

System Development and Evaluation of a Social Robot as a Public Speaking Rehearsal Coach

by

Delara Forghani

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, 2023

© Delara Forghani 2023

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 encompasses first-authored, peer-reviewed content that has been published in the conference proceedings of IEEE RO-MAN and is currently under review for a journal publication.

Research presented in Chapter 3 (Study 1): Delara Forghani, Prof. Moojan Ghafurian, Samira Rasouli, Prof. Chrystopher L. Nehaniv and Prof. Kerstin Dautenhahn designed the study. Delara Forghani implemented the system, performed the experiment, conducted the data analysis, and drafted the manuscripts. Prof. Moojan Ghafurian, Samira Rasouli, Prof. Chrystopher L. Nehaniv and Prof. Kerstin Dautenhahn reviewed and edited the manuscripts, and supervised the project.

Citations:

Delara Forghani, Moojan Ghafurian, Samira Rasouli, Chrystopher L. Nehaniv, and Kerstin Dautenhahn. 2023. What Do People Think of Social Robots and Voice Agents as Public Speaking Coaches? Accepted for conference presentation and publication in Proceedings of IEEE International Conference on Robot and Human Interactive Communication (RO-MAN 2023), 28-31 August 2023, Busan, South Korea.

Delara Forghani, Moojan Ghafurian, Samira Rasouli, Chrystopher L. Nehaniv, and Kerstin Dautenhahn. 2023. Evaluating People’s Perceptions of an Agent as a Public Speaking Coach. Under review for a journal publication.

Research presented in Chapter 3 (Study 2): Delara Forghani, Prof. Moojan Ghafurian, Samira Rasouli, Prof. Chrystopher L. Nehaniv and Prof. Kerstin Dautenhahn designed the study. Delara Forghani implemented the system, performed the experiment, conducted the data analysis, and is currently drafting a manuscript to be submitted to a journal. Prof. Moojan Ghafurian, Samira Rasouli, Prof. Chrystopher L. Nehaniv and Prof. Kerstin Dautenhahn supervised the project.

Abstract

Skills are developed through practice and rehearsal. Mastery and leadership often arise from dedicated effort. Among the skills that demand ongoing attention and practice, public speaking holds a crucial place. It is a skill by which individuals are assessed in various facets of life, be it in academia or as they progress in their professional careers. Nonetheless, when it comes to advancing in this skill, it is notably more time and energy-efficient to receive feedback from an individual capable of providing a reliable analysis of our performance. Ideally, engaging a human public speaking coach would be the optimal choice. However, accessing a human coach may not always be feasible due to a variety of reasons.

In this master’s thesis, I delved into the journey of exploring the potential of using a social robot as a coach for public speaking rehearsals, with the aim of enhancing individuals’ presentation skills.

My MASc project begins by conducting a comparison between two interactive technologies: a voice assistant agent and a social humanoid robot, both considered potential artificial systems for enhancing presentation skills. This comparison explores the embodiment of these technologies and underscores the significance of incorporating social non-verbal behaviours in an agent. To carry out this comparison, we organized an online study through Amazon Mechanical Turk, involving a diverse sample of participants across different age groups. Participants were asked to share their perspectives on various aspects of the agents depicted as coaches for public speaking rehearsals. They were presented with a scenario involving a video showing a student’s brief presentation rehearsal in front of both agents, followed by feedback provided by these agents. Our analysis encompassed both quantitative and qualitative methods. While participants generally displayed a preference for using socially interactive agents as coaches for public speaking rehearsals, the social robot outperformed the voice assistant agent on several metrics. Notably, it received higher rankings in terms of human nature attributes, perceived likeability, and perceived warmth.

Building upon the findings of the first study, and considering the more favourable evaluation of the social robot, we selected the social robot as the coaching agent for public speaking rehearsals involving university students in a single in-person presentation rehearsal session. The system development involved analyzing participants’ speech quality and (simulated) audience orientation and providing verbal feedback based on the calculated performance analysis.

In the second user study, we sought to gauge participants’ interactions with the robot, specifically their level of acceptance and their willingness to use the system in the future. We also administered questionnaires to assess participants’ perceptions of the robot’s behaviour during the session and the accuracy of its feedback. The results of this study

highlighted strong agreement among participants on the use of the social robot as a public speaking rehearsal coach in the university, and their expressed intention to utilize it in the future. Furthermore, participants' perceptions of the robot's competence and usefulness were closely linked to the accuracy of its feedback and the perceived quality of its listening behaviour. Finally, we asked a human public speaking coach to evaluate the robot coach, and presented the results. Future work can improve on this work and provide an accessible public speaking coach that university students can benefit from.

Acknowledgements

I would like to express my deep appreciation to Professor Kerstin Dautenhahn, my Supervisor, and Professor Chrystopher L. Nehaniv, my co-supervisor, for generously providing me with the invaluable opportunity to work within SIRRL. This experience has allowed me to gain profound insights into the world of Human-Robot Interaction (HRI). I extend my appreciation to my supervisors, who consistently shared their wisdom and knowledge with us during our time here. My heartfelt appreciation also goes to Prof. Moojan Ghafurian for her unwavering and remarkable support, and advice throughout my master’s journey and research endeavours. Lastly, I would like to convey my sincere thanks to Samira Rasouli for her assistance and support throughout my research efforts.

I wish to express my gratitude to the thesis review Committee, comprised of Prof. Brandon de Hart and Prof. Shi Cao, for dedicating their valuable time to assess my thesis and offer me their substantial feedback.

Special thanks to Dr. Kent Williams for generously sharing his expertise and providing invaluable advice in the evaluation and enhancement of the system I developed. Dr. Williams, your feedback has been truly precious, and I sincerely thank you for your support.

I want to express my appreciation to my lab members and friends who have been a source of advice and support throughout my master’s journey. Special thanks to Pourya Aliasghari, Ali Noormohammadi Asl, Ellie Sanoubari, Negin Azizi, Jule Margarete Körner, Dr. Melanie Jouaiti, Dr. Alex Aroyo, and Dr. Shruti Chandra.

This research was undertaken, in part, thanks to funding from the Canada 150 Research Chairs Program. I would also like to sincerely thank everyone who participated in the experiments.

Dedication

With heartfelt gratitude, this thesis is lovingly dedicated to my family, Zahra, Mohammad, and Delnaz, who were always by my side and provided me with their warm compassion during the most challenging days of my life.

Table of Contents

Author’s Declaration	ii
Statement of Contributions	iii
Abstract	iv
Acknowledgements	vi
Dedication	vii
List of Figures	xii
List of Tables	xiii
1 Introduction	1
1.1 Motivation	1
1.2 Research Goal and Scope	2
1.3 Summary of Contributions	3
1.4 Thesis Overview	4
2 Background Literature	5
2.1 Public speaking training	5
2.1.1 Public Speaking Anxiety	5

2.1.2	Training Systems for Public Speaking	7
2.1.3	Prosodic Features of Speech	9
2.1.4	Agent Comparison and Social Embodiment	11
2.1.5	Backchanneling	14
3	Study 1: Comparing a Voice Assistant Agent and a Social Robot as a Public Speaking Coach	17
3.1	Research Questions and Hypotheses	17
3.2	Methodology	18
3.2.1	Procedure and Measures	19
3.2.2	Experimental Conditions	24
3.2.3	Participants	25
3.2.4	Statistical Analysis	27
3.2.5	Results	27
3.3	Discussion	44
3.3.1	Perceptions of Agents as Public Speaking Coaches	44
3.3.2	Perceptions of the Human Nature and Social Attributes of Agents .	46
3.3.3	Perceived Intelligence and Likeability of Agents	47
3.3.4	Potential Gender Differences in Perception of Agents	47
3.3.5	Participants' Personality Traits and Differences in Perceptions . . .	48
4	Study 2: User Study Evaluating the Social Robot as a Public Speaking Coach	50
4.1	Research Questions and Hypotheses	50
4.2	Methodology	51
4.3	Experiment	53
4.3.1	System Implementation	53
4.3.2	Procedure and Measures	57
4.3.3	Pre-study Questionnaire	59

4.3.4	Post-study Questionnaires:	60
4.3.5	Participants	64
4.3.6	Results	65
4.4	Discussion	74
4.4.1	Acceptance and Intention to Use	74
4.4.2	Perceived Closeness	75
4.4.3	Perceived Human Nature	76
4.4.4	Evaluation of the Robot as a Public Speaking Coach	77
4.4.5	System Evaluation By a Human Coach	77
5	Conclusion, Limitations, and Future Work	79
5.1	Summary of Findings	79
5.2	Contribution to Knowledge	81
5.3	Limitations and Future Work	82
	References	84
	APPENDICES	98
A	Ethics Clearance Certificates	98
B	Study 1 - Pre and Post-Study Questionnaires	103
B.1	Participants' Demographics	104
B.2	Pre-experimental Questions	105
B.3	Evaluation of the Agent as a Public Speaking Coach	106
B.4	Evaluation of the Human Nature Attributes of the Agent	108
B.5	The Robotic Social Attributes Scale (RoSAS)	109
B.6	Godspeed Perceived Intelligence, and Godspeed Likeability	110
B.7	Participants' Overall Evaluations	111

C Study 2 - Pre and Post-Study Questionnaires	112
C.1 Participants' Demographics	113
C.2 PRCA-PS	114
C.3 TIPI	115
C.4 Almere	116
C.5 Inclusion of Other in the Self (IOS) Scale	119
C.6 Perceived Listening Behaviour and Perceived Usefulness	120
C.7 Open-ended Questions	121
C.8 Negative Attitudes Towards Robot Scale (NARS)	122

List of Figures

3.1	Study 1 -Experimental setup	20
3.2	Study 1 - Google Nest Mini lights	25
3.3	Study 1 - Histogram of participants' ages	26
3.4	Study 1 - Behaviour evaluation fo the agent as a coach	28
3.5	Study 1 - Human nature attributes of the agents	33
3.6	Study 1 - RoSAS scores in different conditions	34
3.7	Study 1 - Godspeed scores in different conditions	36
3.8	Study 1 - Participants' agreements scores on using the agent	38
4.1	Study 2 - Audience pictures and their arrangement as the contextual infor- mation	52
4.2	Study2 - Experimental Setup	58
4.3	Study 2 - The constructs interrelations of the Almere model	61
4.4	Study 2 - Expert evaluation flowchart	64
4.5	Study 2 - Participants' demographics	65
4.6	Study 2 - Almere model constructs scores	66
4.7	Study 2 - Human nature characteristics attributed to the Pepper robot . .	69
4.8	Study 2 - Perceptions of active listening on the robot	69
4.9	Study 2 - General perceptions of the robot as a public speaking rehearsal coach	71
4.10	Study 2 - Human coach's ratings on the robot's feedback	72

List of Tables

3.1	Study 1 - Participants' perceptions of the agents as coaches	21
3.2	Study 1 - Previous attendance in public speaking skill classes	27
3.3	Study 1 - Results of GLM predictions for attitude towards agents in different conditions	29
3.4	Study 1- Results of the GLM on the effects of gender and personality traits	30
3.5	Study 1 -Results of GLM predictions for perceived human nature attributes in various conditions	32
3.6	Study 1 - Results of GLM predictions on ROSAS dimensions in various conditions	35
3.7	Study 1 - Results of GLM predictions on Likeability scores	37
3.8	Study 1 - Results of the TA analysis on people's opinions about the agents	43
3.9	Study 1 - Results of TA analysis on participants' preferences	44
4.1	Study 2 - Speech quality thresholds	56
4.2	Study 2 - Questions on perceived active listening behaviour	62
4.3	Study 2 - Participant Inquiries on Robot Evaluation and Feedback Assessment	62
4.4	Study 2 - IOS scale responses distribution	68
4.5	Study 2 - Results of the quantitative analysis on human coach's feedback .	73

Chapter 1

Introduction

1.1 Motivation

Acquiring the art of public speaking is essential, as it finds utility in both academic [58, 128, 93] and professional contexts, especially in positions that involve interactions with clients [130, 79, 58]. As an example, a study focused on engineering students recognized the importance of integrating communication skills necessary for the workforce into engineering education to improve students' prospects of achieving success in their careers [93]. The significance of thorough preparation is often overlooked according to [47]. Nevertheless, the delivery of a presentation is greatly enhanced by repeated practice of the content and the manner it is delivered [124, 47]. Additionally, for a successful rehearsal leading to an effective speech, it is important to seek feedback. This feedback not only highlights one's strengths to maintain during the actual presentation, but also identifies weaknesses or areas that require improvement. Feedback can be obtained from either a coach or a tutor, but the difficulty in finding a human coach can often lead individuals to proceed without receiving professional feedback. This difficulty arises from factors such as the associated hiring expenses, difficulty in accessing a coach, time constraints, and personal barriers that hinder one's potential, like the fear of judgment and negative evaluations, often referred to as "public speaking anxiety." According to research, public speaking anxiety is a common manifestation of social anxiety [124, 12].

Interactive technologies, such as social robots and voice assistant agents, have the potential to complement existing human support systems. They can assume the role of a feedback system, evaluating presentations and offering individuals constructive feedback comparable to what a public speaking coach would provide. When discussing an interactive

agent, several design elements must be taken into account. First, in a broader context, it is essential to consider the agent’s embodiment and its non-verbal social behaviour (while giving feedback and while listening to the presenters). Subsequently, delving into our particular context, it becomes crucial to evaluate the accuracy of the agent’s feedback in its role as a coach and how instructive the rehearsal session was for the participants. As a final step, we should ask about participants’ acceptance rate and their intention to use the system again.

1.2 Research Goal and Scope

This thesis aims to explore humans’ evaluations and perceptions regarding social robots and voice assistant agents as public speaking coaches in two separate user studies.

In our first experimental study (described in Chapter 3) we conducted an online study with a focus on how the embodiment and non-verbal social behaviour of two different agents would be evaluated by participants. We explored three general Research Questions (RQs) in our first study:

- RQ1 How will participants evaluate two interactive agents with distinct embodiments, one featuring a minimalist design and the other a more complex embodiment capable of displaying various non-verbal gestures and body movements as public speaking coaches?
- RQ2 How can the presence of active listening behaviour in the social robot influence participants’ perceptions of the agents as public speaking coaches, and whether possessing this behaviour is useful in the given context?
- RQ3 How do participants’ personality traits affect their evaluations and ratings of a social agent as a coach?

In this research, we employed a social humanoid robot and a voice assistant agent, both delivering identical verbal content but distinct in their social non-verbal behaviour. Additionally, we included an extra layer of sociability and examined the effects of active listening behaviour by the social robot.

In our second experimental study, elaborated upon in Chapter 4, we extended the findings of the first study. In this phase, we introduced a social humanoid robot equipped with the ability to assess a presenter’s performance during a presentation rehearsal session.

We opted to proceed with the social robot due to its notably better ratings in Study 1 across multiple metrics. Study 2 focused on the system development and evaluation of the system, based on the perspectives of users and an expert. The Research Questions (RQs) we inquired about were:

RQ4 How participants would evaluate the robot’s feedback and its behaviour as a coach, and if they would like to use it in the future?

RQ5 How does a human public speaking coach assess the accuracy of the feedback provided by the robot as a coach?

In this study, participants delivered brief presentations in the experimental room while facing the robot. As they presented, the robot assessed various aspects, including voice modulation skills, speech prosody, acoustic features of their speech, and their ability to maintain eye contact during the presentation. Subsequently, the robot provided verbal feedback based on its analysis of participants’ presentations. We conducted an evaluation session with a human public speaking coach. After showing videos of some of the participants giving presentations and receiving feedback from the robot, we sought the coach’s opinions about the robot’s feedback.

1.3 Summary of Contributions

In this section, we discuss the contributions stemming from the studies presented in this thesis. The utilization of interactive embodied agents within the realm of public speaking training stands as a relatively new and evolving context. Our objective is to investigate not only the essential design strategies but also the system’s effectiveness in enhancing the overall experience in a public speaking rehearsal scenario.

1. The first study explores different physical and behavioural design factors that can contribute to the adoption and preference of a social interactive agent as a public speaking coach and compares the perceptions of different groups of users regarding three interactive agents.
2. In the second study, we introduced the application of a social humanoid robot equipped with a complete physical body and expressive capabilities. This robot was programmed to analyze participants’ presentations in real-time and offer constructive verbal feedback on their delivery. The system was introduced as a prospective

public speaking rehearsal coach, helping university students who prefer rehearsing with a non-human intelligent agent. The goal was to provide them with accurate, instructive, and impartial feedback to enhance their public speaking skills.

1.4 Thesis Overview

The remainder of the thesis is structured as follows:

Chapter 2 presents a comprehensive literature review, addressing pertinent topics such as public speaking training, prosodic features of speech, agent comparison and social embodiment, and backchanneling. Chapter 3, reports our first experimental study that investigates the effect of embodiment and social non-verbal behaviour on participants' perceptions of two agents as public speaking coaches. Chapter 4 describes our second study, focusing on a presentation rehearsal with a social robot and the evaluations provided by participants regarding the robot in its role. Finally, in Chapter 5, we draw the thesis to a close. This chapter provides concise responses to the research questions, outlines the contributions we have made to knowledge in the field, and discusses the limitations inherent in the studies presented.

Chapter 2

Background Literature

This chapter will delve into the literature that forms the background for our research.

2.1 Public speaking training

2.1.1 Public Speaking Anxiety

Public speaking anxiety (PSA) stands as the most prevalent type of anxiety and is the most frequently reported by the general public in the United States [12, 63]. Common negative thoughts linked to public speaking anxiety include the fear of making mistakes, worries about lacking experience, and concerns about facing rejection or humiliation from the audience [11]. Individuals with public speaking anxiety frequently exhibit speech disfluencies and tend to avoid making eye contact with the audience, which often contributes to a negative self-thought [6, 13]. In individuals with high levels of public speaking anxiety, this negative self-talk tends to be even more irrational. [130]. There were studies in the past on alleviating public speaking anxiety in presenters.

Trinh et al. (2015) [125] have developed a virtual agent that was co-presenting with the presenter. Based on the findings, the presence of a co-presenter had a notable impact, leading to a significant reduction in public speaking anxiety among participants and boosting confidence, particularly among non-native English speakers.

Kimani et al. (2019) [71], a virtual agent was created and utilized to engage participants in a Cognitive Behavioral Therapy (CBT) session aimed at guiding them through cognitive restructuring exercises during a presentation. This innovative approach was designed to

address and alleviate their public speaking anxiety. The study involved a comparison of two conditions in which the content of the agent’s conversation varied. In the intervention group (the CBT condition), the agent addressed the topic of public speaking anxiety, while in the control group, the agent engaged in a social chat about places to visit. The results indicated that participants in the CBT condition found the presentation practice more satisfying and enjoyable. Furthermore, the CBT approach led to a significant reduction in speech anxiety and lower self-reported nervousness.

In the study conducted by Wang et al. (2020) [130], the smart speaker Alexa was developed as an educational tutor to guide participants through cognitive reconstruction exercises aimed at reducing their public speaking anxiety. The researchers investigated the impact of varying levels of sociability in the tutor on participants’ perceptions. The results revealed that the sociability of the agents led to an increase in interpersonal closeness, subsequently reducing pre-speech anxiety. Additionally, it boosted participants’ satisfaction levels and their willingness to use the system in the future.

Acoustic characteristics and Public Speaking Anxiety

Goberman and colleagues (2011)) [46] conducted a study exploring the connection between a speaker’s acoustic characteristics and how they assessed their own public speaking anxiety, as well as how listeners evaluated this anxiety. The study demonstrated significant correlations between these variables. It revealed that speakers who perceived higher levels of anxiety in themselves exhibited longer pauses, longer breath groups, and a greater frequency of interjections in their speech. Additionally, there was a strong correlation between variations in fundamental frequency (F0) and how listeners perceived the speaker’s anxiety. The research highlighted a significant difference between self-ratings and listener-ratings of anxiety. Self-ratings tended to be higher, underscoring the “illusion of transparency,” where speakers believed their anxiety was more noticeable to the audience than it actually was. Furthermore, the study emphasized the beneficial impact of practice and rehearsal on speech fluency. Such preparation resulted in reduced speech disfluencies and a decrease in F0 variations.

On the other hand, Hagenaars et al. [53] found that the pitch variability of fearful speech is lower than that of happy speech. According to [88, 112] stress causes a greater muscular tension, which causes the voice to become rigid, thus reducing pitch variability.

Wörtwein et al. (2015) [133] conducted a study investigating the relationship between public speaking anxiety and various acoustic and visual behavioural indicators, aiming to develop an automated assessment of public speaking anxiety. They compared the anno-

tated values with self-assessment measurements using the Personal Report of Confidence as a Public Speaker (PRCS). The study’s results indicated that higher levels of anxiety during public speaking were associated with less eye contact with the audience, reduced voice variability, and an increased frequency of pauses in speech. Unlike [46] they found no correlation between the variation of fundamental frequency (f_0), and the scores on PRCS, but a lower variation of the first formant based on Mel-frequency Cepstral Coefficients (MFCC_0) which is closely related to the intensity of the voice signal was found to be correlated with higher PRCS scores. Also, according to the results anxious speakers tend to have a more monotone speech.

Additionally, Wörtwein et al. (2015) [133] examined how speakers perceived interaction with virtual audiences at varying levels of interaction (Non-interactive or interactive virtual audience), in which there was no significant difference between anxious versus non-anxious participants in their experiences.

2.1.2 Training Systems for Public Speaking

The utilization of interactive technologies for public speaking training is a relatively emerging field of study. Nonetheless, existing literature has highlighted the potential of intelligent user interfaces and interactive agents in assessing and providing feedback on human presentations.

The PitchPerfect system, integrated into Microsoft PowerPoint, was designed with a primary focus on enhancing content coverage and time management during presentations [126]. Beyond just improving the quality of presentations, PitchPerfect also provided assistance in terms of organizing learning content, effectively managing time, and boosting the confidence of presenters when delivering their presentations [126].

Hoque et al. (2013) [64] developed a conversational agent that practiced interviews with participants. It read facial expressions prosodies, and speech content and would provide participants with verbal and non-verbal behaviours in real-time. The system would provide feedback based on total pause duration, speaking rate in syllables per minute, percentage of filler words, pitch variation, the intensity of participants’ smiles, and loudness. The system provided both summary and focused visual feedback, using plots of the measurements. Summary feedback would show the trend of changes in each parameter during the overall interaction and across multiple sessions, and the focused feedback would show the changes in non-verbal behaviour at a time. According to the results, students who received automated affective feedback showed significant improvements in their mock interviews.

Also, participants confirmed that the practice with the embodied virtual coach was useful for them and wanted to practice with the system in the future.

Tanveer et al. (2015) [121] designed a real-time feedback system using a head-mounted display (Google Glass). The device tracked the speaker’s volume and speaking rate and provided participants with feedback using two different strategies, a continuous stream of feedback during the presentation, and sparse recommendations within regular time intervals during the presentation. The goal was to maximize the utility while minimizing potential distractions. The findings from the study indicated that participants expressed significantly higher satisfaction with the sparse feedback strategy. This suggests that delivering feedback at specific intervals during the presentation was more effective and well-received by the participants compared to a continuous stream of feedback.

Schneider et al. (2015) [114] designed a multimodal tool called the Presentation Trainer to give presenters real-time feedback about the non-verbal aspects of the presentation, such as body posture and gestures, voice volume, use of silence pauses as well as filled pauses, and ability to remain grounded without shifting weight from foot to foot, resulting in movements that resemble dancing movements. The tool presented a mirrored image of the presenter on a screen for the control group. Additionally, for the treatment group, the system provided a maximum of one feedback instruction at a time, using both visual and haptic feedback through a graphical interface and a wristband. The system selected which feedback to provide, considering the complexities associated with interpreting real-time feedback by participants.

The study aimed to evaluate participants’ experiences with the Presentation Trainer as they rehearsed for an elevator pitch, with the goal of demonstrating its significant impact on the learning process. The findings revealed that participants perceived practicing with the Presentation Trainer as a more effective learning method compared to traditional classroom settings. Conversely, participants who only saw their mirrored image tended to feel observed and became uncertain about the system’s data collection purpose. Importantly, the performance of the treatment group continued to improve over time, whereas the control group did not experience the same level of improvement.

