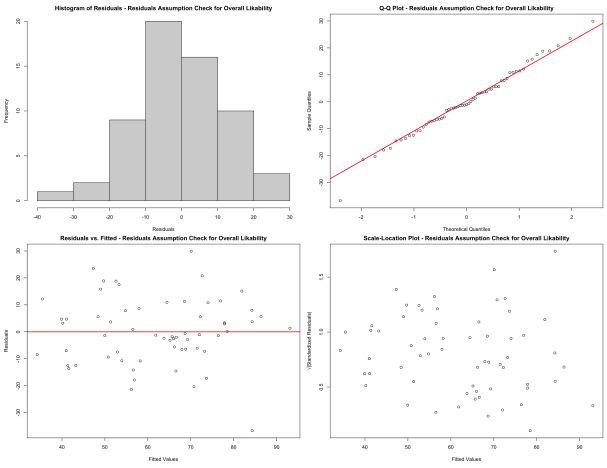



Figure E.10: Residuals Assumption Check for Overall Likability