Trinh et al. (2017) [124] developed a robotic coach using an anthropomorphic robotic head (Furhat). The robot takes the role of a presentation coach and analyzes the presenter’s five aspects of their presentation including content coverage, audience orientation, speaking rate, filler rate, and pitch variety. The robot would give high-level, but comprehensive and organized as well as compassionate feedback to the presenter with the aim of mitigating public speaking anxiety. They also compared the robotic coach with voice-only feedback and a visual-based feedback system using a screen. The results showed that it contributed

to the improvement of presentations compared to the verbal and visual-only feedback. Also, participants who rehearsed with the robotic coach reported high levels of satisfaction and desire to have future rehearsals with the coach. In addition, experts found that robot-assisted presentation rehearsals were significantly more engaging, and significantly better in terms of overall presentation quality and speech quality when compared to non-assisted rehearsals.

Saukh et al (2019) [111] created an app called Quantle that can measure and estimate basic vocal attributes in a person’s speech. The app is mobile-based and does not require an internet connection. Quantle focused on three vocal aspects: pace, pitch, and pause. The system was privacy-preserving as it did not share the speech analysis with any third party and ran locally on a smartphone. Additionally, it could function in noisy environments. Quantle was tested on approximately 1,000 TED speakers and eight speakers at a social event and proved its performance to be comparable to digital coaches that rely on third-party speech-to-text services.

Wang et al. (2020) [131] introduced a user interface for a voice coaching system. This interface focused on analyzing various voice modulation aspects, including pitch, volume, and speed, providing instant recommendations to enhance speech quality as users spoke. The recommendation module was driven by a learning algorithm that identified the most relevant training examples based on the system’s analysis of the user’s speech. The study’s results and performance comparisons were conveniently displayed on the user interface. To train the system, the researchers used a database containing high-quality presentations sourced from TED talks as a benchmark. The recommendation module operated akin to a search engine, retrieving TED talk examples by considering both the speech’s content and the voice modulation skills utilized by the user. This matching process relied on the cosine similarity of high-dimensional data within the system.

The study’s outcomes demonstrated that participants not only acquired knowledge about voice modulation skills but also found the system valuable for improving their skills. Additionally, they expressed satisfaction with their interactions with the system. However, a limitation noted was the system’s inability to offer personalized training.

2.1.3 Prosodic Features of Speech

According to Michelson et al., [83], oratory skills, which are introduced as facial expressions, motion, gestures and vocal features increase the success of a presentation.

Prosody is considered a crucial element in enhancing the effectiveness of a public speech. Numerous public speaking guides recommend that the speaker use a lively vocal delivery,

characterized by variations in intonation, rhythm, and volume, to engage the audience and achieve success [54]. Prior research has developed and examined the proposition that greater variability, specifically in the fundamental frequency (F0), is associated with a heightened perception of liveliness in spoken utterances [123, 59]. Another factor that appears to play a role in the effectiveness of public speaking is the speaker’s rate of speech. Research has demonstrated a stronger correlation between speaking rate and the perception of liveliness compared to pitch variation [123]. In various studies, a rapid speech rate, a lower vocal volume, and heightened speech intensity have been identified as traits associated with self-confidence [51].

Additional attributes thought to enhance the effectiveness of a presentation encompass the speaker’s capacity to connect with their audience, such as through maintaining eye contact, and their awareness of their non-verbal communication, including body language [29].

Hincks (2005) [59] evaluated how prosodic variables, such as pitch variability and speech rate, affected the perception of speaker liveliness. The results showed that pitch variability significantly affected categorizing a speech as being monotone or lively; this effect was stronger in male speakers than in females. The perception of liveliness was highly correlated with a normalized standard deviation of F0, and for female speakers, the mean length of runs was calculated with the number of syllables occurring between pauses of over 250 ms. According to the results, speech rate did not play a significant role in the perception of liveliness.

Over time, the definition of charisma has evolved, but there remains a certain quality that renders a speaker or leader charismatic, and it’s not solely dependent on visual attributes [9]. The charismatic influence of a speaker can be associated with various elements that can be categorized into two main areas 1) the content of the message being delivered and 2) the way through which the message is being delivered. People who have a strong way of delivery are often perceived as charismatic [61].

It was known that the former Apple CEO Steve Jobs was one the most charismatic speakers of the past decades, however, Mark Zuckerberg was generally considered a less charismatic speaker [119, 90]. Niebhur et al. (2016) [90] compared the acoustic features of the “tone of the voice” in Job’s two most well-known product presentations, versus Zuckerberg’s presentation at Facebook events around the same time. The results of the acoustic analyses showed that Steve Jobs had a very high F0 level and a larger F0 range in comparison to the reference values collected from the literature. Mark Zuckerberg had also a slightly high speech rate, that exceeded the range of the reference values and would have caused some phonetic reductions. The acoustic features of Zuckerberg were also higher than the reference values but not as prominent as Steve Jobs. The results generally support

that a charismatic speech is correlated with a high F0 level, a larger F0 range (equivalent to a larger pitch variety) and a relatively fast but not too fast speech rate [90].

Berger et al. (2017) [9] did phonetic research on charismatic speech, they specifically targeted some acoustic parameters, such as fundamental frequency (F0) level, F0 range, and speaking rate and their potential effects on the perceptions of a charismatic speaker. Unlike previous research, there was no relationship between the level and the range of F0 with the perception of charisma, and the only significant effect of F0 that was supported by this study was that the higher F0 leads to less perceived attractiveness. The results also show that the speech rate should not be too fast.

Rosenberg and Hirschberg studied the acoustic features of charismatic speech [108]. In this study, speech rate, intensity, mean f0, the standard deviation of f0, and maximum f0 were considered. These results indicated that all the indicators related to F0 were positively correlated with speakers' perceived charisma. Furthermore, speech intensity and speaking rate (syllables per second) were positively correlated with charisma perception. According to their lexical measures of one's speech, the longer a speaker spoke, the more words he or she used, the more charismatic he or she was perceived to be.

Deng et al. (2020) [30] studied the effects of two groups of speech disfluencies the filled pauses and word fragments in the task of automatic evaluation of fluent and disfluent speech. Based on the results the features based on disfluency are influential for evaluation tasks particularly when coupled with other prosodic features such as silence pauses and speaking rate.

Campione and Vronis (2002) [22] analyzed 5.5 hours of reading and spontaneous speech in five languages including English and categorized silence pauses as brief ($< 200ms$, only occurring in spontaneous speech), medium (200-1000 ms), and long ($> 1000ms$).

Finally, Lubold and Pon-barry (2014) [78] explored if acoustic and prosodic entertainment aspects can enhance rapport in collaborative learning dialogues. They considered three measures of entertainment including convergence, proximity, and synchrony and measured them using several acoustic and prosodic aspects of speech such as intensity, pitch variety, voice quality, and speech rate. They found that people entertain the most with the intensity of the voice. Acoustic-prosodic aspects such as pitch and voice quality appear to be the most significant entertainment factors when rapport is present.

2.1.4 Agent Comparison and Social Embodiment

Interactive agents are more effective when they are physically situated, possessing a tangible representation in their environment [20]. Examples of such physically embodied voice

assistant agents include smart speakers like Amazon Alexa and Google Echo. These devices have a non-anthropomorphic physical form and do not rely on social non-verbal cues in their interactions with humans. Instead, they use visual cues, such as the activation of LEDs, to signal their status to users [84].

Long-term studies have revealed an interesting trend where users, owing to the minimal embodiment and lack of social cues in their voice assistant agents, often take the initiative to decorate these agents in ways that enhance their social presence and responsiveness [25]. These voice interfaces are designed to mimic natural interactions, allowing users to concentrate their attention on a physical embodiment, much like they would in a typical conversation with a human [76].

Indeed, social embodiment differs from physical embodiment in that it encompasses more than just having a physical body. Social embodiment includes the possession of a wide array of non-verbal social signals and interaction cues, which can encompass facial expressions, body movements, gestures, smiles, eye gaze, and various other elements [84].

Common examples of socially embodied agents are social robots such as Pepper, Nao, and Cozmo which employ various social interaction cues to communicate with users. Even Amazon has made developments in robotics and introduced its recent social robot named Astro which is able to move and has huge eyes on a screen [44]. social robots utilize non-verbal behaviour to enhance their communication abilities, which adds an additional layer of pragmatic interaction.

Jung and Lee (2004) [70], showed that a socially embodied robot, functioning as an agent, enhanced the perceived social presence factor and resulted in more positive evaluations when compared to a non-embodied agent.

Luria et al. (2017) [80] designed a home control interface incorporating a social robot with expressive gesture feedback. A social robot named Vyo has been designed utilizing a microscope metaphor considering design factors such as engagement, automation, non-anthropomorphism, politeness, and reliability. The social robot was assessed alongside three other interfaces within the same application: a smart speaker, a mobile application, and a wall-mounted Android tablet device. The findings revealed that participants rated the more familiar interfaces with higher usability scores, while the voice control interface received the lowest score in this regard. The social robot achieved the highest situation awareness score, whereas the voice control interface scored the lowest. Notably, interfaces that ranked higher in terms of engagement and enjoyment tended to be perceived as less easy to use.

Moore et al. (2022) [84] assessed the distracting effects, perceived social presence, and perceived calm of voice user interfaces (VUIs) based on four cognitive tasks. This investi-

gation involved a comparison of three distinct types of voice user interfaces: a voice-only system (a hidden JBL Bluetooth speaker), a physical embodied system without anthropomorphic features (the Amazon Echo Dot smart speaker), and a social embodied system (the social Cozmo robot). Researchers assessed four measures for the cognitive tasks including cognitive speed, working memory, visual memory, and process inhibition. The voice-only VUI was found to have the least disruptive impact on cognitive speed, although it had an adverse effect on visual memory. The physical embodied VUI caused more distractions, followed by the social embodied VUI that caused the most distractions. Participants expressed the most likeability for the social embodied VUI, followed by the physical embodied VUI, with the voice-only VUI ranking as the least likeable. Interestingly, in comparison to the social embodied VUI, the physical embodied VUI was associated with a higher perceived calm. This observation suggests that participants might find greater comfort when designs are streamlined to include only essential features.

Kontogiorgos et al (2019) [73] studied how people responded to various interactive agents while engaging in guided cooking activities. The study employed three different devices: a smart speaker, a robotic head with a human-like face that exclusively communicated through speech, and the same robotic head that incorporated gaze and head movements into its interactions. The study’s findings suggested that the adoption of anthropomorphic characteristics by agents, along with the use of non-verbal cues in communication, may not always yield benefits.

Pollmann et al. (2020) [100] delved into the factors influencing the entertainment value of a robot, specifically in a game scenario. They sought to determine whether this entertainment value was solely associated with voice interactions or if it extended to the physical presence of the robot and its expressiveness conveyed through body gestures, arm movements, and total body motion profiles. They examined a social robot in two modes, inanimate and animated, and compared both to a condition where the robot was absent and only a touchscreen and voice input were available. The results indicated that while the pragmatic aspects of the agents were similar across conditions, the hedonic aspects of the interaction were notably lower in the screen-only condition compared to both the animated and inanimate robot conditions. Participants rated the animated robot as providing the highest entertainment value among the conditions [100].

Laban et al. (2021) conducted a study that emphasized the importance of the agent’s embodiment in a disclosure scenario involving a human, a social robot, and a voice assistant, demonstrating that the embodiment of the agent carries more weight than the content of the disclosure itself. [74]. Their research findings revealed that the embodiment of these agents played a critical role in shaping how individuals perceived the disclosures, the extent of information they shared, and the manner in which they conveyed these disclosures [74].

Notably, participants used significantly more words and devoted more time to disclosure when interacting with a human compared to their interactions with the social robot or voice assistant.

Consequently, when it comes to disclosure scenarios, people tend to favour a human embodiment due to its familiarity and its ability to facilitate more extensive disclosures [74]. Although individuals recognized distinctions between the two artificial agents (the social robot and the voice assistant), these differences did not affect the amount of disclosed information, its perceived quantity as well or its perceived quality [74].

Urakami and Sutthithatip (2021) [127] investigated the effects of the non-verbal and verbal behaviour of a robot in establishing relationships and bonding with a human partner. In their experiment, participants interacted with two robots that differed in the level of expressivity both verbally and non-verbally. The method was that the interactive robot engaged participants in conversations before, during and after the collaborative task, while the non-interactive robot remained passive. According to the results, the robot’s non-verbal and verbal behaviour increased participants’ perceptions of its social agency, and social presence, as well as strengthened the social bonding between the robot and the participants. The results also showed that the perceived collaboration enhanced when the robot maintained eye contact, the robot was given a personality, and the robot appeared alive.

2.1.5 Backchanneling

Backchannels serve as brief cues, both verbal and non-verbal, employed by a listener to express attentiveness and engagement with the speaker’s discourse without interrupting the speaker [134, 39]. They do not aim to answer questions, disrupt the speaker, or provoke specific reactions [39]. Instead, their typical purposes encompass displaying agreement [41], involvement [118], surprise [42], and comprehension [1].

Dennis and Kinney (1998) [31] noted that backchanneling can fulfill four cognitive roles, which encompass indicating comprehension, expressing uncertainty, offering clarification for a message, and concluding a sentence. Nonverbal backchannels are primarily defined by the use of body language, including actions such as nods, shoulder shrugs [33], smiles [21], and gaze [48]. Research conducted by Cathcart et al. (2003) [24] has observed that backchannels typically occur after a brief pause in the speaker’s speech.

Gratch et al. (2007) [52] developed listening behaviour for a virtual agent by analyzing inputs from a speaker, including factors like their head movements, gaze directions, and

acoustic characteristics. Their research revealed that the virtual agent, which offered affirmative listening cues such as head nods, could generate greater engagement and cultivate a stronger rapport compared to a human listener.

Sidner et al. (2005) [115], defined engagement as the dynamic process in which individuals participating in an interaction begin, sustain, and conclude their perceived connection with one another. Their research demonstrated that incorporating backchanneling behaviour in a robot enhanced participants' perceived engagement of the robot during a collaborative task with the robot [115].

Jung et al. (2013) [69] investigated the impact of backchanneling and task complexity on human-robot teamwork, specifically in the context of an Urban Search and Rescue (USAR) task. The study utilized two upper-torso humanoid robots, namely Nexi and Maddox. In the backchanneling condition, the robot would turn towards the participant and nod in sync with the participant's speech in a smooth and swift manner.

The results, obtained through a study involving 96 participants using a between-participants design, revealed that when backchanneling was employed during the complex task, it enhanced team performance and heightened the perception of engagement with the robot. However, it also resulted in a decrease in the perceived competence of the robot [69].

Head nods, as a form of non-verbal backchanneling, are often used to convey both attention and agreement. However, it's important to note that their interpretation can vary based on specific features such as their amplitude, frequency, and pace [17, 94].

Bousmalis [17] conducted a survey highlighting that head nods are considered nearly universal signals of agreement. Nonetheless, their meanings and functions can differ depending on factors like the nod's amplitude, number of cycles, duration, and the context in which they occur [17]. For instance, nods with greater amplitude and higher frequency often signify affirmation, while smaller nods may indicate active participation in the ongoing conversation [3, 109].

Oertel and colleagues [94] found that head nods, in general, conveyed a higher degree of attentiveness compared to audio backchannels. Through the analysis of a corpus of recorded human-human interactions, they identified a total of 77 distinct head nod patterns, taking into account characteristics such as nod duration, the number of oscillations, frequency, amplitude, maximum speed of upward/downward oscillations, and more. Additionally, they observed that various forms of nodding could be perceived as more or less attentive, and combining multiple forms enhanced the perception of attentiveness [94].

Andriella et al. (2020) [2] found that the combination of verbal and nonverbal backchannel cues in entertainment scenarios was legible by participants with analytical thinking and enabled them to make fewer mistakes.

More recently, models designed for generating backchannels have been assessed for their ability to anticipate the timing of backchannel responses, with the aim of improving the natural flow of conversations with an agent [2, 32, 95].

Park et al. (2017) [95] conducted a study in which they examined a rule-based backchanneling model in a storytelling context involving children aged 4-6 years old. Their model called the Backchannel Opportunity Prediction (BOP) model, identified four specific cues from the speaker and generated backchannels in accordance. When the BOP model was used in the robot in interactions with children, it led to a noticeable increase in the perception of attentiveness by the children. Additionally, the children exhibited a preference for telling stories to the BOP model robot [95].

Murali et al. (2021), as detailed in their study [87], created a virtual agent to assist students during oral presentations. This virtual agent expressed attentive listening through various non-verbal cues, including head nods, smiles, and raised eyebrows when the presenter made eye contact with it. Additionally, the virtual agent used an “OK” gesture whenever the presenter transitioned to the next slide. Their research involved both self-reported data and physiological measurements, including heart rate (HR) and heart rate variability (HRV). The results demonstrated significant reductions in public speaking anxiety among participants when compared to a control condition that only a text display [87].

Based on the literature discussed it remains uncertain to what degree of complexity in the embodiment of a socially interactive agent is preferred and liked by humans in the context of public speaking. Given the widespread use of smart speakers as voice assistant agents in various settings, we were interested in comparing one with a humanoid social robot capable of resembling non-verbal social behaviour. The aim was to assess their effectiveness as coaches for public speaking rehearsals, both in their social and professional aspects. While non-verbal backchanneling as a means of active listening has been explored in counselling, its examination in coaching remains limited. We aim to address this gap by investigating the effects of non-verbal backchanneling when implemented in a social robot functioning as a coach.

With a focus on the social robot equipped with a complete physical body and a broad range of expressiveness (not just an anthropomorphic head), we sought to explore university students’ perspectives on the social robot considering different metrics during a presentation rehearsal with it. The existing literature also lacks insights from experts within the field regarding their opinions and evaluations of feedback accuracy and potential enhancements. Therefore, our objective is to incorporate an expert evaluation of the developed system using a human coach in public speaking training.

Chapter 3

Study 1: Comparing a Voice Assistant Agent and a Social Robot as a Public Speaking Coach

3.1 Research Questions and Hypotheses

This study aims to answer the following research questions (RQs). In this chapter, we present detailed versions of the first three general questions we discussed in Chapter 1.

RQ1 How do participants evaluate a social robot and a voice assistant in terms of

RQ1-1 agent behaviour as a public speaking coach

RQ1-2 perceived human nature

RQ1-3 perceived social attributes

RQ1-4 perceived intelligence

RQ1-5 perceived likeability

RQ2 How does non-verbal backchanneling, expressed through head nods in the robot while listening to the presentation, affect participants' evaluations of

RQ2-1 robot behaviour as a public speaking coach

RQ2-2 perceived human nature

RQ2-3 perceived social attributes

RQ2-4 perceived intelligence

RQ2-5 perceived likeability

RQ3 How do participants' personality traits affect their evaluations and ratings of a social agent as a coach?

Based on the research questions mentioned previously, we proceed to examine the following hypotheses:

H1 Participants will evaluate the social robot more positively than the voice assistant agent in terms of its

H1-1 behavioural measures as a public speaking coach

H1-2 human nature aspects

H1-3 social attributes

H1-4 perceived intelligence

H1-5 perceived likeability

H2 Participants perceive the active listener robot that uses non-verbal backchanneling more positively than the passive listener robot in terms of its

H2-1 behavioural measures

H2-2 human nature aspects

H2-3 social attributes

H2-4 perceived intelligence

H2-5 perceived likeability

H3) We hypothesize that the personality traits have correlations with how people rate the agent.

The hypotheses are derived from various sources of prior research. **H1** draws on previous studies concerning an agent’s human-likeness, as evidenced in references such as [73, 80, 124]. **H2** is based on research exploring the effect of backchanneling and its positive impacts on users’ overall interaction experiences with agents, as exemplified by references like [52, 115, 87]. Lastly, **H3** is built upon earlier research concerning interpersonal similarity and attraction, highlighting significant correlations between the Big-Five personality traits and the Godspeed questionnaire, as documented in [27].

3.2 Methodology

We conducted an online video study through Amazon Mechanical Turk (MTurk) using a custom online interface developed with HTML, JavaScript, SQL, and the Go programming languages. This online web interface had been previously employed and tested in our laboratory’s online studies but was modified to suit the specific requirements and content of our study. To address our research questions, we compared three different experimental conditions, each involving distinct agents acting as public speaking coaches.

(a) **VoiceAgent** condition, in which we utilized the Google Nest Mini, which provided verbal feedback following the presentation [49].

- (b) **ActiveListenR** condition, in which we employed the Pepper robot [116], which displayed non-verbal backchanneling cues during the presentation and delivered verbal feedback accompanied by hand gestures after the presentation, during the feedback phase.
- (c) **PassiveListenR** condition, in which the Pepper robot remained immobile during the presentation, and it delivered verbal feedback accompanied by gestures in the post-presentation feedback phase.

The study adopted a between-participant design, where participants were randomly assigned to one of the three aforementioned conditions.

3.2.1 Procedure and Measures

Upon reading and providing their consent to the study information form, participants proceeded by following these steps:

Step 1 - Demographics and Pre-experimental Questionnaires: Participants were first requested to provide information regarding their age and gender. Furthermore, we inquired about their prior involvement in public speaking classes and workshops, as well as their self-perceived necessity to attend such events. Additionally, participants were asked to report their Big Five personality traits according to the TIPI framework [50]. It’s important to note that responding to all demographic inquiries was voluntary for participants, as mandated by our Human Research Ethics Board. For more information please refer to Appendix B.2.

Step 2 - Watching the Video: Participants were provided with instructions for viewing a video and subsequently answering questions related to it. The video featured an individual delivering a presentation on the topic of the “History of Canada Day.” The content for this presentation was sourced from the official website of the Government of Canada (canada.ca). The presenter’s speech lasted approximately one minute and thirty seconds. It’s worth noting that the agent displayed different behaviours during this presentation period in accordance with the specific experimental condition, which will be elaborated upon in Section 3.2.2.

Throughout the presentation, slides relevant to the topic appeared in the upper left corner of the video and were changed as the speaker progressed in their speech, as illustrated in Figure 3.1. Following the conclusion of the student’s presentation, the agent proceeded to provide feedback to the student, which lasted for approximately two minutes.

The agent, in its role as the coach, initiated the feedback session by expressing appreciation for the presentation and emphasizing its interest in the content. The feedback

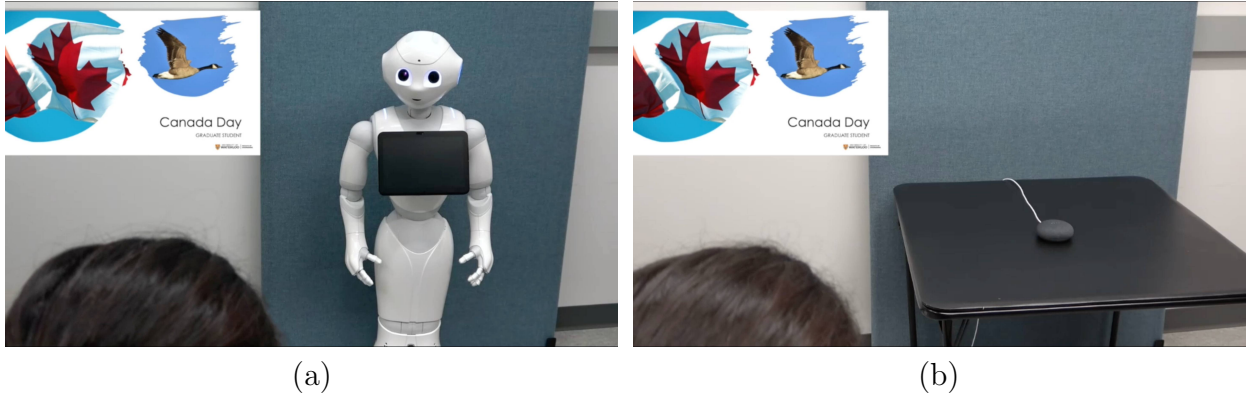


Figure 3.1: Picture of the experimental setup with (a) the social robot, and (b) the voice assistant agent.

primarily centred around the evaluation of vocal modulation skills and acoustic parameters during the presentation. This encompassed aspects such as speech rate, volume, and pitch variety, as well as the usage of rising intonations within sentences and falling intonations at the end of sentences.

During the feedback session, the agent highlighted positive aspects of the presentation with comments such as, “*You demonstrated a good speech rate in your presentation!*” or “*I liked the pitch variety in your speech!*” Following these positive remarks, the agent provided constructive suggestions for enhancing the speech, including advice like, “*Please avoid using fillers! Try not to use pauses too often or for too long!*”, or “*Try not to speak too fast or too slow!*” Finally, the agent appreciated the presenter’s efforts, creating a supportive and encouraging atmosphere.

After the agent’s feedback had been delivered, the video presentation concluded. At this point, participants were given the option to either replay the video or continue on to answer the subsequent questions. To prevent participants from doing multiple tasks at once while the video was playing, if the current window or tab was minimized or changed, the video playback would freeze, and it would automatically resume playing once the window or tab was refocused. This approach aimed to maintain participants’ focus and attention on the video content.

The audio recordings of both the presenter and the agent, as well as the content of the presentation and the feedback, remained consistent across all scenarios. These recordings were prepared in advance and played during the video recording. To replicate an adult voice, we adjusted the pitch and frequency of the Pepper robot’s voice, as it was originally

Table 3.1: Questions designed to assess participants’ perceptions of the agent’s behaviour as a public speaking coach. Answers were given on a continuous scale. The wording used at both ends of the scales is shown in the table.

Num	Question items	Scale labels
1	I found the agent as a public speaking coach:	(Not satisfying at all - Very satisfying)
2	I found the interaction with the agent as a public speaking coach:	(Not engaging at all - Very engaging)
3	I found the behaviour of the agent as a public speaking coach:	(Not natural at all - Very natural)
4	I found the agent as a public speaking coach:	(Not attentive at all - Very attentive)
5	I found the agent as a public speaking coach:	(Not competent at all - Very competent)
6	I found the agent as a public speaking coach:	(Not humane at all - Very humane)
7	How would you rate the coach’s feedback?	(Not trustworthy at all, Very trustworthy)
8	How well did you find the feedback provided by the coach?	(Very difficult to understand, Very easy to understand)
9	How helpful did you find this coach for students to improve their presentation skills?	(Very unhelpful, Very helpful)
10	How likely are you to rehearse with this coach for future presentations?	(Not at all, Very much)
11	How well do you think the coach can build rapport with a human partner?	(Not at all, Very much)

designed to sound childlike.

To prevent potential biases and maintain anonymity, the video camera, as shown in Figure 3.1, was consistently placed in the same position behind the presenter’s head for all experimental conditions. This ensured that the identity and age of the presenter remained undisclosed, to avoid any possible biases from occurring.

Step 3 - Post-experimental Questionnaires: To analyze participants’ perceptions of the agents in their roles as coaches, we employed a series of questionnaires, some of which were taken from established questionnaires commonly utilized in Human-Robot Interaction (HRI) research. Below, we provide a comprehensive explanation of each questionnaire employed in the study.

Agent Behaviour: This questionnaire was designed to focus on participants’ perspectives regarding the behaviour of the agents in their coaching roles and how they interacted with the students. The questionnaire comprised 11 questions, which are detailed in Table 3.1. Participants were asked to provide their responses on a continuous scale, ranging from 0 to 1000. For further information please refer to Appendix B.3.

Human Nature (HN) Attributes: Haslam et al. [56] have proposed two separate perceptions of humanness: Human Nature (HN) and Human Uniqueness (HU), each of which is characterized by specific attributes, and the absence of these attributes can signify dehumanization or denial of humanness. Human Uniqueness attributes distinguish humans from non-human animals and encompass qualities that make us humans such as civility, refinement, moral sensibility, higher cognition, and maturity.

In contrast, Human Nature attributes, which include emotionality, warmth, openness, agency (liveliness), and depth, are fundamental to humans and do not differentiate them

from non-human animals. When something is lacking in the realm of Human Nature, it tends to make an entity appear more machine-like [56].

These attributes have previously been employed to assess individuals’ perceptions of the HN and HU attributes of virtual agents [43]. In our study, we utilized the same questions to assess the HN dimensions of the agents as coaches. We asked participants to rate the agents on four dimensions of HN: emotionality, warmth, openness, and agency (liveliness). As with the previous questionnaire, participants provided their responses on a continuous scale, resulting in values ranging from 0 to 1000. For more information please refer to Appendix B.4.

The Robotic Social Attributes Scale (RoSAS): RoSAS, which stands for Robot Social Attribute Scale, is a standardized metric in the field of Human-Robot Interaction (HRI). It was developed by drawing inspiration from the Godspeed questionnaires [8] and leveraging insights from social psychological research on social perceptions of robots [23]. RoSAS considers three key social dimensions using 18 attributes: warmth, which is concluded based on measuring six attributes, including organic, sociable, emotional, compassionate, happy, and feeling (in scoring HN, as mentioned above, warmth is rated directly on a predetermined scale), competence, and discomfort. These dimensions are measured through a total of 18 attributes. For more information please refer to Appendix B.5.

One notable advantage of RoSAS is its versatility, as it is designed to be applicable to a wide range of robots. It’s important to note that RoSAS is not intended to replace previous HRI measurements like Godspeed [23], instead, it complements them. RoSAS does not mandate a specific scale for responses, but the authors do recommend including a neutral or midpoint value, typically by using an odd number of possible responses in Likert-type items. In our study, we opted to employ a seven-point Likert scale. This choice aligns with the approach taken in other studies [117, 97], where RoSAS was utilized to assess participants’ perceptions of a robot’s social behaviour, particularly in relation to warmth, competence, and discomfort.

Perceived Intelligence and Likeability (Godspeed): The Godspeed questionnaire is used to assess various aspects related to human-robot interaction, including anthropomorphism, likeability, animacy, perceived intelligence, and perceived safety [8]. It can be used alongside RoSAS when specific attributes need to be measured that are not fully covered by RoSAS [23, 8].

In our study, we focused solely on evaluating the perceived intelligence and likeability of the agents using the Godspeed questionnaire. We excluded the assessment of safety, anthropomorphism, and animacy aspects from our research because they were not directly relevant to our study’s research questions. It’s worth noting that in one of our experimental

conditions, the agent was a voice assistant that lacked physical goal-oriented movement and a human-like physical body. Additionally, previous research on animacy perception has indicated that it is more influenced by physical interactions with robots rather than by observing or watching human-robot interaction scenarios [40]. As a result, we concentrated exclusively on the subscales for perceived intelligence and likeability when utilizing the Godspeed questionnaire. For reading the question items please refer to Appendix B.6.

Open-ended Questions: Finally, participants were asked to offer their opinions through two open-ended questions concerning the potential utilization of robots and voice assistant agents as coaches for public speaking. In the first open-ended question, depending on their experimental condition (ActiveListenR, PassiveListenR, or VoiceAgent), they were directed to share their opinions on either robots or voice assistant agents used as public speaking coaches. Participants were requested to rate a statement, either “*Voice assistant agents can be good coaches*” or “*Robots can be good coaches*,” using a continuous scale that ranged from “Strongly disagree” to “Strongly agree.” Additionally, participants were encouraged to provide explanations or justifications for the ratings they assigned.

Additionally, participants were queried about their preferences regarding a coaching agent, with four available options: (1) the same robot or voice assistant agent they had viewed in the video, (2) a different robot or voice assistant agent than the one they had seen, (3) an alternative form of an interactive system (for example, a voice assistant agent for those in robot conditions and vice versa), or (4) a human coach. Participants were requested to explain why they favoured a particular agent or a human coach. In Section 3.2.5, we present thematic analyses of participants’ comments regarding their preferences. To read the question items please review Appendix B.7.

To enhance the readability and organization of the questionnaires, each of the previously described questionnaires was presented on a separate page within the online interface. Participants completed one questionnaire before proceeding to the next. Following each questionnaire, we included an attention check to ensure participants’ engagement and attentiveness. These attention checks consisted of True-False questions, such as “*The presentation topic was about a city in Canada: True - False.*” We posed five attention checks in total.

It’s important to note that participants’ payment did not depend on their performance in these attention checks. However, in our data analysis, we decided to exclude data from participants who answered more than two out of five attention check questions incorrectly, as this could indicate a lack of engagement or attentiveness during the study.

3.2.2 Experimental Conditions

In the two conditions involving the Pepper robot, the robot was positioned in front of the student with its head and torso oriented toward the student, as depicted in Figure 3.1(a). In the voice assistant condition, a Google Nest Mini device was placed on a table in front of the student, as illustrated in Figure 3.1(b). As mentioned previously, we took steps to control for various factors that could potentially influence participants’ perceptions of the agents. These included using identical audio across all three conditions and strategically positioning the camera to ensure that the “student presenter” in the video remained anonymous and unidentifiable to participants.

While the verbal behaviour of the agents was consistent, the non-verbal behaviour of the agents differed depending on the experimental condition as follows:

Condition 1 - ActiveListenR: Throughout the presentation, the Pepper robot utilized non-verbal backchanneling through head nodding, in order to demonstrate active listening. During recording the video, the Wizard of Oz method was used, where one of the researchers controlled the timing of the Pepper robot’s head nods. It was intended for the nods to be synchronized with the end of utterances and brief pauses during the presenter’s speech as a reference for when the robot should nod. It is important to note that the study did not examine how participants perceived the timing of these nods, and they were added at the researcher’s discretion.

As outlined in [94], variations in the amplitude and frequency of head nods can convey different meanings. To minimize the potential impact of constant identical movements on the study’s outcomes, and to make the listening behaviour seem more natural we implemented two modes of head nodding for the robot in a randomized order to serve as non-verbal backchanneling. One mode consisted of a slow vertical movement of the robot’s head, while the other mode comprised two consecutive vertical movements with a relatively quicker pace and a smaller amplitude of the robot’s head joint.

In the video, following the presentation and during the feedback phase, the robot employed non-verbal gestures, which were consistent for both robot conditions. These gestures and arm movements of the Pepper robot were meticulously designed using the Choregraphe [26] software, and their timing was fine-tuned to align with the verbal content. The inclusion of these gestures served two main purposes: first, to leverage the expressive capabilities of the robot’s actuators, and second, to establish a clear distinction between the physical embodiment of the voice assistant agent and the social robot.

Condition 2 - PassiveListenR: Throughout the presentation, the Pepper robot remained in a static position with an immobile head, and it did not exhibit any backchan-



Figure 3.2: A display of the lights of the Google Nest in the video

neling behaviour. The robot’s positioning and orientation with respect to the “student presenter” were identical to those in Condition 1. Following the presentation and during the feedback phase, the hand gestures, arm movements, and timing were consistent with those in Condition 1.

Condition 3 - VoiceAgent: Participants observed a Google Nest Mini device positioned on a table in front of the student. Throughout the presentation, the Google Nest device remained silent and was in listening mode. In this mode, Google Nest typically activates four lights. Following the presentation, these lights were deactivated, and the Google Nest device commenced providing the same verbal feedback as in the preceding two conditions.

The inclusion of these lights was intended to align the interactions with the voice assistant agent to the common behaviour of a Google Nest device, which could impact participants who were familiar with the device. In the video, the intensity of the lights was moderate to not cause undue distractions (please refer to Figure 3.2), but we still wished for them to be noticeable to those who were attentive.

3.2.3 Participants

A total of 184 participants were recruited through the MTurk platform, with specific inclusion criteria. Participants were limited to those residing in Canada and the United States, and they were required to have an approval rate of at least 80% based on a minimum of 50 HITs completed. Out of the total participants, 130 were from Canada, and 54 were from the United States.

Among the 184 participants, 16 records were excluded due to failing to pass more than

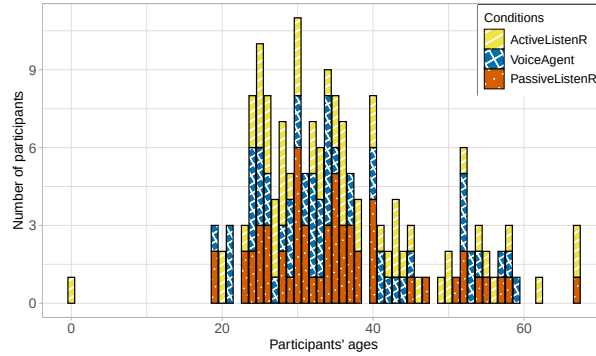


Figure 3.3: The histogram plot depicts participants’ age distributions across different conditions.

two attention checks. Consequently, the final dataset included 168 participants, with 38 from the United States and 130 from Canada. Of these participants, 66 self-identified as female, 100 as male, and 2 chose not to disclose their gender. The participants’ ages ranged from 19 to 67 years old.

The distribution of participants across the three experimental conditions was as follows:

- ActiveListenR condition: 58 participants (28 self-identified as female, 28 as male, 2 did not share gender information; age range: 20 to 67 years old, mean age: 36.24).
- PassiveListenR condition: 56 participants (18 self-identified as female, 38 as male; age range: 23 to 67 years old, mean age: 35.32).
- VoiceAgent condition: 54 participants (20 self-identified as female, 34 as male; age range: 21 to 59 years old, mean age: 35.20).

For a visual representation of the age distributions of participants in the three experimental conditions, please refer to Figure 3.3.

The study was estimated to take approximately 20 minutes to complete, and participants were compensated with a payment of 4 USD upon finishing the study. It’s important to note that the study received approval from the Human Research Ethics Board of the University of Waterloo. Participants were required to use one of three specific web browsers (Chrome, Safari, or Firefox) that had undergone extensive testing to ensure compatibility with our interface.

Additionally, participants were limited to using laptops or desktop computers for their interactions with the system. This decision was made to ensure that participants could engage with videos and questions on larger screens, as opposed to smaller smartphone screens. It is worth noting that all elements on each page were dynamically adjusted to fit the size of the participant’s screen, enhancing the user experience and readability.

Table 3.2: Participants’ Responses on Previous Public Speaking Classes Experiences - The first number denotes the count of participants, while the second number indicates the corresponding percentage of participants.

Experimental Condition	Previous Experience		Perceived Necessity	
	Yes	No	Yes	No
VoiceAgent	16 - 29.63%	36 - 66.66%	20 - 37.04%	34 - 62.96%
PassiveListenR	14 - 25%	40 - 71.43%	23 - 41.07%	32 - 57.14%
ActiveListenR	18 - 31.03%	37 - 63.79%	25 - 43.10%	34 - 58.62%

In Table 3.2, you can find information about the number of participants in each condition who had prior attendance in public speaking classes. Additionally, it notes whether they expressed the need to attend public speaking classes or workshops. It’s important to mention that some participants opted not to share this particular information.

3.2.4 Statistical Analysis

In our analysis, we employed Generalized Linear Models (GLMs) [89], specifically considering a Gaussian family. These models were utilized to explore the significant effects of the experimental conditions. Additionally, we used GLMs to account for potential confounding variables, including those gathered during the pre-experimental phase.

Our approach involved a step-wise simplification of the GLMs, aiming to create models that minimized Akaike’s Information Criterion (AIC) [18]. This process helped us identify the most relevant subset of factors to incorporate into the model for a more accurate and informative analysis of the dependent variables assessed through the questionnaires.

3.2.5 Results

In this section, we present the results of the analyses of participants’ responses to the questionnaires. Throughout this section, we will also address the Research Questions (RQs) by providing insights into the statistical analyses conducted.

Behaviour Evaluation of the Agent as a Coach: Participants’ assessments of the agents’ behaviours in their roles as coaches, addressing RQ1-1 and RQ2-1, are visually presented in Figures 3.4. The specific questions corresponding to these evaluations were

provided in Table 3.1. Table 3.3 displays the outcomes of a Generalized Linear Model (GLM) that predicted the ratings for each item according to the experimental condition.

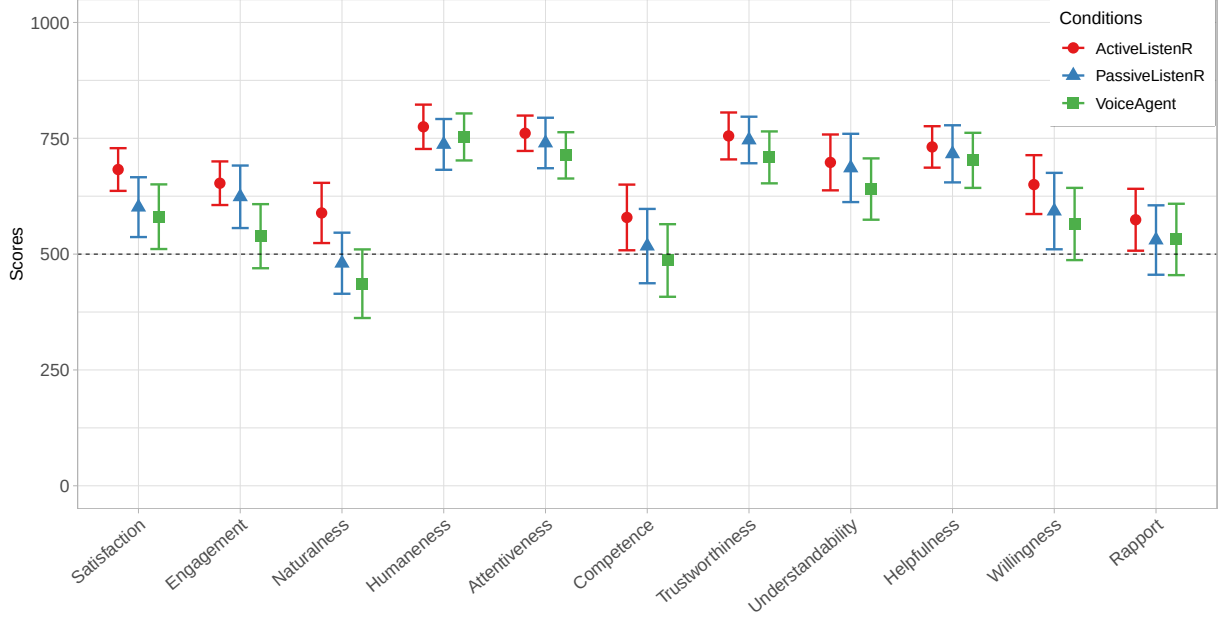


Figure 3.4: Generalized linear model predicting participants' scores regarding the agents' evaluations. The dashed line specifies the neutral choice at the centre of the continuous scale (data obtained ranging from 0 to 1000). Error bars indicate 95% confidence intervals.

The analysis revealed that participants rated the ActiveListenR agent as a coach significantly higher in terms of satisfaction ($se = 42.47, t = 2.642, p < .01$) and engagement ($se = 43.41, t = 2.535, p < .05$) compared to the VoiceAgent. Additionally, participants perceived the ActiveListenR agent as significantly more natural ($se = 46.98, t = 3.535, p < .001$) and humane ($se = 52.363, t = 1.986, p < .05$) than the VoiceAgent.

However, there were no significant differences observed between the PassiveListenR agent and the VoiceAgent in terms of satisfaction, engagement, naturalness and humaneness, as indicated in Table 3.3. Pairwise comparisons between the ActiveListenR and PassiveListenR conditions revealed that participants perceived the ActiveListenR agent as significantly more satisfying ($se = 41.99, t = 2.041, p < .05$) and more natural ($se = 46.62, t = 2.466, p < .05$) compared to the PassiveListenR agent.

The analysis did not uncover any significant differences among the three agents with regard to the other items in the Questionnaire shown in Table 3.1. These items include

Table 3.3: Generalized linear model predicting ratings of Satisfaction, Engagement, Naturalness, and Humaneness. Factors included in the model for each dependent variable (each column) were selected based on the Akaike information criterion (AIC). If a row is empty for a dependent variable, that means the factor was not included in the final reduced GLM predicting that variable. The estimates were according to comparisons to the baseline level (VoiceAgent). (* = $p < .05$; ** = $p < .01$; *** = $p < .001$)

Covariate	Satisfaction				Engagement			
	Estimate	SE	t	p	Estimate	SE	t	p
Condition								
ActiveListenR	112.197	42.47	2.642	<.01**	110.04	43.41	2.535	<.05*
PassiveListenR	26.492	42.276	0.627	.532	78.78	43.89	1.795	.074
Age								
Age	-2.184	1.557	-1.403	.162				
Gender								
Male	-68.647	35.836	-1.916	.057				
TIPI								
Agreeableness					23.18	15.77	1.470	0.143
Conscientiousness	-38.048	15.016	-2.534	<.05*				
Emotional Stability	35.771	13.506	2.648	<.01**				

Covariate	Naturalness				Humaneness			
	Estimate	SE	t	p	Estimate	SE	t	p
Condition								
ActiveListenR	166.080	46.98	3.535	<.001***	103.999	52.363	1.986	<.05*
PassiveListenR	51.135	46.867	1.091	.277	46.471	53.065	0.876	.382
Age								
Age	-3.697	1.733	-2.134	<.05*	-2.929	1.906	-1.537	.126
Gender								
Male	-75.161	41.014	-1.833	.069				
TIPI								
Extraversion					21.558	13.575	1.588	.114
Agreeableness	37.818	18.501	2.044	<.05*	53.885	19.816	2.719	<.01**
Conscientiousness	-58.65	16.861	-3.478	<.001***	-54.461	18.177	-2.996	<.01**
Emotional Stability	27.792	15.302	1.816	.071				

attentiveness, competence, trustworthiness, feedback understanding, helpfulness, rapport, and willingness to use the agent in the future.

The study’s findings offer partial support for H1-1 and H2-1. Specifically, participants rated the ActiveListenR agent more favourably than the VoiceAgent across four behavioural measures related to their role as a public speaking coach, supporting H1-1. Additionally, participants perceived the ActiveListenR agent as a more satisfying and natural coach compared to the PassiveListenR agent, providing support for H2-1.

Table 3.4: The results of generalized linear models show that participants’ gender and TIPI personality dimensions are significantly correlated with evaluations of agents’ attentiveness and competence, as well as the willingness of future use and feeling of rapport with the technology. Factors included in the model for each dependent variable (each column) were selected based on the Akaike information criterion (AIC). (* = $p < .05$; ** = $p < .01$; *** = $p < .001$)

Covariate	Attentiveness				Competence			
	Estimate	SE	t	p	Estimate	SE	t	p
Gender								
Male	-61.87	30.30	-2.042	<.05*	-55.43	28.30	-1.959	.052
TIPI								
Agreeableness	28.22	13.14	2.148	<.05*	25.58	12.27	2.084	<.05*
Covariate	Willingness				Rapport			
	Estimate	SE	t	p	Estimate	SE	t	p
Gender								
Male	-102.26	43.89	-2.330	<.05*	-86.80	42.29	-2.052	<.05*
TIPI								
Agreeableness	40.96	19.76	2.073	<.05*	41.62	19.04	2.186	<.05*
Conscientiousness	-50.11	17.89	-2.801	<.01**	-55.12	17.05	-3.233	<.01**
Openness to Experiences	28.64	16.63	1.722	.087				

As indicated by the results presented in Table 3.3 and Table 3.4, and in response to RQ3, we observed a positive correlation between the agreeableness personality trait and participants’ perceptions of naturalness ($se = 18.501, t = 2.044, p < .05$), humaneness ($se = 19.816, t = 2.719, p < .01$), attentiveness ($se = 13.14, t = 2.148, p < .05$), and competence ($se = 12.27, t = 2.084, p < .05$) in the agent acting as a coach. Furthermore, their willingness to use the agent in the future ($se = 19.76, t = 2.073, p < .05$) and the level of rapport they reported experiencing with the agent ($se = 19.04, t = 2.186, p < .05$) exhibited positive correlations with the agreeableness trait.

The conscientiousness personality trait displayed several significant negative effects. It had a negative impact on participants’ perceptions of the agent’s naturalness ($se =$

16.861, $t = -3.478, p < .001$) and humaneness ($se = 18.177, t = -2.996, p < .01$). Additionally, conscientiousness was negatively correlated with participants' satisfaction with the agent as a coach ($se = 15.016, t = -2.534, p < .05$), their willingness to use it in the future ($se = 17.89, t = -2.801, p < .01$), and the level of rapport they reported feeling with the agent ($se = 17.05, t = -3.233, p < .01$). Additionally, participants' emotional stability exhibited a positive correlation with their perception of satisfaction ($se = 13.506, t = 2.648, p < .01$).

The results regarding the correlations between ratings and the Big Five personality traits are not influenced by the experimental conditions, as evidenced by the non-significance of the two-way interaction between Big Five personality traits and the experimental conditions. These findings provide support for H3.

The analysis of two-way interactions between the gender factor and the experimental conditions factor revealed some correlations. These results are summarized below:

- Male participants found the VoiceAgent significantly less engaging ($se = 43.41, t = -2.535, p < .05$) than the other two agents.¹
- Male participants willingness to use the ActiveListenR agent was significantly higher compared to the PassiveListenR agent ($se = 105.33, t = 2.178, p < .05$).
- Male participants reported feeling a higher level of rapport towards the ActiveListenR agent in comparison to the PassiveListenR agent ($se = 100.95, t = 2.802, p < .001$).
- Regardless of the experimental conditions, male participants tended to find the agent's feedback significantly less trustworthy than female participants ($se = 30.59, t = -2.266, p < .05$).
- Female participants, irrespective of experimental conditions, found the agents to be significantly more attentive ($se = 30.30, t = 2.042, p < .05$) and expressed a higher willingness to use them in the future ($se = 43.89, t = 2.330, p < .05$).
- Additionally, female participants reported significantly higher levels of rapport with the agents ($se = 42.29, t = 2.052, p < .05$).

These findings highlight gender-related differences in participants' perceptions and interactions with the agents across the various experimental conditions.

Perception of Human Nature Attributes of the Agents: We assessed four Human Nature (HN) attributes for each agent, which include emotional responsiveness, warmth, openness, and agency (liveliness).

The results, addressing RQ2-1 and RQ2-2, indicate that perceptions of the Human Nature (HN) attributes were significantly influenced by the existence of the social robot,

¹"Male" refers to participants who self-identified as male, and "female" refers to participants who self-identified as female.

Table 3.5: Generalized linear model predicting the agents’ human nature on four attributes (emotionality, warmth, openness, agency). Factors included in the model for each dependent variable were selected based on the Akaike information criterion (AIC). If a row is empty for a dependent variable, that means the factor level did not exist in the final reduced GLM predicting that variable. The estimates were according to comparisons to the baseline level (VoiceAgent). (* = $p < .05$; ** = $p < .01$; *** = $p < .001$)

Covariate	Emotionality				Warmth			
	Estimate	SE	t	p	Estimate	SE	t	p
Condition								
ActiveListenR	147.49	46.19	3.193	<.01**	172.01	46.47	3.701	<.001***
PassiveListenR	106.43	46.12	2.308	<.05*	112.91	46.39	2.434	<.05*
Gender								
Male	-110.23	39.50	-2.790	<.01**	-121.42	39.74	-3.055	<.01**
TIPI								
Agreeableness	58.43	17.60	3.321	<.01**	53.20	17.70	3.005	<.01**
Conscientiousness	-48.26	15.88	-3.038	<.01**	-34.91	15.98	-2.185	<.05*

Covariate	Openness				Agency			
	Estimate	SE	t	p	Estimate	SE	t	p
Condition								
ActiveListenR	145.74	49.59	2.939	<.01**	122.93	44.15	2.784	<.01**
PassiveListenR	108.86	49.51	2.199	<.05*	100.63	44.08	2.283	<.05*
Gender								
Male	-131.91	43.24	-3.051	<.01**	-67.31	37.76	-1.783	.076
TIPI								
Agreeableness	34.69	19.43	1.785	.076	48.57	16.82	2.888	<.01**
Conscientiousness	-28.26	17.82	-1.586	.115	-30.04	15.18	-1.979	<.05*
Emotional Stability	29.45	16.01	1.840	.068				

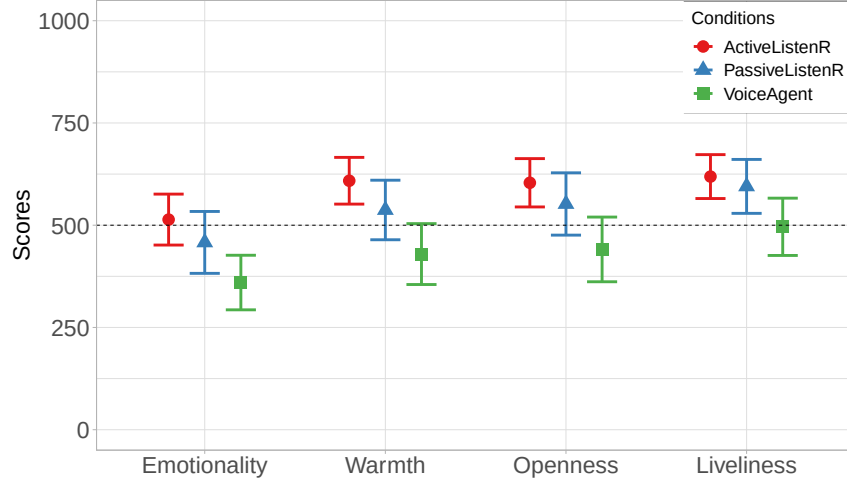


Figure 3.5: Human Nature scores of the agent in different conditions. The dashed lines specify the neutral choice at the centre of the continuous scale (ranging from 0 to 1000). Error bars indicate 95% confidence intervals.

both in the ActiveListenR and PassiveListenR conditions. As illustrated in Table 3.5 and Figure 3.5, when compared to the VoiceAgent, both the ActiveListenR and PassiveListenR agents received significantly higher ratings on all four HN attributes.

According to Table 3.5, in comparison to the VoiceAgent, the ActiveListenR and PassiveListenR agents were scored higher on the following HN attributes: Emotional Responsiveness (ActiveListenR: $se = 46.19, t = 3.193, p < .01$; PassiveListenR: $se = 46.12, t = 2.308, p < .05$), Warmth (ActiveListenR: $se = 46.47, t = 3.701, p < .001$; PassiveListenR: $se = 46.39, t = 2.434, p < .05$), Openness (ActiveListenR: $se = 49.59, t = 2.939, p < .01$; PassiveListenR: $se = 49.51, t = 2.199, p < .05$), and Agency (ActiveListenR: $se = 44.15, t = 2.784, p < .01$; PassiveListenR: $se = 44.08, t = 2.283, p < .05$). These findings suggest that the presence of the social robot in both ActiveListenR and PassiveListenR conditions had a significant positive impact on participants' perceptions of these HN attributes when compared to the VoiceAgent.

The results offer support for H1-2, indicating that the social robot in both the ActiveListenR and PassiveListenR conditions was perceived as more human-like with respect to the Human Nature attributes compared to the voice assistant agent (VoiceAgent). However, H2-2 was not supported as there were no significant differences observed between the ActiveListenR agent and the PassiveListenR conditions regarding perceptions of HN attributes.

In response to RQ3 and in alignment with H3, both agreeableness and conscientiousness exhibit significant correlations with perceptions of Human Nature (HN) attributes. Agreeableness displays a positive correlation with the following HN attributes: (Emotional Responsiveness: $se = 17.60, t = 3.321, p < .01$, Warmth: $se = 17.70, t = 3.005, p < .01$, Agency: $se = 16.82, t = 2.888, p < .01$). On the other hand, conscientiousness shows a negative correlation with the same HN attributes: (Emotional Responsiveness: $se = 15.88, t = -3.038, p < .01$, Warmth: $se = 15.98, t = -2.185, p < .05$, Agency: $se = 15.18, t = -1.979, p < .05$).

It is worth noting that irrespective of the experimental condition, gender played a role in participants' perceptions of Human Nature (HN) attributes. Specifically, men tended to perceive the agent as significantly less emotional ($se = 39.50, t = -2.790, p < .01$), less warm ($se = 39.74, t = -3.055, p < .01$), and less open ($se = 43.24, t = -3.051, p < .01$) when compared to women.

Perceived Social Attributes of the Agent:

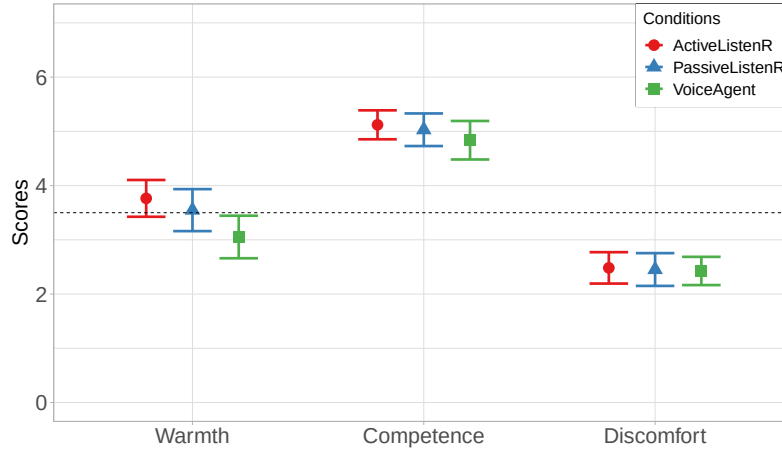


Figure 3.6: The scores for the three social dimensions (warmth, competence, discomfort) defined in RoSAS across the three experimental conditions. The dashed line indicates the neutral choice at the center of the seven-point Likert scale. The error bars indicate 95% confidence intervals.

In addressing RQ3-1 and RQ3-2, we employed a Generalized Linear Model (GLM) on the ratings obtained from the RoSAS questionnaire. The outcomes of this analysis are presented in [Figure 3.6](#) and detailed in [Table 3.6](#).

According to the results, both the ActiveListenR ($se = 0.248, t = 2.796, p < .01$) and

Table 3.6: Generalized Linear Models predicting the RoSAS dimensions about the agent (Warmth, Competence, Discomfort). Factors included in the model for each dependent variable were selected based on the Akaike information criterion (AIC). If a row is empty for a dependent variable, that means the factor level did not exist in the final reduced GLM predicting that variable. The estimates were according to comparisons to the baseline level (VoiceAgent).(* = $p < .05$; ** = $p < .01$; *** = $p < .001$)

Covariate	Warmth				Competence				Discomfort			
	Estimate	SE	t	p	Estimate	SE	t	p	Estimate	SE	t	p
Condition												
ActiveListenR	0.692	0.248	2.796	<.01**								
PassiveListenR	0.505	0.247	2.045	<.05*								
Age												
Age									0.013	0.007	1.779	.077
Gender												
Male	-0.638	0.216	-2.954	<.01**					0.346	0.166	2.083	<.05*
TIPI												
Extraversion									0.079	0.049	1.606	.110
Agreeableness	0.328	0.097	3.387	<.001***	0.2	0.077	2.583	<.05*	-0.121	0.075	-1.625	.106
Conscientiousness	-0.315	0.089	-3.541	<.001***								
Emotional Stability	0.155	0.08	1.943	.054					-0.201	0.06	-3.315	<.01**

the PassiveListenR ($se = 0.247, t = 2.045, p < .05$) agents received higher ratings on the warmth dimension compared to the VoiceAgent. It's important to note that warmth is one of the Human Nature (HN) attributes explicitly measured as one of the four attributes. However, in RoSAS, warmth is a social dimension calculated based on six attributes in the questionnaire (Organic, Sociable, Emotional, Compassionate, Happy, Feeling).

On the other hand, there were no significant differences found between the VoiceAgent and the other two agents in terms of the competence dimension (ActiveListenR: $se = 0.221, t = 0.792, p = 0.429$; PassiveListenR: $se = 0.223, t = 0.649, p = 0.517$) and the discomfort dimension (ActiveListenR: $se = 0.195, t = 0.711, p = 0.478$; PassiveListenR: $se = 0.197, t = 0.967, p = 0.335$). As a result, H1-3 is partially supported by the results.

However, H2-3 was not validated, as no significant differences were identified between the ActiveListenR and PassiveListenR agents in relation to the dimensions of RoSAS.

In response to RQ3, the study revealed that two personality traits, agreeableness and conscientiousness, had notable impacts on participants' assessments of the agent, thereby partially supporting H3-3. Specifically, agreeableness exhibited a positive and significant effect on perceived social warmth ($se = 0.097, t = 3.387, p < .001$) and perceived social competence ($se = 0.077, t = 2.583, p < .05$). On the other hand, conscientiousness had a significant negative effect on perceived social warmth ($se = 0.089, t = -3.541, p < .001$). Emotional stability had a significant negative effect on perceived social discomfort ($se =$

0.06, $t = -3.315$, $p < .05$).

Participants who self-identified as male rated the agent’s social warmth significantly lower than those who self-identified as female ($se = 0.216$, $t = -2.954$, $p < .01$). The perceived level of social discomfort was significantly higher for males than for females ($se = 0.166$, $t = 2.083$, $p < .05$).

Perceived Intelligence, and Likeability of the Agent: The results pertaining to Godspeed perceived intelligence and Godspeed likeability are presented in [Figure 3.7](#).

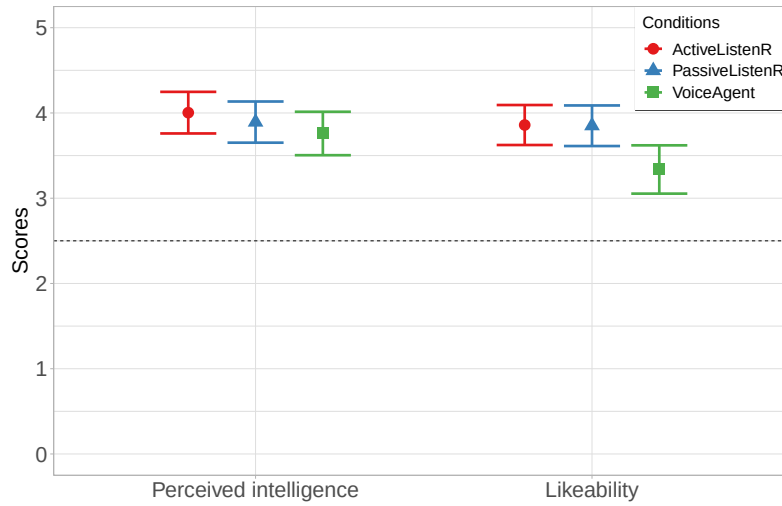


Figure 3.7: The average of the Godspeed scores on the dimensions of perceived intelligence and likeability across all experimental conditions. The dashed line indicates the neutral choice at the center of the five-point Likert scale. The error bars indicate 95% confidence intervals.

In response to RQ1-4 and RQ2-4, our data did not provide support for H1-4 and H2-4, as no significant differences were observed between the agents in terms of perceived intelligence. When using the VoiceAgent as the reference level, the predicted results from the GLM model for the ActiveListenR agent were not statistically significant ($se = 0.180$, $t = 1.093$, $p = 0.276$), and for the PassiveListenR agent, they were also not significant ($se = 0.182$, $t = 0.844$, $p = 0.4$).

However, as indicated in [Table 3.7](#), both the ActiveListenR and PassiveListenR agents were rated significantly more likable than the VoiceAgent (ActiveListenR: $s = 0.176$, $t = 2.887$, $p < .01$; PassiveListenR: $se = 0.178$, $t = 2.685$, $p < .01$), thereby supporting H1-5.

Table 3.7: Generalized Linear Models predicting the Godspeed likeability. Factors included in the model for each dependent variable were selected based on the Akaike information criterion (AIC). The estimates were according to comparisons to the baseline (VoiceAgent). (* = $p < .05$; ** = $p < .01$; *** = $p < .001$)

Covariate	Likeability			
	Estimate	SE	t	p
Condition				
Active listening robot	0.508	0.176	2.887	<.01**
Passive listening robot	0.477	0.178	2.685	<.01**
Age				
age	-0.011	0.006	-1.715	.088
TIPI				
Agreeableness	0.138	0.064	2.142	<.05*

Nevertheless, H2-5 was not supported since there was no significant difference found in likeability scores between the ActiveListenR agent and the PassiveListenR agent.

Additionally, in addressing RQ3, the agreeableness personality trait had positive correlations with likeability scores ($se = 0.064, t = 2.142, p < .05$).

Thematic Analysis: The analysis of participants’ answers to open-ended questions followed the six-step Thematic Analysis (TA) method developed by Braun and Clarke, as outlined in their work [19]. Initially, codes, which are labels assigned to specific pieces of text, were generated from extracted excerpts. For one of the analyses, specifically the first open-ended question where we had a greater number of comments, a theme was considered valid if we identified a minimum of 20 codes related to it across all conditions. In the second analysis, which focused on discerning the prominent reasons for selecting a choice in a multiple-choice question, we set a threshold of at least 4 codes occurrences to accept a theme, as some options (e.g., a different agent) were less frequently selected than others.

Themes underwent a review and revision process, and sub-themes were formulated, primarily based on whether the attitudes expressed within each theme were positive or negative. One author conducted the entire TA of the dataset, while another author independently analyzed 20% of the data. Inter-rater reliability was calculated using kappa’s agreement score [75]. The results of calculating weighted kappa indicated a score of 0.873 ($kappa = 0.823, SE = 0.054, 95\%, CI = [0.717, 0.929]$), reflecting “almost perfect agreement.”

Figure 3.8 illustrates the participants’ ratings regarding their sentiments about using

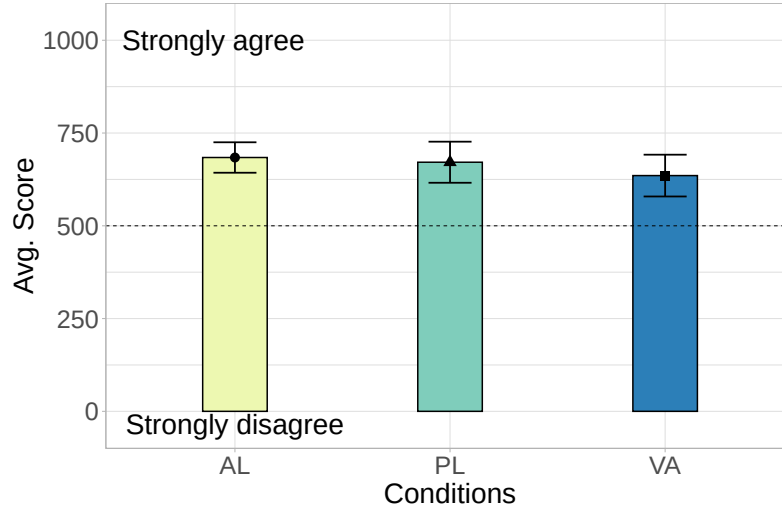


Figure 3.8: Ratings on whether participants agreed with using the agent they saw as a public speaking coach (rated from “Strongly disagree” to “Strongly agree”). Mean values and error bars (95% confidence intervals) are shown. The dashed line indicates the neutral choice. Recorded responses changed from 0 to 1000. AL: ActiveListenR, PL: PassiveListenR, VA: VoiceAgent.

the agent they were exposed to as a public speaking coach. On average, the ratings are quite elevated (around 700 out of 1000). This indicates that participants were supportive of the concept of employing either a social robot or a voice assistant agent as a public speaking coach.

A total of six main themes were discerned, each of which was further categorized with 3 to 5 sub-themes to provide a more comprehensive representation of the comments. Table 3.8, presents the themes and their associated sub-themes, and these will be elaborated upon in the subsequent sections. It’s important to note that the quotes provided serve as illustrative examples and may be drawn from any of the three experimental conditions (VA: VoiceAgent, AL: ActiveListenR, PL: PassiveListenR), as they represent themes that were consistent across all experimental conditions.

Theme 1 - Technology Acceptance

125 participants provided their views on this theme. In this theme, we describe how people perceive technology and their willingness to use it in the future. There are four sub-themes associated with this theme.

1.1 Interested in the technology: 76 Participants expressed explicit interest after watching the video demonstration. *“I’ve truly never thought about it. I didn’t even realize that this type of coaching existed. It was pretty neat to see a robot listening and giving feedback in that way (AL).” “I think this is a great idea. It is a way for people to practice and improve before having to actually present (VA).”*

1.2 Open to the technology 23 comments did not suggest that they would object to using/trying this technology and were open to it. *“I would not mind to give it a try and see if that would improve my public speaking skill (AL).” “I feel somewhat positive about having Voice Assistant agents as public speaking coaches (VA).”*

1.3 Skeptical about accepting the technology: 12 Participants expressed doubt and skepticism. *“I have mixed emotions when it comes to a robot for a public speaking coach, I am not sure they are able to pick up on every detail (PL).” “Just basing it off the robot’s feedback in the video, I’m not sure they make good public speaking coaches (AL).”*

1.4 Not interested in the technology: 14 comments suggested that they were not interested in using such technology. They cited reasons such as discomfort with interacting with social agents in general and preferring to interact with humans. Additionally, some participants did not believe that the technology was useful or valuable. *“I like to see people, I do not like interacting with robots. It feels silly (AL).” “It seems like it wouldn’t help. I would not feel reassured or engaged by a voice assistant providing feedback. Person-to-person is a much better experience as it allows you to engage and empathize with the coach’s feedback (VA).”*

Theme 2 - Advantages/Shortcomings of Technology

There were 30 comments about this theme. Participants discussed what this technology can and cannot do, as well as how certain issues may or may not be addressed by it.

2.1 A useful tool to help nervousness/public speaking anxiety/increasing confidence: 15 participants remarked that the system could be useful for people who may get nervous when delivering a public presentation. *“Often times, I am nervous to practice talking in front of friends and family, but I would like the idea of using a voice assistant (VA).” “I feel like it would be a good first step for people who are afraid of speaking in front of others (AL).”*

2.2 Does not help with fear/anxiety of public speaking: 4 participants stated that the agent does not help with the fear of public speaking in front of a live audience. *“They wouldn’t help too much with the fear of speaking before a live audience (PL).” “I think it*

can be a helpful starting point to get you used to the cadence of public speaking, but I don't think it will actually help you get over the anxiety of it (AL)."

2.3 Useful tool for beginners/students/for practicing basic skills: 21 participants remarked on the application of such agents to practice some of the fundamentals of delivering a presentation. *"I think that it could be very beneficial for University students (PL)."* *"I think they could be helpful for people who are struggling with the general basics of public speaking (AL)."*

Theme 3 - Attitude towards Agent

A total of 43 participants discussed their attitudes towards the agent they were exposed to.

3.1 Agent addresses only a few aspects of the presenter: According to 18 participants, the agent was not able to pick up every detail, such as non-verbal cues, and provide the student with more personalized feedback. *"I'm not sure it took into account that the speaker sounded to be an ESL. So the comment about inflections may be due to that, as opposed to public speaking skill (VA)."* *"There's a lot of important nonverbal stuff like body language and vocal tone that a robot couldn't really teach (PL)."*

3.2 Agent was strange/creepy/unnatural/robotic: According to 19 participants, the agent was unfriendly and creepy. Since the Pepper robot's voice was used in all conditions, we received some comments that the agent's voice was robotic. *"Kind a weird. I think it would be better if (you at least had the option) of it emailing you a report or something (VA)."* *"I think it's awkward and a bit creepy, to be honest (AL)."*

3.3 Agent does not understand human emotions: 13 participants said that public speaking involves an emotional aspect, and since the agent had no human emotions, it could not provide feedback on this aspect. *"I wonder what the possibilities and limits are when it comes to being able to sense and evaluate the passion and emotion of a speaker, along with the relative context of a subject (PL)."* *"They don't take human emotion into consideration (AL)."*

3.4 Agent was unresponsive/weird during presentation: 5 comments (all in the PassiveListenR condition) were regarding the agent being unresponsive during the presentation. *"The only negative was I felt like the robot was unresponsive during the presentation. I felt like some kind of emotion or gestures during the presentation would have been nice (PL)."*, *"The robot didn't seem very engaged during the speaking - it would be more human if it showed that it was listening somehow (PL)."*, *"I feel that to give a public presenta-*

tion you need to 'feel' your audience. The robot from the presentation is stiff during the student's presentation; not good (PL)."

Theme 4 - Interaction Perceptions

A total of 21 participants provided input regarding how they envisioned and perceived the interaction in general.

4.1 Interaction is odd: 4 comments in the ActiveListenR and the PassiveListenR conditions were related to interactions with the social robot and speaking with it felt strange. *"It feels somehow strange but I believe it would be helpful to consider it (AL)."* *"However, it still feels strange having the interaction with a robot instead of a human. I think I would become more comfortable with them over time (PL)."*

4.2 Interaction is easy and comfortable (no fear of being judged): 12 participants stated that the agent provided a reasonable level of comfort to practice the presentation, and it felt easier than standing in front of a human. *"I think they can be very good as it's much less intimidating than a real human. They seem to be easier to learn from and get feedback from (PL)."* *"I think it's good because it's less nerve-racking presenting in front of a machine (VA)."*

4.3 Interaction is different from real life: 9 participants said that the interaction differed from the real-life experience, i.e., presenting in front of a live audience. *"Presenting to a robot vs a real human would be easier and not representative of how you would feel in front of a human audience (AL)."*; *"It is different than being in front of live people so it may not fully prepare someone for that experience (PL)."*

Theme 5 - Agent vs. Human

29 comments compared the agent they observed as a coach with a human coach.

5.1 Better than humans in some aspects: 4 participants indicated that the agent performed well, even better than a human. *"They may pick up things, not every human public speaking coach can pick up (PL)."* *"You can practice public speaking and receive advice instantly anytime you want while you cannot do the same with another person (PL)."*

5.2 A human coach should be next to the agent: 3 participants indicated having a human coach beside the agent could be beneficial. *"I think though that it would be harder to only work with a robot, I think both a robot and human coach would work best together"*

(PL).” “It might be better, in some instances, for students to have one other human person in the room to help with additional feedback (AL).”

5.3 Human coach/feedback would be preferred: 17 participants stated that this system could not replace a human coach. “I don’t think a robot could replace a person as a public speaking coach (PL).” “That is good advice, and they have the advantage of scrubbing through hundreds of thousands of data sets, but there will always be some less technical issues that only a human could understand (VA).”

5.4 Agent is a good alternative when no human coach is available: 8 participants remarked on the benefit of having such a coach in a situation where no human coach can be easily accessed. “It seems to be a very effective way to provide feedback when no alternative affordable professional help is available (AL).” “I think if you have no one else available, then it would be okay (VA).”

Theme 6 - Feedback

There were 46 comments regarding the feedback content, as well as the agent’s voice (the same under all conditions).

6.1 Helpful feedback: 30 comments were on how helpful and appropriate the feedback was. “I think it can be useful and provide some generally decent feedback (VA).” “It’s different, but I think they can help and process things differently and provide great feedback for you to improve (AL).”

6.2 Feedback/voice was not understandable: It was noted by 11 participants that the feedback was difficult to understand. “I like it, but if they can talk in ways that you can understand them (AL).” “I don’t mind, but the robot voice was a bit difficult to understand (PL).”

6.3 Feedback is generic: According to 6 participants, the feedback provided was quite generic and did not offer any ground-breaking information. “I think they could give good general advice, but it is hard to imagine they could successfully take into account all the different variables that are specific to different people’s situations (PL).” “I am unsure of how it can handle specific, in-depth feedback (VA).”

6.4 Feedback is unbiased/impersonal/honest: 10 participants expressed satisfaction with their impression of the feedback provided by the agent as unbiased and consistent. “The benefit of a robot is that it will give very consistent objective critique so you can actually track the progress based on the skill more so than the quality of content (AL).” “I feel like it could benefit people learning, less risk of a coach being mean (AL).”

6.5 Cannot (completely) trust the agent’s feedback: 8 participants expressed a lack of trust in the agent’s performance evaluation. “*The robot seems to be pre-programmed when giving feedback. I would not trust the robot’s assessment of my performance (AL).*” “*I’m not very confident that they would be able to provide good, accurate feedback though (VA).*”

Table 3.8: Thematic Analysis results on opinions about the agents. The percentage of responses related to each sub-theme is provided for each condition. The total percentage of participants commenting on each theme is shown below the dashed lines. While one participant comment might be counted in more than one sub-theme, depending on the content. As a result of this overlap, sub-theme percentages may not add up to their respective total percentages.

Themes	Sub-themes	VoiceAgent	ActiveListenR	PassiveListenR
Technology Acceptance	Interested in the technology	38.89%	58.62%	36.21%
	Open to the technology	11.11%	12.07%	17.24%
	Skeptical about accepting the technology	5.56%	5.17%	10.34%
	Not interested in the technology	11.11%	6.9%	6.9%
Total percentage of participants who commented on technology acceptance		66.67%	82.76%	73.21%
Advantages/Shortcomings of Technology	Useful tool to help nervousness /public speaking anxiety/ increasing confidence	9.26%	8.62%	8.62%
	Does not help with fear/anxiety of public speaking	5.56%	0%	1.72%
	Useful tool for beginners/students/for practicing basic skills	14.81%	12.07%	10.34%
	Total percentage of participants who commented on advantages/shortcomings of technology	24.07%	20.69%	8.93%
Attitude Towards Agent	Agent addresses only a few aspects of the presenter	12.96%	5.17%	13.79%
	Agent was strange/creepy/unnatural/robotic	20.37%	5.17%	8.62%
	Agent does not understand human emotions	5.56%	5.17%	12.07%
	Agent was unresponsive/weird during presentation	0%	0%	8.62%
Total percentage of participants who commented on attitude towards agent		25.93%	12.07%	39.29%
Interaction Perceptions	Interaction is odd	0%	5.17%	1.72%
	Interaction is easy and comfortable (no fear of being judged)	9.26%	1.72%	10.34%
	Interaction is different from real life	1.85%	5.17%	8.62%
Total percentage of participants who commented on interaction experience		9.26%	10.34%	17.86%
Agent vs. Human	Better than humans in some aspects	0%	3.45%	3.45%
	A human coach should be next to the agent	0%	3.45%	1.72%
	Human coach/feedback would be preferred	9.26%	10.34%	10.34%
	Agent is a good alternative when no human coach is available	7.41%	3.45%	3.45%
Total percentage of participants who commented on agent vs. human comparison		14.81%	17.24%	19.64%
Feedback Evaluation	Helpful feedback	14.81%	22.41%	15.52%
	Feedback/voice was not understandable	0%	8.62%	10.34%
	Feedback is generic	5.56%	3.45%	1.72%
	Feedback is unbiased/impersonal/honest	3.7%	6.9%	6.9%
	Cannot (completely) trust the agent’s feedback	5.56%	5.17%	3.45%
Total percentage of participants who commented on feedback evaluation		16.67%	36.21%	28.57%

A final question asked participants whether they preferred interacting with the same agent again, with a different agent, or with a human coach. Table 3.9 presents the findings from the thematic analysis, encompassing the rationales articulated by participants for their preferences, along with the number of participants who supported each of the options offered for a public speaking coach.

Table 3.9: Participants’ agent preferences in each condition. Numbers are divided by themes. The “total number (percentage)” of participants preferring the same agent, a different agent, another interactive system, or a human is shown for each condition.

	Themes	VoiceAgent	ActiveListenR	PassiveListenR
Same agent	because of objectivity and quality of feedback	10	17	16
	because the agent is non-judgemental, helps overcome fear	1	3	1
	because it was warm, friendly, natural	3	4	4
	Total number and percentage of participants selecting this choice	17 (31.48%)	25 (43.10%)	22 (39.28%)
Human	as they can provide more comprehensive feedback	10	11	10
	as they are compassionate, can communicate personally, is comfortable	7	9	7
	as the agent seems unnatural	4	2	2
	as they are more reliable, knowledgeable, experienced	5	6	6
	as they listen, and respond during presentation	0	0	4
	as I have to present to humans in real life	2	3	1
	Total number and percentage of participants selecting this choice	28 (51.85%)	30 (51.72%)	23 (41.07%)
Different agent	one that looks and sounds more human	6	0	3
	Total number and percentage of participants selecting this choice	7 (12.96%)	0 (0%)	6 (10.71%)
Another interactive system	Prefer more inanimate, interactive system such as a Google assistant	0	2	6
	Total number and percentage of participants selecting this choice	2 (3.7%)	2 (3.44%)	5 (8.92%)

3.3 Discussion

In this online study, participants were asked to assess the performance of three agents taking on the role of public speaking coaches. To explore participants’ perceptions of these diverse agents in the context of public speaking coaching, we examined two distinct agent types, differing in their physical embodiments (a voice assistant agent and a social robot). Furthermore, for one of the agents (the social robot), we introduced modifications to its non-verbal behaviour by incorporating head nods to signify active listening.

3.3.1 Perceptions of Agents as Public Speaking Coaches

Our study revealed that the ActiveListenR agent, which exhibited backchanneling behaviour, was perceived as more natural and satisfying compared to the PassiveListenR agent (which lacked backchanneling behaviour) and the VoiceAgent (a voice assistant agent). However, there were no significant differences in satisfaction and naturalness between the PassiveListenR agent and the VoiceAgent. These findings highlight the effectiveness of active listening behaviour in the agent, aligning with previous research that emphasizes the positive impact of “active engagement” on relational satisfaction [37, 14]. When individuals feel heard and understood, they tend to have higher satisfaction in their relationships, which can have substantial implications for their overall well-being and physical health [14].

Several studies have put forth probabilistic and reinforcement learning models aimed

at predicting the optimal timing for generating backchannels in human-agent dyadic conversations [65, 66, 85]. In our research, we employed a Wizard-of-Oz method to introduce backchannels without specifically considering the proper timing for their occurrence. Nevertheless, we observed that the mere act of head nodding by the agent while listening, even without precise timing considerations, had a significant impact on how the agent was perceived as being natural. As in a related study [86], researchers explored the use of random verbal backchannels in a voice assistant agent, showing that this approach resulted in prolonged interactions and increased user engagement.

Regarding agent engagement, participants perceived the ActiveListenR agent as being more engaging in its interaction with the student compared to the VoiceAgent. However, there was no significant difference in perceived engagement between the PassiveListenR and the VoiceAgent. While this could partially be attributed to the novelty effect of the social robot, the lack of significant difference between the PassiveListenR agent and the VoiceAgent suggests that the novelty effect isn't the sole factor influencing perceived engagement. There were no significant differences between the ActiveListenR agent and the PassiveListenR agent.

This corresponds to our **H1** but exclusively highlights the benefits of the ActiveListenR agent over the VoiceAgent. Our findings also reinforce prior research on the effect of active listening on engagement in human-agent interactions [115]. June et al. (2013) [69] observed that integrating backchanneling into robots led to an enhanced perception of engagement in the robots.

Notably, the ActiveListenR agent achieved significantly higher ratings for satisfaction, engagement, naturalness, and humaneness in comparison to the VoiceAgent. Conversely, we did not observe significant differences between the PassiveListenR agent and the VoiceAgent in these aspects. This observation is interesting since both the ActiveListenR and PassiveListenR agents employed nonverbal social cues, such as hand gestures and arm movements when providing feedback. However, it was only the agent demonstrating active listening behaviour that was perceived as a noteworthy improvement in the mentioned criteria compared to the VoiceAgent.

The research did not uncover any significant differences among the agents regarding other metrics more related to their effectiveness as coaches, such as helpfulness, attentiveness, competence, and trustworthiness. This observation aligns with the participants' overall agreement scores regarding the suitability of these agents as coaches, which remained consistently high across all conditions.

One plausible explanation for these findings could be that the perceived quality and accuracy of the agent's feedback had a more significant impact on people's evaluations of

its coaching abilities than its social behaviour. Additionally, the study found no significant differences between the participants’ sense of rapport, or their willingness to use the agents across different conditions. Perhaps, participants’ familiarity with smart speakers, their user-friendliness, and their widespread use, factors contributing to their popularity, may explain this lack of distinction [106].

Additionally, it’s possible that individuals who were exposed to the voice assistant agent couldn’t easily envision interacting with a social robot or incorporating social nonverbal cues into their feedback process. Consequently, they may have considered the voice-only feedback provided by the voice assistant agent to be sufficient for their needs.

3.3.2 Perceptions of the Human Nature and Social Attributes of Agents

Utilizing a social robot as the coach in both the ActiveListenR and PassiveListenR conditions resulted in higher ratings for HN (Human Nature) attributes, in comparison to the VoiceAgent condition. Hence, it appears that the agent’s embodiment, particularly the inclusion of anthropomorphic features and non-verbal social behaviours during feedback, may have had a positive impact on the perception of HN attributes. We didn’t discover solid evidence supporting the idea that active listening behaviour significantly influenced the perception of these HN attributes. Possibly the influence of non-verbal backchanneling behaviour during listening time was overshadowed by the extent of non-verbal behaviour employed in both robotic conditions for feedback time. In a similar study [73], the anthropomorphic robotic head was perceived as being more present than the non-anthropomorphic voice assistant agent, even when non-verbal behaviours like gaze and head movements were not involved.

The findings from the RoSAS questionnaire indicated that among the three dimensions—warmth, competence, and discomfort—both the ActiveListenR and PassiveListenR agents were perceived as possessing higher “warmth” in comparison to the VoiceAgent.

The heightened perceptions of warmth and friendliness observed in the robot can be attributed to its social capabilities. Human beings tend to exhibit a cognitive preference for social interactions in the presence of subtle social cues and anthropomorphic characteristics [73]. Establishing familiar channels of communication is especially pertinent for tasks that necessitate in-person collaboration [73]. According to [77], participants rate a robotic head with a human-like appearance as more socially engaging than a voice assistant agent. This suggests that the presence of anthropomorphic features in agents may elicit a sense of familiarity akin to natural human communication [77].

3.3.3 Perceived Intelligence and Likeability of Agents

The results suggest that there were no notable distinctions among the agents in terms of perceived intelligence. This could possibly be attributed to the stronger influence of the feedback phase and its content on participants’ perceptions of the agent’s intelligence. One plausible explanation for the lack of significant differences in perceived intelligence, as well as other factors like attentiveness and competence among the agents, might be that the effect of verbal feedback dominated the impact of non-verbal behaviour. It’s plausible that participants primarily assessed the agents as coaches based on the quality of the feedback they provided.

It’s possible that participants in our study might have gradually lost focus on the phase during which the robot actively listened to the student and its corresponding behavior. This tendency could be linked to the fact that participants were presented with surveys and questionnaires immediately after viewing the video, which shifted their attention closer to the moment when they observed the coach’s feedback.

Nevertheless, the outcomes of the study revealed that both the ActiveListenR and PassiveListenR agents were more likeable when compared to the VoiceAgent. This aligns with a previous study conducted by Moore et al. (2022) [84], which investigated various voice user interfaces (VUIs) and potential distraction effects. They found that likeability scores were higher for the socially embodied VUI (the social robot), in contrast to the physically embodied VUI (the smart speaker) [84].

3.3.4 Potential Gender Differences in Perception of Agents

The results of our analysis revealed a notable influence of gender on how the agent was perceived as a coach. Specifically, participants who identified themselves as female rated the agents significantly more positively than those who identified themselves as male in several aspects, including attentiveness, warmth, as well as human nature attributes, such as emotional responsiveness, warmth, and openness.

This aligns with prior research on gender differences in non-verbal communication, which has proposed that females tend to employ non-verbal cues more frequently than males. Additionally, studies have indicated that females are able to assess emotions and personalities based on nonverbal cues more accurately compared to males [55].

However, as indicated in the study [113], males typically perceive robots as being more human-like than females do, which stands in contrast to our study’s findings. Male participants were assigned significantly lower ratings than female participants when it came to

their intention to use the agents and their sense of rapport with them. Additionally, males rated the feedback provided by the agents as less trustworthy compared to females. These findings, especially in the two conditions involving a social robot, appear to be at odds with previous research, which has typically shown that males tend to exhibit a stronger preference for robots than females [105, 113].

It’s important to note that our study only included participants who identified as either male or female, with a small number of participants choosing not to disclose their gender. Consequently, we were unable to make comparisons with gender identities beyond the male and female categories.

It’s worth emphasizing that, despite the variations in ratings between genders, it was specifically the male participants who expressed a higher inclination to use the ActiveListenR agent compared to the PassiveListenR agent. Additionally, males reported a stronger sense of rapport with the ActiveListenR agent in particular.

3.3.5 Participants’ Personality Traits and Differences in Perceptions

An extensive meta-analysis investigating the influence of the Big Five personality traits on robot acceptance revealed that agreeableness and conscientiousness were among the least commonly studied variables [35]. Agreeableness is associated with how warm and friendly an individual tends to be [99]. In contrast, conscientiousness pertains to the extent to which a person demonstrates diligence, meticulousness, and awareness of their actions and their potential outcomes [122].

Our research outcomes highlight that individuals characterized by high levels of agreeableness displayed a notably greater inclination to consider using the agents as presentation rehearsal coaches in the future and reported a stronger sense of rapport with them. Additionally, participants with higher levels of agreeableness perceived the agent as being natural and assigned human nature attributes to it. Prior studies have established positive correlations between agreeableness and robot acceptance [72, 110]. As suggested by Bernotat et al. [10] individuals with higher agreeableness levels tend to have a higher degree of trust. This enhanced trust can subsequently diminish perceived risk and foster greater acceptance of new technologies [98].

In contrast, conscientiousness exhibited a negative correlation with evaluations of the agents as coaches. Participants who scored higher on conscientiousness tended to perceive the agent as less warm and less emotional. They were also less inclined to consider using

the agent in the future for presentation rehearsals and reported feeling less rapport with the agent. Looije et al [77], observed that individuals with higher levels of conscientiousness tended to have a lower liking for social robots. Similarly, Kimoto et al. [72] identified negative correlations between the conscientiousness personality trait and robot acceptance. Additionally, Syrdal et al. High conscientious people were less likely to allow a robot to approach them closely [120]

Our study encountered certain limitations that were brought to our attention by the participants. One such limitation pertained to the quality of the voice employed by the agents, with some participants finding it to be robotic and challenging to comprehend. It's worth mentioning that the Pepper robot's voice was consistently used in all conditions, and we made efforts to test various pitches and frequencies within its capabilities to enhance its adult-like and comprehensible qualities as much as possible.

As mentioned earlier, it is possible that the feedback provided by the robot overshadowed the robot's behaviour during the listening phase. This might have led to participants forgetting how the robot behaved while the student was delivering the presentation, potentially affecting the comparisons between the robot displaying backchanneling and the one that remained motionless during the presentation. To address this limitation in future studies, we could consider separating the evaluation of the robot's behaviour during the listening phase from the feedback phase. This approach would enable us to place greater emphasis on participants' assessments of the robot's listening skills and gain a more comprehensive understanding of how participants perceive the robot throughout the entire interaction.

Online studies have demonstrated the potential to yield results comparable to those obtained in in-person studies [7, 68]. Nevertheless, it's worth noting that our findings may not fully align with outcomes from live human-robot interaction (HRI) scenarios. When participants engage with a social agent in person, they have the opportunity to observe and comprehend a broader range of the agent's behaviors [129]. This includes the ability to better detect backchannels and head nods, which are more discernible when participants are physically present with the robot.

One valuable suggestion raised by certain participants pertains to the format of the feedback delivered by the robot. These participants expressed a preference for receiving written feedback in the form of brief notes, which would facilitate tracking and retaining the information provided. To address this preference, future studies could explore the possibility of offering both verbal and written feedback. This could be achieved through an external monitor or screen, or by utilizing Pepper's tablet.

Chapter 4

Study 2: User Study Evaluating the Social Robot as a Public Speaking Coach

In the second study, we conducted an in-person experiment in which participants engaged in a single presentation session in the presence of the Pepper robot [116]. The primary objective was to determine participants' levels of acceptance, interpersonal closeness, and their perceptions of the Pepper robot as a public speaking coach. Each participant had one in-person session dedicated to delivering a brief presentation and receiving feedback from the Pepper robot.

4.1 Research Questions and Hypotheses

The research questions (RQs) were structured as follows. In this chapter, we present detailed versions of the last two general questions we discussed in Chapter 1.

RQ4-1 What are participants' attitudes about the robot as a public speaking coach particularly in terms of its acceptance and their intentions to use it?

RQ4-2 What level of rapport and interpersonal closeness have participants experienced with the robot as a public speaking coach?

RQ4-3 What are participants' perceptions of the human nature of the robot as a public speaking coach?

RQ4-4 What are participants' overall agreement scores regarding the use of the Pepper robot for public speaking rehearsals at the university?

RQ4-5 How do participants evaluate the relevance of the feedback provided by the robot?

RQ5 What are the views and opinions of a professional public speaking coach regarding the feedback provided by the robot?

The following hypotheses can be proposed to answer the aforementioned research questions, although the research is exploratory.

H4-1 Participants will generally display a positive attitude towards the Pepper robot and its functionalities as a public speaking rehearsal coach. We expect that they will be willing to use the system in the future.

H4-2 Participants will experience a relatively high level of interpersonal closeness with the robot as a coach, as they are unlikely to feel uncomfortable with being judged and have no fear of negative evaluation.

H4-3 Based on the expressiveness potential of the Pepper robot, we expect that people will anthropomorphize and ascribe human nature attributes to it that will distinguish it from machine-like behaviour.

H4-4 RQ4 asks participants for a general evaluation of the robot’s performance as a coach, and we will inquire about it from participants after the study. As such, we expect that if participants assign high scores to the “Almere” model constructs, they will generally concur and perceive the Pepper robot as a useful asset for their presentation rehearsal needs.

H4-5 As the speech processing module, the speech transcription module, and the audience orientation module were implemented, and the Pepper robot gives feedback based on the results of the calculations provided by these modules, we expect that participants will observe relatively high relevance in the robot’s feedback. While this greatly hinges on participants’ personality traits, their previous experiences, and their self-image about their presentation, it is important to acknowledge that some participants may not readily accept the robot’s assessment.

Note, that we do not hold any particular expectations concerning the insights gathered from the evaluation session with the human coach regarding the developed system.

4.2 Methodology

We conducted an in-person user study with university students interacting with the Pepper robot. The session centred around delivering a presentation in the presence of the robot, which assumed the role of a public speaking rehearsal coach. The topic of the presentation



Figure 4.1: The arrangement of pictures of the ‘audience’ on the wall behind the robot’s head and shoulders.

was the “History of Canada Day”, which was adopted from the official website of Immigration, Refugees and Citizenship Canada (canada.ca). The presentation content delved into significant events from various years that led to the establishment and celebration of Canada Day on July 1st.

Throughout the presentation, the student would stand in front of the Pepper robot. The robot did not disrupt the delivery of the presentation; instead, it exhibited active listening behaviour towards the presenter.

Following the completion of their presentations, the robot would offer feedback, which typically lasted for about one minute. Feedback was based on an analysis of participants’ vocal modulation skills and their audience orientation. To enhance the contextual information in the presentation session, we incorporated copyrighted audience images that depicted only the upper bodies of people (<https://www.freepik.com/>). Figure 4.1 displays the arrangement of pictures depicting a simulated audience around the robot’s head and shoulders. The space between the first row of pictures was reserved for the robot to stand, and the arrangement was designed to simulate a room of audience members with multiple rows.

Our program analyzes participants’ quality of speech and speech prosodies, as well as their audience orientation during their presentation, to provide participants with feedback. Processing participants’ speech quality was done using the Praat software [101]. Praat is a free computer software for phonetic analysis of speech. The software was designed by Paul Boersma and David Weenink of the University of Amsterdam and is still being developed and updated [101]. We used a Python library to use Praat functionalities in Python which

is called Parselmouth.

The study’s objective encompasses both asking for participants’ evaluations about the rehearsal session with the Pepper robot as well as seeking an expert evaluation. Participants answered questionnaires before and after the presentation (as illustrated in [subsection 4.3.2](#)). Regarding the expert evaluation, upon the completion of the study, we randomly selected nine recorded video sessions of participants’ presentation rehearsals (as depicted in [section 4.3.6](#)). We then invited a professional public speaking coach to the lab, where, through a feedback session, we sought their insights concerning the quality of feedback provided by the robot. Specifically, we inquired about aspects that were effectively addressed and those that might need improvement in the future. Additionally, we requested the coach to assess the robot’s feedback considering multiple metrics.

4.3 Experiment

4.3.1 System Implementation

To analyze the acoustic and prosodic aspects of speech, we took into account several factors, including articulation rate (speech rate excluding silent pauses), voice volume, pitch variation, and filler rate.

Articulation Rate: In the phonetics literature, the terms “speech rate”, and “articulation rate” are often used interchangeably, while they convey different meanings. Even though both indicate “the number of output units per unit of time”, the speech rate is inclusive of pause intervals while the articulation rate is not [\[67\]](#).

The articulation rate specifically governs the pace at which speech segments are pronounced, disregarding speaker-specific nuances in conveying information, such as hesitations, pauses, emotional expressions, and so forth [\[67\]](#). Conversely, speech or speaking rate encompasses broader speaker attributes, including the frequency of pauses and the use of laughter or fillers like “you know” or “I mean.” These elements, when inserted into the speech, disrupt fluency and define a unique communication style specific to the speaker [\[67\]](#).

In this study, measuring only articulation rate will provide us with a more accurate assessment of cross-dialectal disparities in speech tempo, and it eliminates the pauses as an extra within-speaker variable [\[104\]](#).

To calculate the articulation rate using Parselmouth and Praat software functionalities, we employed a method based on detecting syllable nuclei [\[28\]](#). We utilized an open-source

script developed for the Praat software, which is designed to identify syllables in continuous speech [28]. Syllable nuclei, which are peaks preceded and followed by dips in intensity, are the center of a syllable. The algorithm uses the intensity contour of the voice signal to find peaks and locate the occurrences of these nuclei.

A threshold of -25 dB or -20 dB is used as the silence threshold, and any segment of the intensity signal with intensity below this threshold is considered unvoiced and is excluded from the analysis to only remain with the voiced segments. In the original paper, potential syllables are defined as those segments that are at least 2 dB above the median of the intensity [28]. Parselmouth proposes a modified version based on the 0.99 quantile of the intensity minus the silence threshold rather than its median [38]. Among potential syllables, those that are preceded by at least 2dB dips in intensity are counted as syllable nuclei [38]. To consider the articulation rate, we calculate the number of syllable nuclei per phonation time (excluding pauses).

Pitch Variety: The pitch contour is derived through autocorrelation [15], utilizing default values in Praat, with a minimum pitch of 75 hertz and a maximum pitch of 500 hertz. To assess pitch variety, we computed the difference between the 95th quantile and the 5th quantile of the pitch contour, expressed in hertz, and reported this difference as the pitch variation.

Voice Loudness: As a method of measuring voice volume, we used the intensity contour and compared the mean intensity for the entire presentation time with some ground truth thresholds that we determined based on testing the appropriate sound volume for the experiment room area. A minimum appropriate loudness of 52 dB and a maximum appropriate loudness of 70 dB were determined empirically in the experiment room and remained constant throughout the whole experiment. Therefore less than 52 dB was considered as “quiet”, and more than 70 dB was considered as “loud”.

Filler Rate: We identified filler words (such as “um,” and “uh”) by employing the speech disfluencies detection module within the AssemblyAI API. AssemblyAI utilizes speech-to-text transcription models to transcribe speech data. We record the presenter’s voice using Python’s Pyaudio package and subsequently upload the local audio file to the API for processing using the API token and an HTTP request. The response is returned in a JSON format, containing the transcribed data as well as meta-data (information about the data). One can set specific attributes for the JSON output to retrieve the information they need about the data.

We set the disfluencies parameter to “true” to maintain them in the transcript when submitting the audio file for processing. Having determined the number of filler words in the transcribed data, we divide it by phonation time (speaking time without pauses) in

minutes to compute the filler rate.

It is important to note that the AssemblyAI API automatically deletes uploaded files from their servers either once the transcription process has concluded or after 24 hours following the initial file upload, and it is no longer stored on their servers.

Audience Orientation: The purpose of audience orientation is achieved by estimating the head pose of participants during their presentation. To this aim, we attached a webcam on the top of the Pepper robot’s tablet to have a front view of the participant’s upper body. The head pose estimation algorithm uses the Perspective-n-Point(PnP) solution to estimate the position of the head from its two-dimensional image. PnP solution results in a translational and rotational matrix. At each time, the head pose estimation algorithm returns a three-dimensional angle.

The angle with respect to the y-axis shows the yaw head movement can be a good estimation of the presenter’s audience orientation at each time. The amount of time that the participant’s head orientation, or specifically their head yaw, is less than a threshold from where the robot stands is calculated. Through tests, we decided to choose 13° for the head orientation threshold. Therefore, we calculated the time during which the participant’s head yaw was within -13 to 13 degrees of the robot’s position during the whole presentation.

Listening Behaviour, and Robot’s Engagement: Pepper was programmed to exhibit head nodding as a response to the participant’s presentation, conveying active listening. The robot would nod every six seconds while the head motion behaviour was designed in two distinct modes, randomly alternating between them.

The first mode involved a relatively slow vertical head movement with a small range of motion, mimicking a subtle nod. The second mode entailed two consecutive head movements at a relatively faster pace, with the second motion having a smaller range compared to the first. This design aimed to make the nodding gesture appear as natural as possible, similar to human nodding behaviour, and to prevent participants from perceiving a repetitive pattern in the robot’s head. Additionally, to demonstrate attentiveness and focus, the robot would lean forward every forty seconds.

Furthermore, Pepper’s face-tracking module was activated, allowing it to track participants’ faces throughout their presentations, enhancing interaction and engagement.

Feedback Generation: All of the parameters described above are being calculated as the participants deliver their presentation, and when the presentation is over, the results of the calculations will be compared to the ground truth values that are supported by research to ensure that the presentation is delivered effectively [124]. Depending on whether

Table 4.1: Speech quality thresholds - These thresholds (except for the voice volume) are directly adopted from [124]

Metrics	Range		
Speech rate (syl/s)	[0,3] slow	(3,5) good	[5,∞) fast
Pitch Variety (Hz)	[0,120) monotonous	[120,∞) good	
Voice Volume (dB)	[0,52) quiet	[52,70] good	(72,∞) loud
Audience Orientation	[0,0.4) low	[0.4,∞) good	
Filler rate (Fillers/minute)	[0,5) low	[5,15) moderate	[15,∞) high

the value of each parameter falls within the acceptable range, the robot gives positive or negative feedback. Table 4.1 shows the acceptable ranges we considered for each parameter that is adopted from a previous study [124].

We used the well-known feedback sandwich technique [102] to make the robot’s feedback coherent for all participants. The feedback sandwich is interpreted as embedding corrective feedback within two general positive statements. It is shown that it has a positive effect on the performance of the under-studied population. The effectiveness of the feedback sandwich is based on the assumption that people accept negative feedback better if it is followed by a compliment, and the combination of both positive and negative feedback tends to reduce the stress and discomfort in accepting criticism [102].

Therefore in our study, the feedback is structured as follows:

1. A brief opening statement that expresses appreciation for how interesting the presentation was.
2. Positive comments indicating encouragement regarding aspects of the presentation that were successfully conducted.
3. Comments regarding negative aspects of the presentation, followed by suggestions for improvement.
4. A final expression of appreciation for the presenter and hope that the rehearsal was fruitful.

In addition, we created written versions of the feedback that corresponded to each positive or negative feedback. These written notes were displayed on Pepper’s tablet during the feedback session according to the time of the specific feedback. This approach allows participants to have a concise visual representation of the robot’s feedback about their presentation, aiding in better recall and understanding. Positive feedback was displayed on a green-coloured screen, while negative feedback was presented on a red-coloured screen.

4.3.2 Procedure and Measures

To recruit participants for the study, we spread the study posters throughout the university campus. Students who were interested in participating were directed to an online booking platform known as “Calendly” to select their preferred time slots from the available options. Calendly allows event organizers to list their available time slots, and participants can conveniently choose the time that suits them best. The system automatically manages overlaps and ensures that selected time slots are no longer available. A calendar invitation email and a reminder email were configured to be sent to participants. As part of the experiment and afterward, each participant would receive a unique code that would be used as their identification.

The researcher prepared the presentation slides and presentation notes in advance. To keep the presentation brief, we kept the content to less than one page and had only four slides. Out of these, two slides contained informative data, while the remaining two served as the introduction and appreciation slides.

To maintain an appropriate level of complexity and challenge, we deliberately chose a topic that incorporated historical and political terminologies, such as “The British North-America Act,” “Constitution Act”, and “Anniversary of Confederation.” Nevertheless, we aimed to minimize the cognitive load on the presenter and to achieve this, we simplified the text content and crafted presentation notes to be engaging, and less boring. The presentation slides and presentation notes would be shared with the participants about two or three days before their session, and we would ask them to review these materials before taking part in the experiment. Additionally, participants would be asked to fill out the consent form, read the information sheet, and fill out a pre-study questionnaire.

The experiment room was designed to be similar to a rehearsal session for a presentation. A podium was provided for the presenter, and a laptop was set up on the podium with the presentation slides displayed on it. The slides were projected on the wall behind the presenter using a video projector. The Pepper robot was placed in front of the podium approximately one, and a half meters from where we asked participants to stand, and pictures depicting audience members were attached to the wall behind the Pepper robot, in a way that they would not be occupied by the robot body and would almost surround its head and shoulders.

Before participants entered the experiment room, the researcher would present the slides on the laptop located on the podium. The slides were identical to what had been sent to participants. A video projector was connected to the laptop to project the slides

on the wall behind the presenter. The podium was placed at an angle relative to where the presenter would stand so that whenever the presenter looked at the laptop during their presentation, their head yaw angle exceeded the accepted range, indicating that they exited from the orienting towards the audience. Participants may have also exited from orienting towards the audience due to looking at the projected slides on the wall behind them.

In the experiment room, another laptop with post-study questionnaires was placed on a separate desk, so that participants could answer the questions after the experiment.

One camera was positioned only in front of the presenter and the other was positioned on one side of both the presenter and the robot to capture the entire experiment room. The researcher would start recording on both cameras just before the participant entered the experiment room. [Figure 4.2](#) shows the experiment setup from the point of view of both cameras.



Figure 4.2: Experimental Setup

As participants entered the experiment room, they signed the consent form in hard copy and were provided with some instructions regarding the experiment. They were

provided with a wireless microphone to attach to themselves. The researcher instructed participants not to move the podium or laptop during their presentations to maintain consistency. Additionally, the researcher ensured that the webcam captured the upper body of participants effectively as part of the setup process. Finally, the researcher would ask participants to begin their presentation.

The presentation would last approximately two to three minutes, after which the robot would thank participants and ask for a brief period to review their presentation. In part, this was due to the Assembly AI speech-to-text transcription for the analysis of filler words, and the fact that the recorded audio file had to be uploaded to the Assembly AI API. Nevertheless, this waiting time was less than one minute, depending on the duration of the participants' presentations, and for most participants, this waiting time was around twenty seconds. The robot would then start giving feedback to the participant, which would last around two minutes followed by the final appreciation.

Finally, the Pepper robot would ask participants to answer post-study questions using the laptop on the desk. This arrangement was primarily designed to minimize interaction between the experimenter and participants. Consequently, participants had little or no interactions with the experimenter throughout the experiment, until the conclusion of the post-experimental phase.

4.3.3 Pre-study Questionnaire

The pre-study and post-study questionnaires were generated using Qualtrics [103], a versatile platform for creating online surveys, managing responses, and analyzing data. The pre-study questionnaire covered various aspects, including demographics, participants' experience with robot programming, and their prior interactions with robots.

Participants were asked whether English was their second language (ESL) or if they were native English speakers. Additionally, they were asked to specify the total number of languages they could speak. The questionnaire also inquired about the number of presentations participants had delivered in the past two years, whether they had taken any public speaking classes, and whether they felt the need for such classes. For more detailed information please visit Appendix C.1. Following that, participants were asked the following two questionnaires (Please note that participants were given the option to not answer any of the pre-study questions they wished.)

Personal Report of Communication Apprehension, Public Speaking Subscale (PRCA-PS): Six items are included in the public speaking subscale of PRCA. The items are rated on a 5-point Likert scale from 1 (strongly agree) to 5 (strongly disagree).

There is a minimum score of 6 and a maximum score of 30, the higher the score, the more fear one indicates toward giving a public speech. Scores between 13.75 and 20.75 indicate moderate levels of anxiety, while scores between 20.75 and 30 indicate high levels of anxiety. Literature has demonstrated good validity and reliability of these scores, as indexed by Cronbach’s alpha which is used to show an internal consistency [81]. For more detailed information please visit Appendix C.2.

Ten-Item Personality Inventory questionnaire (TIPI): A brief measure of assessing the Big-Five Personality dimensions scales using ten items, which are extraversion, agreeableness, conscientiousness, emotional stability, and openness to experiences with only two items per scale. It consists of ten items on a 7-point Likert scale from 1 (Disagree strongly) to 7 (Agree strongly). For further details about question items please refer to Appendix C.3.

4.3.4 Post-study Questionnaires:

Here we present the post-study questionnaires that we asked participants to answer.

Acceptance and Intention to Use: The Almere model is an extension of the Unified Theory of Acceptance and Use of Technology (UTAUT) questionnaire, originally developed to evaluate the acceptance of assistive social robots among elderly individuals [57]. The study done with older adults regarding the acceptance of assistive social robots showed the Almere constructs to be solid to use in different settings [57]. We made slight adaptations to certain phrases to align them with the context of giving presentations. However, during the rephrasing process, we aimed to maintain as close a connection to the original sentences as possible. The questionnaire is shown in Appendix C.4.

The original model encompasses 11 constructs: Anxiety, Attitude towards technology, Facilitating Conditions, Intention to Use, Perceived Adaptiveness, Perceived Enjoyment, Perceived Ease of Use, Perceived Sociability, Perceived Usefulness, Social Influence, Social Presence, Trust, and Use. Figure 4.3 illustrates the interrelationships among these constructs, specifically, it has been shown that Attitude towards technology is a crucial factor in determining the Intention to Use.

Perceived Closeness: The Inclusion of Other in the Self (IOS) Scale is a tool that prompts participants to choose an image that most accurately reflects their level of connection and interpersonal closeness with another individual or a group. This scale comprises seven Venn-like diagrams, each depicting varying degrees of overlapping circles in a visual format [4]. In our specific context, we modified this scale and asked participants to evaluate their interpersonal closeness with the robot.

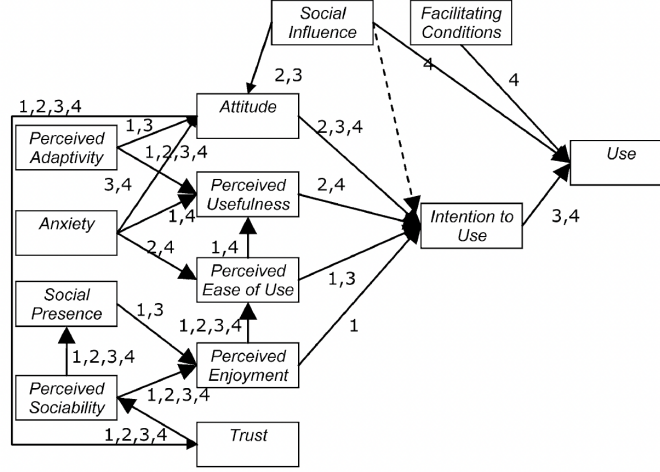


Figure 4.3: Figure taken from [57] illustrates the interrelation between the Almere model constructs. The numbers show the experiment in which the interrelation between the constructs was confirmed by regression scores. However, no regression analysis confirmed the dotted lines.

It has been evidenced that the IOS scale is a psychometric and substantive suitable measure of closeness, that can be done quickly and can be applied to various circumstances [4]. The questionnaire is shown in Appendix C.5.

Perceived Human Nature: The characteristics of human nature can be summed up as emotional responsiveness, interpersonal warmth, cognitive openness, agency, and depth. When these traits are denied, it is referred to as mechanistic dehumanization, which causes a person or social group to be perceived as inert, cold, rigid, passive, and superficial [56].

In our study, we used the same attributes to evaluate the HN dimensions of the Pepper robot. Participants were asked to rate the Pepper robot on these HN dimensions using a continuous scale, with values ranging from 0 to 1000. The questionnaire is shown in Appendix B.4.

Perceived Listening Behaviour: To assess the robot’s active listening behaviour during participants’ presentations, we designed a set of three questions for participants to answer. Understanding participants’ perceptions of the robot’s capacity for active listening was crucial, as the robot serves as a coach for students. This role necessitates a certain level of comprehension and human-like behaviour, rather than merely mechanistically calculating parameters. Moreover, participants tended to feel as if they were in a relatively realistic scenario when they perceived that the agent was actively and explicitly engaged in listening to them. As there was no established standard questionnaire for evaluating the

Table 4.2: Questions on perceived active listening behaviour - These questions were posed to participants to gauge their perception of the robot’s active listening behaviour.

Questions	Response scale
While I was presenting, the robot was:	(Not listening at all, Very listening)
While I was presenting, the robot was:	(Not attentive at all, Very attentive)
While I was presenting, the robot was:	(Not engaged at all, Very engaged)

Table 4.3: Participant Inquiries on Robot Evaluation and Feedback Assessment

Questions	Response scale
Evaluation of the Robot	
I found the robot as a public speaking coach:	(Not competent at all, Very competent)
I found the robot as a public speaking coach:	(Not helpful at all, Very helpful)
How do you feel about having robots as public speaking coaches?	(Strongly disagree, Strongly agree)
Feedback Evaluation	
How well did you find the feedback provided by the coach?	(Not understandable at all, Very understandable)
How relevant was the robot’s feedback to you?	((Not relevant at all, Very relevant))

listening behaviour of the robot, we utilized the questions presented in [Table 4.2](#), which participants answered on a continuous scale. The questionnaire is shown in [Appendix C.6](#).

General Evaluation of the Robot’s Behaviour and Its Feedback: Lastly, participants were asked to provide a comprehensive assessment of the Pepper robot’s performance as a public speaking coach. This evaluation encompassed inquiries, including the robot’s competence, and its feedback accuracy. Also, we asked about the degree to which participants agreed to have the Pepper robot for public speaking training. To evaluate the quality of the robot’s feedback, we specifically targeted feedback comprehensibility and relevance. The responses to all of these questions were provided on a continuous scale. The questions and the responses are shown in [Table 4.3](#) and [Appendix C.7](#).

The Negative Attitudes towards Robots Scale : The Negative Attitudes Towards Robots Scale (NARS) is specifically designed to measure anxiety towards robots [\[91\]](#). The term robot anxiety refers to a fear or negative emotion that prevents someone from interacting with robots in a dyadic manner. There are 14 items in this questionnaire, divided into three subscales: S1: Negative Attitude Toward Situations of Interaction with Robots (6 items), S2: Negative Attitude Toward Social Influence of Robots (5 items), and S3: Negative Attitude Toward Emotions in Interaction with Robots. Responses are based on a Likert scale from “Strongly disagree” to “Strongly agree”.

Since two of the subscales of the Negative Attitudes towards Robots Scale (NARS) are phrased in a negative tone and one in a positive tone, they can potentially introduce implicit biases in participants before they interact with the robot. To minimize this potential bias, we positioned the NARS questionnaire as the last one in our survey. Please note that we primarily collected data from the NARS to explore potential correlations between the NARS subscales and other metrics in the post-experimental phase. The questionnaire is shown in Appendix C.8.

Expert Evaluation: Upon completion of the user study, we invited a public speaking coach to assess the relevance and accuracy of the robot feedback from an expert perspective. To facilitate this evaluation, we selected nine videos from a collection of fifty, which included participants delivering presentations in front of the robot, followed by the robot’s feedback. These video selections were made randomly.

I invited the human coach to our laboratory, where we jointly viewed the videos on a television screen. To ensure optimal audio clarity and eliminate background noise or echoes, the human coach was provided with a headphone during the viewing. Each video had a duration of approximately five minutes.

Following the viewing of each video, I posed three questions to the human coach. The first two questions were displayed on a tablet, allowing the human coach to select his answers, while the last question was asked verbally, and the response was recorded.

1) “What do you think about the robot’s feedback, given the presenter’s performance?” The coach provided a rating on a continuous scale ranging from “Very Irrelevant” (0) to “Very Relevant” (1000).

2) “What do you think about the content of the feedback provided by the robot?” The coach’s response was indicated on a continuous scale from “Not instructive at all” (0) to “Very Instructive” (1000).

These questions aimed to gauge the human coach’s perspective on the relevance of the robot’s feedback in relation to the presenter’s performance and the learning value of the feedback content.

The third question was posed after each video as an open-ended question, and the coach’s response was audio-taped: 3) “What specific aspects of the feedback could be improved?”

Following watching all nine videos, we posed two questions to the human coach on the tablet, aimed at obtaining an overall assessment of the robot’s performance.

1) “How would you evaluate the performance of the robot as a coach?” The response

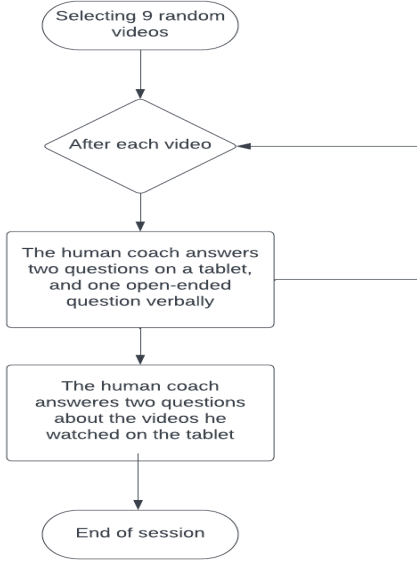


Figure 4.4: The diagram shows the procedure we followed for the session with the human coach

was recorded on a continuous scale, spanning from “very unhelpful” (0) to “very helpful” (1000).

2) “Do you think the robot is useful to be used in the University of Waterloo for presentation rehearsals?” The response to this question was rated on a continuous scale, ranging from “Not at all” (0) to “very much” (1000). [Figure 4.4](#) depicts the procedure we followed as well as the order of questions we asked the human coach.

4.3.5 Participants

We recruited a total of 50 university students for our study. The participants had a mean age of 20.82 years. Among them, 23 were undergraduate students, 21 were master’s students, 4 were Ph.D. students, and 2 preferred not to disclose their academic status. In terms of gender distribution, 30 participants identified as female, 18 as male, and 2 preferred not to disclose their gender.

Among the participants, 33 reported having previous interactions with robots, while 17 had no prior experience with robots. Of those who had previous interactions, 19 had interacted with human-like robots, 2 with animal-like robots, 19 with robots that were

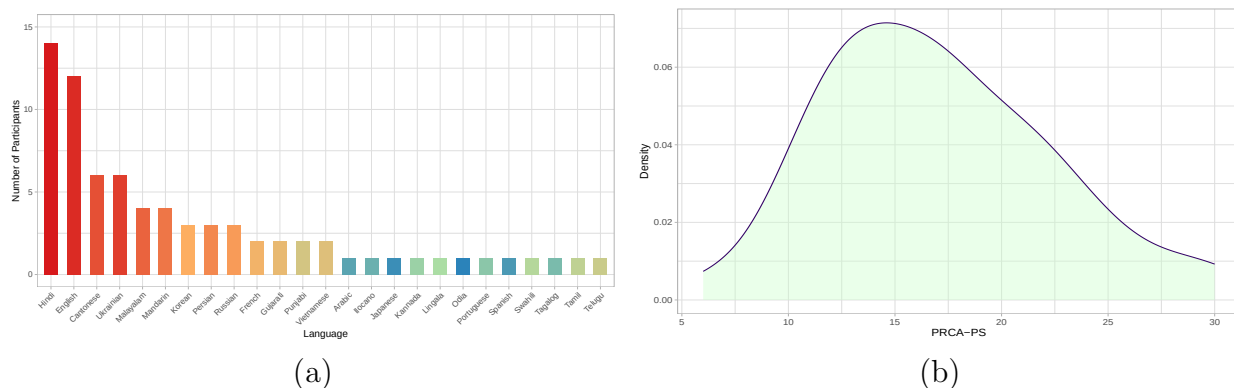


Figure 4.5: (a) Distribution of home languages spoken by participants, showing the number of participants for each language. (b) Probability density distribution plot of participants’ PRCA-PS scores

neither human-like nor animal-like, and 4 with other types of robots, such as lower-limb exoskeletons and serial and parallel manipulators.

A total of 11 participants have attended prior classes concerning public speaking skills, while a total of 38 have not. Also, 28 participants expressed a need to attend public speaking classes and rehearse presentations, while 20 did not, and 2 were unsure.

Regarding language proficiency, 36 participants spoke English as a second language, while 14 spoke it as their first language. 11 participants had attended previous public speaking classes, while 38 had not (one participant did not provide an answer). The participants in the study came from various countries and had a range of languages spoken at home, with a significant number of them being multilingual (Figure 4.5 (a) displays the distribution of the languages participants reported as their primary spoken languages at home).

Participants also reported their public speaking anxiety with the PRCA-PS questionnaire (Figure 4.5 (b) shows the probability density function of the PRCA-PS scores).

The study received ethics clearance from the University of Waterloo Human Research Ethics Board.

4.3.6 Results

In this section, we provide an overview of the general findings from a single-session presentation rehearsal with the Pepper robot.

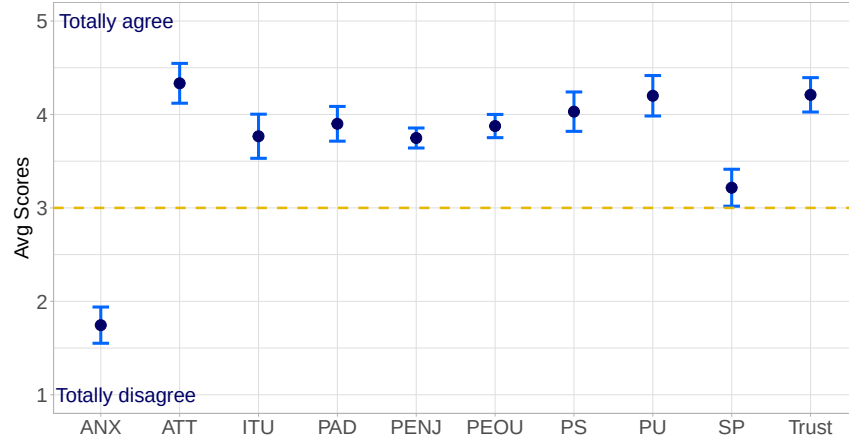


Figure 4.6: Almere model constructs scores. Dark blue points show the mean score for each construct and error bars indicate 95% confidence intervals. The yellow line represents the neutral choice.

Acceptance and Intention to Use: Among 13 constructs that were defined in the Almere model, we used 10 constructs, as the other 3 were not deemed very relevant to our study (Anxiety (ANX), Attitude towards technology (ATT), Intention to Use (ITU), Perceived Adaptiveness (PAD), Perceive Enjoyment (PENJ), Perceived Ease of Use (PEOU), Perceived Sociability (PS), Perceived Usefulness (PU), Social Presence (SP), Trust). Also, based on the model description, some of the constructs have a significant role in determining the Intention to Use and Actual Use, which are Perceived Usefulness, Perceived ease of Use, Perceived Enjoyment, Trust, Attitude towards technology, Social Influence and Facilitating Conditions 4.3 (The last two constructs were not used in our analysis). Anxiety, Perceived Sociability, Perceived Social Presence and Perceived Adaptiveness are believed not to directly determine the Intention to Use.

The original reference includes some question items associated with each construct. The responses were on a 5-point Likert scale from “Totally disagree” to “Totally agree”. The items were presented in a random order to participants. Figure 4.6 shows the results of the score for each of the constructs.

According to Figure 4.6, the mean score and the standard deviation of anxiety towards the robot during the interaction was ($M = 1.74, sd = .68$). This indicates that participants were not intimidated by using the robot. The results also show the scores for Attitude towards the agent ($M = 4.33, sd = .75$), Perceived Sociability ($M = 4.03, sd = .74$), Perceived Usefulness ($M = 4.2, sd = .76$), and Trust ($M = 4.21, sd = .65$) were all

relatively high and above 4. The average scores for Intention to Use ($M = 3.76, sd = .83$), Perceived Adaptiveness, ($M = 3.9, sd = .65$), Perceived Enjoyment ($M = 3.75, sd = .38$), Perceived Ease of Use ($M = 3.88, sd = .44$) were above 3.5 and close to 4. The Perceived Social Presence was the only construct that was reported lower than other constructs ($M = 3.22, sd = .69$).

Based on the numbers reported above and [Figure 4.6](#) H4-1 is supported, as participants generally gave a high score towards the ATT, PU, PS, and Trust (above 4), PENJ, and PEOU were close to 4 indicating a relatively high score. In addition, the constructs of ATT, PU, PEOU, PENJ, and Trust are directly related to intent to use, as outlined by [Figure 4.3](#).

Post-hoc analysis During the posthoc analysis and applying linear models, we took into account 43 participants since the records for 7 participants for the PRCA-PS questionnaire were incomplete.

In order to examine the confounding effects of NARS and PRCA-PS scores, TIPI, previous interactions with robots, previous participation in public speaking classes, and level of preparedness on participants' ratings of the robot, we applied regression analysis using linear models. We confirmed the correctness of the use of linear models via tests such as the normality of residuals. The factors included in the linear models were selected in conjunction with the minimizing multicollinearity and minimization of the Akaike Information Criterion (AIC).

According to the results, the anxiety towards the robot ($se = .006, t = 4.324, p = < .001$) was positively correlated with the NARS score, and the attitude towards the robot ($se = .032, t = -3.042, p = < .01$) was negatively correlated with the NARS score.

The intention to use the system in the future for presentation rehearsals was negatively correlated with NARS scores ($se = .036, t = -2.862, p = < .01$), previous interactions with robots ($se = .591, t = -2.99, p = < .01$), PRCA-PS scores ($se = .053, t = -2.693, p = < .05$), and previous attendance in public speaking skill classes ($se = .711, t = -2.548, p = < .05$). Participants' perceived usefulness ($se = .055, t = -2.721, p < .01$), and the trust towards robot's feedback ($se = .033, t = -2.231, p < .05$) were negatively correlated with the PRCA-PS score .

Participant's perceived enjoyment displayed positive correlations with participants' level of preparedness ($se = 0.0009, t = 3.753, p < .001$).

Perceived Interpersonal Closeness: Respondents were tasked with choosing a pair of circles from a selection of seven, each denoting varying degrees of overlap. The scoring system in the main reference introduces a numerical score according to the amount of overlap: 1 = no overlap; 2 = little overlap; 3 = some overlap; 4 = equal overlap; 5 = strong overlap; 6 = very strong overlap; 7 = most overlap.

The results of the IOS scale showed the mean score of $M = 3.42$, and the standard deviation of $sd = 1.51$ across all participants. Table 4.4 shows the number of participants for each diagram. As a result of analyzing the IOS scale, H4-2 is partially supported by the available data, as 70% of the participants were selected from “Some Overlap” to “Most overlap”.

Table 4.4: Number of Participants Choosing Different Diagrams in the Inclusion of Other in the Self (IOS) Scale

Diagram	Frequency
No overlap	5 (10%)
Little overlap	10 (20%)
Some overlap	11 (22%)
Equal overlap	14 (28%)
Strong overlap	4 (8%)
Very strong overlap	5 (10%)
Most overlap	1 (2%)

Perceived Human Nature Attributes: The results of the human nature attributes of the Pepper robot are shown in Figure 4.7. As depicted, the Pepper robot’s emotional responsiveness was perceived as relatively lower than other human nature characteristics ($M = 528.64, sd = 275.22$). In contrast, the Pepper robot was perceived as warm and friendly ($M = 716.84, sd = 234.79$). Additionally, the ratings for perceived openness were close to the ratings for warmth ($M = 685.8, sd = 261.79$). Notably, the attribute of agency or liveliness received the highest ratings among all human nature attributes ($M = 742.66, sd = 229.38$).

The findings provide support for H4-3, as the robot’s scores on the dimensions of Human Nature attributes leaned more towards the human-like end rather than the machine-like end, and most of these attributes garnered an average score exceeding 500, which represents the midpoint of the scale. Specifically, perceived Agency, Warmth, and Openness were rated around 700, while perceived Emotional Responsiveness received a score surpassing 500.

Listening Behaviour Evaluation: We asked participants questions to find out if they perceived active listening behaviour in the Pepper robot. The results of the participants’ scores are summarized in Figure 4.8.

The mean score and the error bars show that participants did perceive active listening behaviour in the robot. The scores for the first question regarding the robot being listening ($M = 829.98, sd = 154.841$), being attentive ($M = 837.6, sd = 156.58$), and being

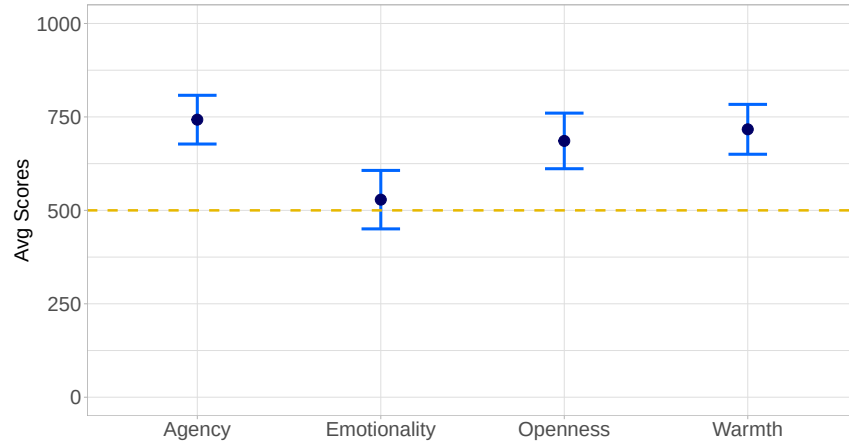


Figure 4.7: Human nature characteristics attributed to the Pepper robot. Dark blue points show the mean score for each attribute and error bars indicate 95% confidence intervals. The yellow line shows the midpoint on the slider.

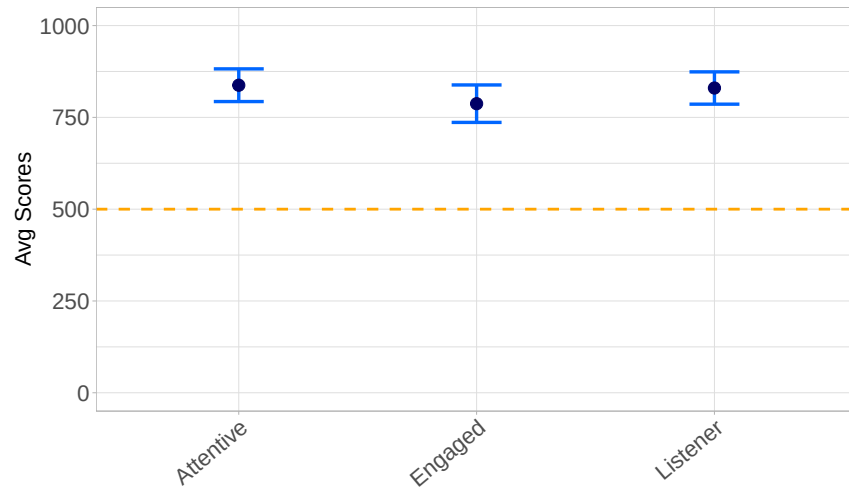


Figure 4.8: Perceptions of active listening on the robot. Dark blue points show the mean score for each item and error bars indicate 95% confidence intervals. The orange line shows the midpoint on the slider.

engaged ($M = 787.32, sd = 179.435$) were relatively high. Shows the impact of non-verbal backchannelling in the perception of active listening.

Post-hoc analysis

The results of the linear regression analysis revealed significant correlations between perceptions of the robot's listening behaviour and the assessment of Human Nature attributes in the robot. According to the results perceived non-verbal listening behaviour in the robot has significant positive correlations with perceived emotional responsiveness ($se = 0.227, t = 3.232, p < .01$), perceived warmth ($se = 0.19, t = 4.294, p < .001$), perceived openness ($se = .209, t = 4.717, p < .001$), and perceived agency ($se = .22, t = 2.439, p < .05$) of the robot.

Also, according to the results of the linear regression participants' scores in the IOS questionnaire are positively correlated with perceived listening behaviour in the robot Robot ($se = .001, t = 2.061, p < .05$).

General Perception of the Robot as a Public Speaking Coach: A general evaluation of the robot as a public speaking rehearsal coach is shown in [Figure 4.9](#). We assessed the participants' perceptions of the robot's competence, its perceived helpfulness, and the extent to which participants agreed to the application of the Pepper robot as a public speaking rehearsal coach. Feedback was evaluated based on its relevancy to the participants' own performance and the level of comprehensibility. Responses were given on a continuous scale. According to the results the mean score and standard deviation for perceived competence were $M = 818.02, sd = 201.92$, perceived helpfulness $M = 912.16, sd = 126.21$, overall agreement score on using the system $M = 834.2, sd = 178.2$, perceived feedback comprehensibility $M = 827.5, sd = 193.31$ and perceived feedback relevancy $M = 855.56, sd = 165.12$.

These findings provide support for H4-4 and H4-5, as participants assigned high ratings to the Pepper robot, both in terms of their general evaluations and the relevance of the robot's feedback. Additionally, participants expressed significant agreement concerning the system's potential as a coach for public speaking rehearsals, perceiving it as competent and highly useful.

Post-hoc Analysis: The scores of competence are significantly correlated with the feedback relevancy scores ($se = .106, t = 6.99, p < .001$), and perception of attentiveness in the robot ($se = .112, t = 2.51, p < .05$). Similarly, the usefulness of the system was significantly correlated with the feedback relevancy ($se = .113, t = 3.374, p < .01$), and has a significant negative correlation with the NARS scores ($se = 2.319, t = -2.24, p < .05$).

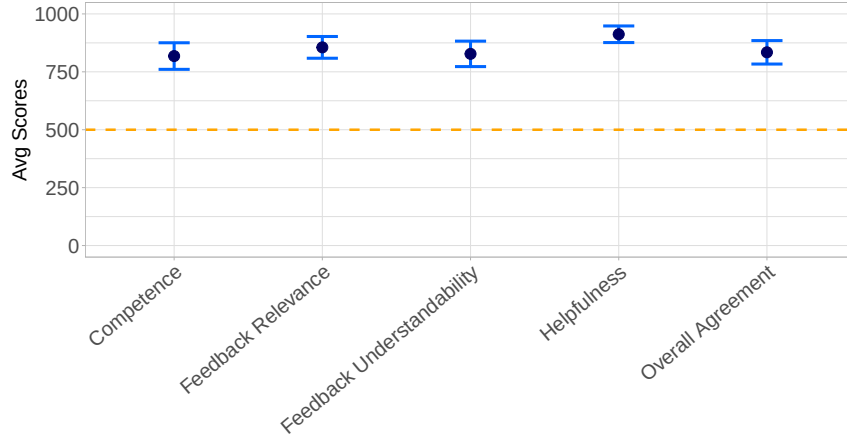


Figure 4.9: General perceptions of the robot as a public speaking rehearsal coach and the feedback provided by the robot. Dark blue points show the mean score of each item and error bars indicate 95 % confidence intervals. The yellow line indicates the neutral choice.

Furthermore, the agreement scores for using the robot as a public speaking rehearsal coach at the university have significant positive correlations with perceived feedback relevancy ($se = .099, t = 7.459, p < .001$), and significant negative correlations with NARS ($se = 2.036, t = -2.436, p < .05$)

Expert Evaluation: To assess the accuracy and relevancy of the feedback provided by the robot and in response to the RQ5, we sought the assistance of a human coach who specialized in public speaking.

Figure 4.10 shows the expert’s average scores for the nine videos, specifically focusing on feedback relevance and instructional quality of the feedback. As per the data, the scores for feedback relevance are ($M = 843.22, sd = 107.07$), while the scores for the feedback being instructive are ($M = 842.11, sd = 92.76$). These scores indicate a substantial level of agreement between the human coach and the feedback provided by the robot. It suggests that, according to the expert, the robot’s feedback was deemed highly accurate and relevant concerning the presenter’s performance and, importantly, that it was instructive and beneficial for the presenter’s rehearsal sessions with the coach.

We conducted a qualitative analysis of the responses to the open-ended question. The human coach’s responses recorded after each video for improving the robot’s feedback were transcribed into a textual format using the AssemblyAI speech-to-text transcription API, and then we proceeded with the qualitative analysis of the provided text. Through this analysis, we were able to uncover valuable insights regarding the robot’s performance and

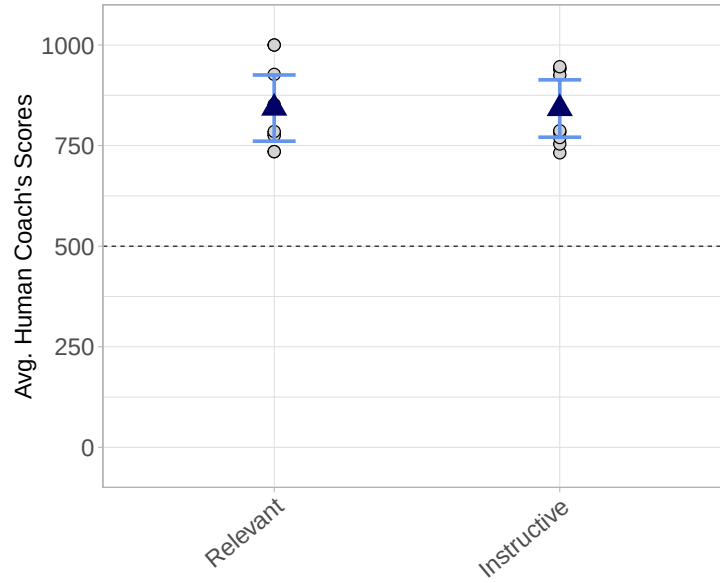


Figure 4.10: Perceptions of the human coach regarding the feedback relevancy and feedback instruction quality provided by the robot as a coach. The dark blue triangle shows the average human coach’s ratings over nine videos, and the error bars indicate 95% confidence intervals.

identify areas where improvements could be made in terms of feedback, as per the expert human coach’s perspective.

The results of the final two questions aimed at obtaining an overall assessment of the robot’s performance are as follows.

1) “How would you evaluate the performance of the robot as a coach?” Responses were recorded on a continuous scale, spanning from “very unhelpful” (0) to “very helpful” (1000). The score provided by the human coach to the question was **918**, indicating his overall positive evaluation of the robot’s performance as a coach.

2) “Do you think the robot is useful to be used in the University of Waterloo for presentation rehearsals?” Responses to this question were also rated on a continuous scale, ranging from “Not at all” (0) to “very much” (1000). This question was answered by the human coach with a score of **838**, demonstrating his general support for having an agent at the university to assist with presentation rehearsals.

Table 4.5: The results of the qualitative analysis of the open-ended question after each video in the feedback session with the human expert, where points of agreement and disagreement were identified

Participant1	
Points of Agreement	Robot detected clear volume, it was louder than speaking voice Robot detected suitable speech rate
Points of Disagreement	Fillers may not be associated with anxiety, and there were not too many fillers There were some indications of nervousness in his body movement, but not in his voice The presenter couldn't sustain eye contact really well
Participant2	
Points of Agreement	The robot could provide a good feedback on fillers The robot mentioned confidence and preparation for lack of fillers I agree as they are key elements for reducing filler words The robot is particularly good at evaluating speech rate The robot suggested breaking the sentences into shorter phrases to improve speech rate
Points of Disagreement	The only thing I would tweak is to suggest the idea of pausing more
Participant3	
Points of Agreement	I would definitely agree with the robot that the speech was monotonous There wasn't a huge pitch range. A lot of the words within sentences were not being stressed
Points of Disagreement	I don't agree that having a good volume can solely make the speech be followed without any difficulty One could scream but you can still have difficulty following the speech The robot contradicts itself when it says it had difficulty due to lack of pitch variety I didn't totally agree with the eye contact statement She could have done a better job in terms of maintaining eye contact
Participant4	
Points of Agreement	I use the kind of sandwich approach that the robot does The robot gives feedback about the whole presentation
Points of Disagreement	Whereas there was a definite change in how she presented from the first 30 seconds to the rest Therefore, it would be good if there were a way to take into account different parts; the introduction can be very telling I would provide overall feedback, and then I would talk about specific areas
Participant5	
Points of Agreement	I did like the robot's observations about a large number of filler words, and its instructions The robot did a good job on assessing the articulation rate, as she did have a suitable articulation rate
Points of Disagreement	The problem is that all these things are kind of tied together The speech rate is a composite measure of speed and breakdowns together She did not have a good speech rate I disagree with the robot's feedback about pitch variety, as I thought her speech was quite monotonic
Participant6	
Points of Disagreement	I would disagree about the intonation variety. It depends on the listener's expectations though Pitch variety is very culturally determined, and in some cultures, low amounts of pitch variety are totally acceptable But if I were the Robot, I would advise trying to expand the pitch range a bit more
Participant7	
Points of Agreement	I thought the feedback was quite accurate in terms of what I would say as well I agree with the robot that her speech rate was in the optimal zone I thought it was interesting that the robot pointed out her lack of pitch this time
Points of Disagreement	With accented speech, there is a U-shaped curve in terms of comprehensibility Too fast speech is more incomprehensible, and too slow speech is too hard to sustain attention From my perspective, you cannot really highly accurately measure pitch variety
Participant8	
Points of Agreement	I agree with the feedback on articulation rate and pitch variety
Points of Disagreement	Due to the accent and articulation, there were portions that I was unable to understand Regarding volume, there were times when he talked to himself The first language plays an important role in how you pronounce the vowels and intonation Some languages are stress-timed and some are syllable-timed which reflects tone, prominence and stress The robot does not pick up on distinguishing between languages
Participant9	
Points of Disagreement	I would ask them to embrace the idea of the pause, a bit more It advised speaking in shorter sentences, but taking longer pauses at critical points should be mentioned She appeared to be very confident, but at the same time, it seemed that she was just trying to get it over with

4.4 Discussion

The purpose of this study was to evaluate how participants perceived a social humanoid robot acting as a public speaking coach. We incorporated various functionalities into the robot that mimic the tasks of a human public speaking coach. The robot’s responsibilities included analyzing the quality of participants’ speeches and their engagement with the audience during presentations. It provided verbal feedback, highlighting both the positive aspects of their presentations and the areas that required improvement. For this research, we employed the Pepper robot, selected for its expressive capabilities, anthropomorphic design, and a wide repertoire of gestures and arm movements.

Our evaluation of the robot encompassed various dimensions, including participants’ acceptance of the robot and their willingness to use it for the study’s purpose using the Almere model, perceived interpersonal closeness between participants and the robot, perceived human nature qualities of the robot, and general perceptions of the robot’s competence, usefulness, as well as evaluations of the feedback provided by the robot.

4.4.1 Acceptance and Intention to Use

Further exploration of the Almere questionnaire reveals that participants reported low anxiety scores, particularly on the Anxiety (ANX) subscale, during their interactions with the robot. This finding has both positive and negative implications for our experiment.

On the positive side, participants experienced reduced anxiety when addressing the robot, likely because it lacks the capacity to make negative judgments, diminishing the fear of negative evaluation in comparison to working with a human coach. Therefore, it provides a conducive environment for individuals to unleash their full potential. It is worth noting that, participants with higher NARS scores, indicating overall negative attitudes towards robots, expressed more anxiety when interacting with the Pepper robot. On the other hand, a potential downside or counterfactual to consider is that the scenario might not have effectively replicated a true presentation experience. Consequently, participants may not have experienced anxiety in the first place, as they were aware that this was not a realistic presentation rehearsal.

The results regarding participants’ Attitude Toward the Agent (ATT), Perceived Usefulness (PU), Trust in the robot’s assessments (Trust), and Perceived Sociability (PS) were above the “Agree” score. Also, Perceived Ease of Use (PEOU), Perceived Enjoyment (PENJ), and Perceived Adaptiveness (PAD) fell close to the “Agree” score. These results

collectively indicate an inclination, or an intention to use and the acceptance of the system. As indicated by the findings from [57], and as illustrated in Figure 4.3, the constructs mentioned are significantly linked to the intention to use the system.

Individuals with higher public speaking anxiety levels (PRCA-PS) exhibited less trust in the robot’s assessments, and their perceived usefulness was lower as compared to those with lower levels. Also, having higher levels of public speaking anxiety contributed to a reduced tendency to use the system in future rehearsal sessions. The negative correlation that was observed doesn’t necessarily imply that individuals with high public speaking anxiety hold a negative view of the system. This is supported by the fact that the error bars for all of the subscales displayed a small range, indicating that participants, in general, had a similar range of ratings. It; however, indicates their lower preference for the system when compared to participants with lower levels of public speaking anxiety, which can also be attributed to their discomfort with the mere act of delivering a presentation.

Participants who exhibited high levels of negative attitudes toward robots, as measured by the Negative Attitudes toward Robots Scale (NARS), also demonstrated increased anxiety in their interactions with robots, as measured through the Anxiety (ANX) construct in the Almere model. This aligns with prior research that has explored the positive interconnections between negative attitudes toward robots, anxiety, and subsequent behaviours toward robots, often resulting in communication avoidance between humans and robots [92]. NARS was also negatively associated with the attitude towards technology(ATT), and intention to use (ITU) in the Almere model.

4.4.2 Perceived Closeness

In this study, the results of the IOS scale indicated that a significant portion of the participants, specifically 70%, selected the diagrams depicting “some overlap” and beyond, as a representation of their interpersonal closeness with the Pepper robot. Furthermore, 20% of participants selected diagrams indicating “strong overlap,” “very strong overlap,” and “most overlap.” These findings highlight that a substantial majority, or 70% of the participants, experienced a sense of rapport and interconnection with the Pepper robot during their interactions.

Furthermore, participants who identified active listening behaviour in the robot exhibited a significantly greater sense of personal closeness with it. This observation aligns with previous research that has demonstrated how the perception and recognition of active listening can contribute to the development of positive interpersonal relationships [34, 45].

As indicated in a prior study [60], the perception of active listening triggers the activation of a specific brain region known as the ventral striatum. The striatum plays a significant role in processing various types of rewards, and the ventral striatum is capable of encoding a wide array of rewards, including abstract rewards such as the “warm glow,” which is the pleasant feeling experienced following receiving acts of kindness [60]. Consequently, the recognition of active listening can be considered a form of reward.

4.4.3 Perceived Human Nature

The impact of human-robot interaction (HRI) is influenced not only by the degree of a robot’s anthropomorphism but also by the manner in which individuals anthropomorphize it [135]. Human nature (HN) traits serve as distinguishing factors between humans and machines. It is conceivable that when we attribute these traits to a robot, it not only becomes more human-like in our perception but also appears less machine-like at the same time [135].

In our assessment, we considered the perceived human nature of the agent, with the hypothesis that the more human nature attributes the agent is perceived to possess, the higher the likelihood that participants would actively engage in interactions with the agent, as the agent becomes more human-like and can foster a sense of familiarity with participants. This notion is supported by previous research on self-disclosures, stating that anthropomorphism significantly influences how people perceive disclosures, the volume of information shared, the manner in which information is communicated, and the duration of the interaction. According to [74], participants would prefer to communicate with humans due to their familiarity with the communication experience [74].

Among the human nature attributes, emotional responsiveness was notably lower than other attributes. This observation aligns with the concept of empathy, which involves the ability to recognize one’s emotional state and respond with appropriate affect [96]. In this particular scenario, the robot’s role is to objectively evaluate participants’ presentations without engaging in affective responses. Hence, it is entirely reasonable that participants did not perceive any emotionality or empathy in the robot. It is worth noting that designing the robot’s personality or emotional responses was not within the scope of this study.

In our observations, we noted particularly high scores in human nature attributes related to warmth, agency, and openness. Specifically, we identified strong positive correlations between the perceived human-likeness of the agent and the robot’s listening behaviour during the presentations. This suggests that participants who perceived the robot as actively listening and attentive during the presentations were more likely to rate the robot as

possessing human nature attributes. According to [36], perceptions of anthropomorphism and the sense of human nature in an agent can be linked to the agent’s embodiment as well as its verbal and non-verbal social behaviours.

4.4.4 Evaluation of the Robot as a Public Speaking Coach

The assessment of the robot’s competence, helpfulness, and agreement scores as a public speaking rehearsal coach indicates a notably high level of acceptance of the system’s ability to evaluate presenters. Our findings reveal that participants’ ratings of the robot’s competence, its helpfulness, and participants’ willingness to use the system are positively correlated with the relevance and accuracy of the feedback provided, as well as the perceived attentiveness of the system. It can be inferred that participants’ evaluations of the system’s effectiveness as a rehearsal coach are linked to both the agent’s non-verbal active listening behaviour and its performance in providing appropriate verbal feedback for participants.

4.4.5 System Evaluation By a Human Coach

Our aim was not only to assess the participant’s acceptance of the robot’s ability to evaluate their presentations but also to obtain expert opinions regarding the system’s performance in relation to the presenters’ performances. The results of the questions posed to the human coach reflected a high level of acceptance with respect to the relevance and informative nature of the robot’s feedback.

This is promising, especially given that the human coach gave notably high scores concerning the system’s future use and overall perceived usefulness. A qualitative analysis of the human coach’s feedback after reviewing each video unveiled some limitations and highlighted areas for future improvement. According to the human coach’s observations, the robot performed well in evaluating articulation rate and assessing voice volume/intensity during presentations. However, in the case of the robot’s assessment of pitch variety, there were instances where the human coach disagreed with the robot’s positive evaluation of intonations and pitch variation, indicating an expectation of greater pitch variation.

While acknowledging the cultural and subjective aspects of evaluating pitch variation, it was pointed out that, in North America, a charismatic and engaging speech typically involves a broad range of pitch variation and specific stress on keywords. It was noted that the robot appeared to be relatively lenient in evaluating some participants, even though they had the potential to improve and incorporate more variation in their presentations.

Similar considerations apply to the evaluation of eye contact, as there were cases where the human coach anticipated a higher degree of eye contact maintenance during the presentations. While acknowledging cultural variations in eye contact preferences for different countries, the human coach stressed the importance of maintaining eye contact with the audience, particularly in North America. As a result, it might be worthwhile for future studies to explore the analysis with higher cut-off thresholds for both eye contact and pitch variety.

Additionally, the robot could be adjusted to better analyze accented speech. Given that certain languages exhibit stress-timed or syllable-timed characteristics [107] that might be more or less comfortable with varying pitch, the robot can adapt its feedback generation to accommodate accented speech effectively. Additionally, accented speech can affect the speech rate, where a high speech rate can make the speech incomprehensible.

Moreover, one of the limitations of the robot was its reliance on average-based calculations for analysis. In reality, there are common scenarios where presenters may act differently at the start of their presentation compared to the rest, or where their performance varies over the course of the presentation. To address this, the robot can offer feedback on specific time intervals when the speaker could enhance their performance during those particular segments of the presentation. Furthermore, it is worth considering an analysis of speech rate, as well as articulation rate, and incorporating the rate and duration of silent pauses in speech. As mentioned earlier, speech rate serves as a valuable measure for assessing speech fluency. Therefore, the robot can analyze and identify silent pauses in the presenter’s speech, and provide suggestions for when the presenter should take a deliberate pause to reduce the overuse of filler words.

Chapter 5

Conclusion, Limitations, and Future Work

5.1 Summary of Findings

In this section, we will provide concise answers to the research questions on a broad scale.

RQ1 - How will participants evaluate two interactive agents with distinct embodiments, one featuring a minimalist design and the other a more complex embodiment capable of displaying various non-verbal gestures and body movements as public speaking coaches?

The social robot was notably more anthropomorphized than the voice assistant agent across all human nature attributes, underlining the pivotal role of embodiment in assessing human likeness and humanness. The RoSAS did not reveal substantial differences between the agents, except for the perception of the social robot as warmer and friendlier, underscoring the importance of social embodiment and non-verbal social behaviour in conveying warmth. Additionally, the social robot was found to be more likeable, although not perceived as more intelligent. It's worth noting that these differences were related to the social aspects of the agent rather than the agent's ability to accomplish its task. In terms of the task at hand (presentation evaluation), verbal feedback appeared to be a more crucial factor, with all agents being equally effective in this regard.

RQ2 - How can the presence of active listening behaviour in the social robot influence participants' perceptions of the agents as public speaking coaches, and whether possessing this behaviour is useful in the given context?

In terms of the agent’s engagement in the presentation, naturalness, participants’ satisfaction, and the humane quality of the agent, both the embodiment and active listening behaviour played significant roles. Notably, active listening behaviour appeared to be particularly important in these aspects, as differences were observed primarily between the ActiveListenR agent and the VoiceAgent, with no noticeable distinctions between the PassiveListenR agent and VoiceAgent. The ActiveListenR agent was perceived as more natural and satisfying compared to the PassiveListenR agent.

RQ3 - How do participants’ personality traits affect their evaluations and ratings of a social agent as a coach?

Our research findings emphasize that individuals who exhibit high levels of agreeableness showed a significantly stronger tendency to use socially interactive agents as coaches for future presentation rehearsals. They also reported experiencing a deeper sense of rapport with the agents. Furthermore, participants with higher agreeableness scores perceived the agent as more natural and attributed higher human-nature attributes to it.

Conversely, conscientiousness showed an inverse correlation with assessments of the agents as coaches. Participants with higher conscientiousness scores tended to view the agent as less warm and emotionally expressive. They also displayed a lower inclination towards considering the use of the agent in the future for presentation rehearsals. They also reported experiencing a lower level of rapport with the agent.

Thematic analysis results indicated that participants generally welcomed the idea of using an socially interactive agent as a public speaking coach, primarily to mitigate subjective and biased evaluations and to practice fundamental public speaking skills. However, for more detailed feedback and a deeper understanding of human emotions, participants acknowledged the necessity of a human coach.

RQ4 - How participants would evaluate the robot’s feedback and its behaviour as a coach, and if they would like to use it in the future?

According to the results of the experiment, the Pepper robot emerged as a viable choice for a public speaking rehearsal coach. Participants expressed their acceptance and their intention to use the robot. They also reported a sense of interpersonal closeness and rapport with the robot, with nearly half of the participants indicating an equal or greater connection between themselves and the robot. Furthermore, participants attributed human-like qualities to the robot, particularly in relation to its active listening behaviour.

Perceived usefulness, willingness to use the robot as a presentation rehearsal coach, and the assessment of the robot’s competence as a coach were strongly linked to the accuracy of the feedback and participants’ perception of the robot’s attentiveness. Notably,

participants who reported negative attitudes towards the robot, and high levels of public speaking anxiety reported less intention to use as well as less acceptance towards the robot.

RQ5 - How does a human public speaking coach assess the accuracy of the feedback provided by the robot as a coach? The human coach’s assessment generally demonstrated a high level of feedback relevance and informativeness. However, there were some suggested improvements to the robot’s performance that are illustrated as suggestions for future work.

5.2 Contribution to Knowledge

We believe that the results presented in this thesis contribute valuable insights to the field of human-robot interaction, human-agent interaction, speech processing, and public speaking training. The findings can be particularly beneficial for designers of interactions seeking to design interactive agents for educational, coaching, and well-being purposes. We know that humanness has two dimensions, Human Uniqueness (HU) and Human Nature (HN), and as the results showed, a social robot’s perceived HN is more than that of a voice assistant agent. Therefore, social robots can be used for purposes specifically aimed for agents possessing high levels of humanness.

Also, we discovered that in a social robot, the presence of active listening in the form of non-verbal backchanneling can enhance the perceived naturalness of the interaction. This finding holds particular significance when considering the use of social robots in counselling or coaching roles. This is due to the fact that previous studies [74] have demonstrated that perception of humanness is particularly important in the aforementioned areas since humans are more likely to engage in communication when they sense familiar interactions, much akin to human-to-human interactions.

Addressing the second study, we integrated the robot with the necessary features to demonstrate its capability in its role, we developed a semi-autonomous method for analyzing speech quality, disfluencies, and audience orientation and the robot generates feedback relatively fast in less than a minute. We integrated insights from speech processing, linguistics, and communication into human-robot interaction, uncovering useful findings and suggesting potential directions for future research. Our study revealed a strong preference for using social robots in the realm of public speaking training and communication skills, socially interactive agents—such as social robots were highly welcomed as rehearsal coaches for university students. The speech processing module contributes to automating the analysis of speech during public speaking, allowing for the continuous monitoring of

appropriate ranges of speech parameters. This is particularly relevant when considering the use of the PRAAT software and its Python API. However, it is more beneficial to maintain constant monitoring of speech parameters rather than relying on averages.

We had the opportunity to seek an expert evaluation of the developed system, and the results indicated a strong compatibility between their assessments of the presentations and the evaluations provided by the robots. Notably, expectations regarding a wide range of pitch variety and consistent eye contact are particularly important in North America; consequently, the thresholds introduced can be considered even higher. Moreover, the first language participants speak plays a significant role in their ability to maintain a wide range of pitch variety, just as cultural background influences eye contact. Thus, in future, the robot’s feedback delivery can be tailored to account for the participant’s cultural background, in a personalized manner.

5.3 Limitations and Future Work

Our system exhibited certain limitations that can guide future work. One notable limitation is the evaluation of presentations based solely on the overall performance of participants, without addressing minor specific flaws in particular segments of the presentation, as long as the overall performance meets the acceptance thresholds. There is room for improvement in the system to adopt a more granular approach to presentation analysis, providing assessments for the introduction, the main body, and the conclusion of the presentation, offering a more comprehensive evaluation.

The system currently does not consider accented speech, which is a significant factor in a culturally diverse society like Canada. To enhance its inclusivity, the robot can adopt a more flexible approach by taking into account participants’ varying levels of English language proficiency. The first language spoken by participants, as well as the culture in which they grew up, significantly influences the manner in which they deliver presentations and their patterns of varying intonations and eye contact during speech. In North America, there is a general preference for more pitch variation and maintaining eye contact with the audience during presentations. Students residing in North America need to adapt to these preferences to advance their job prospects, etc.

Regarding pitch variation, future development might adopt a more granular approach by analyzing the intonation in relation to specific utterances. For example, there is the phenomenon called “uptalk”, which involves speaking with a rising intonation primarily to convey uncertainty, as if the speaker cannot make up their mind, and to suggest that they

want the listener to do that for them. It involves employing a rising intonation pattern after a declarative statement to seek feedback or reassurance from the listener [132, 62]. On the other hand, a reassuring phrase can be spoken with a falling tone indicating the speaker’s confidence [82].

Another limitation that can be taken care of in future studies is that we could ask participants in the demographics questionnaire not only about their prior interaction experiences with a robot as users but also about any experience in the development and design of robots and/or of interactions with robots, as these could significantly affect the overall perceptions.

Additionally, a potential direction for future research could involve improving the robot’s capacity to comprehend the content of presentations. One method for the robot to demonstrate its understanding of the content would be by being able to pose questions related to the presentation on different slides after the presentation concludes. This capability could help the robot explicitly demonstrate its attentiveness and understanding of the content presented by the participants. This approach would align with the practice of some participants in our study who concluded their presentations by inviting the robot to ask questions.

Additional analysis that can be addressed as future work is the time of the presentation delivery in front of the robot that participants took, to analyze their anxiety levels and to assess if they wanted to get through the presentation as quickly as possible. The time duration participants took to deliver presentations in front of the robot, could be one possible indicator of their engagement in the presentation.

Another future, valuable addition to the rehearsal coach could involve enhancing its ability to establish synchrony with the participants, thus improving the sense of interpersonal closeness. Previous studies have highlighted the significance of synchrony in communication and entertainment, where speakers adjust their speech in harmony with their partners, including factors such as pitch, intensity, and jitter [16]. This aspect warrants further analysis and exploration in future research.

To elevate the agent’s functionality, it can be fine-tuned to analyze body language and mannerisms. Additionally, integrating the agent with the capability to display facial expressions could enhance its naturalness and engagement, improving its interaction with participants.

References

- [1] John Algeo. *British or American English?: A Handbook of Word and Grammar Patterns*. Studies in English Language. Cambridge University Press, 2006.
- [2] Antonio Andriella, Rubén Huertas-García, Santiago Forgas-Coll, Carme Torras, and Guillem Alenyà. Discovering sociable: Using a conceptual model to evaluate the legibility and effectiveness of backchannel cues in an entertainment scenario. In *2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, pages 752–759, 2020.
- [3] Michael Argyle. *Bodily communication*. Routledge, 2013.
- [4] Arthur Aron, Elaine N Aron, and Danny Smollan. Inclusion of other in the self scale and the structure of interpersonal closeness. *Journal of personality and social psychology*, 63(4):596, 1992.
- [5] Reza Asadi, Ha Trinh, Harriet J. Fell, and Timothy W. Bickmore. Intelliprompter: Speech-based dynamic note display interface for oral presentations. In *Proceedings of the 19th ACM International Conference on Multimodal Interaction, ICMI '17*, page 172–180, New York, NY, USA, 2017. Association for Computing Machinery.
- [6] Joe Ayres. Coping with speech anxiety: The power of positive thinking. *Communication Education*, 37(4):289–296, 1988.
- [7] Christoph Bartneck, Andreas Duenser, Elena Moltchanova, and Karolina Zawieska. Comparing the similarity of responses received from studies in amazon’s mechanical turk to studies conducted online and with direct recruitment. *PLOS ONE*, 10(4):1–23, 04 2015.
- [8] Christoph Bartneck, Dana Kulić, Elizabeth Croft, and Susana Zoghbi. Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence,

and perceived safety of robots. *International Journal of Social Robotics*, 1(1):71–81, Jan 2009.

- [9] Stephanie Berger, Oliver Niebuhr, and Benno Peters. Winning over an audience—a perception-based analysis of prosodic features of charismatic speech. In *Proc. 43rd Annual Conference of the German Acoustical Society, Kiel, Germany*, pages 1454–1457, 2017.
- [10] Jasmin Bernotat and Friederike Eyssel. A robot at home — how affect, technology commitment, and personality traits influence user experience in an intelligent robotics apartment. In *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pages 641–646, 2017.
- [11] Amy M. Bippus and John A. Daly. What do people think causes stage fright?: Naïve attributions about the reasons for public speaking anxiety. *Communication Education*, 48(1):63–72, 1999.
- [12] Anke W. Blöte, Marcia J.W. Kint, Anne C. Miers, and P. Michiel Westenberg. The relation between public speaking anxiety and social anxiety: A review. *Journal of Anxiety Disorders*, 23(3):305–313, 2009.
- [13] Graham D. Bodie. A racing heart, rattling knees, and ruminative thoughts: Defining, explaining, and treating public speaking anxiety. *Communication Education*, 59(1):70–105, 2010.
- [14] Graham D Bodie. Listening as positive communication. *The positive side of interpersonal communication*, pages 109–125, 2012.
- [15] Paul Boersma and Vincent Van Heuven. Speak and unspeak with praat. *Glott International*, 5(9/10):341–347, 2001.
- [16] Stephanie A Borrie, Nichola Lubold, and Heather Pon-Barry. Disordered speech disrupts conversational entrainment: a study of acoustic-prosodic entrainment and communicative success in populations with communication challenges. *Front Psychol*, 6:1187, August 2015.
- [17] Konstantinos Bousmalis, Marc Mehu, and Maja Pantic. Towards the automatic detection of spontaneous agreement and disagreement based on nonverbal behaviour: A survey of related cues, databases, and tools. *Image and Vision Computing*, 31(2):203–221, 2013. Affect Analysis In Continuous Input.

- [18] Hamparsum Bozdogan. Model selection and akaike’s information criterion (aic): The general theory and its analytical extensions. *Psychometrika*, 52(3):345–370, Sep 1987.
- [19] Virginia Braun and Victoria Clarke. Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3:77–101, 01 2006.
- [20] Rodney A. Brooks. Elephants don’t play chess. *Robotics and Autonomous Systems*, 6(1-2):3 – 15, 1990.
- [21] Lawrence J. Brunner. Smiles can be back channels.: Semantic scholar, Jan 1979.
- [22] Estelle Campione and Jean Véronis. A large-scale multilingual study of silent pause duration. In *Proc. Speech Prosody 2002*, pages 199–202, 2002.
- [23] Colleen M. Carpinella, Alisa B. Wyman, Michael A. Perez, and Steven J. Stroessner. The robotic social attributes scale (rosas): Development and validation. In *2017 12th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 254–262, 2017.
- [24] Nicola Cathcart, Jean Carletta, and Ewan Klein. A shallow model of backchannel continuers in spoken dialogue. In *Proceedings of the Tenth Conference on European Chapter of the Association for Computational Linguistics - Volume 1*, EACL ’03, page 51–58, USA, 2003. Association for Computational Linguistics.
- [25] Minji Cho, Sang-su Lee, and Kun-Pyo Lee. Once a kind friend is now a thing: Understanding how conversational agents at home are forgotten. In *Proceedings of the 2019 on Designing Interactive Systems Conference*, DIS ’19, page 1557–1569, New York, NY, USA, 2019. Association for Computing Machinery.
- [26] Softbank Robotics Documentation what’s new in naoqi 2.5? http://doc.aldebaran.com/2-5/index_dev_guide.html. Accessed: 2023-11-08.
- [27] Bart Craenen, Amol Deshmukh, Mary Ellen Foster, and Alessandro Vinciarelli. Do we really like robots that match our personality? the case of big-five traits, godspeed scores and robotic gestures. In *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pages 626–631, 2018.
- [28] Nivja H. de Jong and Ton Wempe. Praat script to detect syllable nuclei and measure speech rate automatically. *Behavior Research Methods*, 41(2):385–390, May 2009.
- [29] Michael A DeCoske and Sara J White. Public speaking revisited: delivery, structure, and style. *American journal of health-system pharmacy*, 67(15):1225–1227, 2010.

- [30] Huaijin Deng, Youchao Lin, Takehito Utsuro, Akio Kobayashi, Hiromitsu Nishizaki, and Junichi Hoshino. Integrating disfluency-based and prosodic features with acoustics in automatic fluency evaluation of spontaneous speech. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 6429–6437, Marseille, France, May 2020. European Language Resources Association.
- [31] Alan Dennis and Susan Kinney. Testing media richness theory in the new media: The effects of cues, feedback, and task equivocality. *Information Systems Research*, 9:256–274, 09 1998.
- [32] Zijian Ding, Jiawen Kang, Tinky Oi Ting HO, Ka Ho Wong, Helene H. Fung, Helen Meng, and Xiaojuan Ma. Talktive: A conversational agent using backchannels to engage older adults in neurocognitive disorders screening, 2022.
- [33] Allen T. Dittmann and Lynn G. Llewellyn. Relationship between vocalizations and head nods as listener responses. *Journal of Personality and Social Psychology*, 9(1):79–84, 1968.
- [34] Changming Duan and Clara E Hill. The current state of empathy research. *Journal of counseling psychology*, 43(3):261, 1996.
- [35] Connor Esterwood, Kyle Essenmacher, Han Yang, Fanpan Zeng, and Lionel Peter Robert. A meta-analysis of human personality and robot acceptance in human-robot interaction. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI ’21, New York, NY, USA, 2021. Association for Computing Machinery.
- [36] Friederike Eyssel, Frank Hegel, Gernot Horstmann, and Claudia Wagner. Anthropomorphic inferences from emotional nonverbal cues: A case study. In *19th International Symposium in Robot and Human Interactive Communication*, pages 646–651, 2010.
- [37] Thijs Fassaert, Sandra van Dulmen, François Schellevis, and Jozien Bensing. Active listening in medical consultations: development of the active listening observation scale (ALOS-global). *Patient Educ Couns*, 68(3):258–264, August 2007.
- [38] Parselmouth praat scripts in python. <https://osf.io/6dwr3/>. Accessed: 2023-11-08.
- [39] Alexander Flint. *The effects of interlocutor backchannels and L1 backchannel norms on the speech of L2 English learners*. PhD thesis, University of Oxford, 2016.

- [40] Haruaki Fukuda and Kazuhiro Ueda. Interaction with a moving object affects one’s perception of its animacy. *International Journal of Social Robotics*, 2(2):187–193, Jun 2010.
- [41] Rod Gardner. Between speaking and listening: The vocalisation of understandings1. *Applied Linguistics*, 19, 06 1998.
- [42] Rod Gardner. *When Listeners Talk: Response Tokens and Listener Stance*. John Benjamins, Amsterdam, 2001.
- [43] Moojan Ghafurian, Neil Budnarain, and Jesse Hoey. Improving humanness of virtual agents and users’ cooperation through emotions. *IEEE Transactions on Affective Computing*, 2021.
- [44] Samuel Gibbs. Amazon launches home robot astro and giant alexa display. *The Guardian*, 2021.
- [45] Gerald A. Gladstein. Empathy and counseling outcome: An empirical and conceptual review. *The Counseling Psychologist*, 6(4):70–79, 1977.
- [46] Alexander Goberman, Stephanie Hughes, and Todd Haydock. Acoustic characteristics of public speaking: Anxiety and practice effects. *Speech Communication*, 53:867–876, 07 2011.
- [47] Andy Goodman and Cause Communications. *Why Bad Presentations Happen to Good Causes: And how to Ensure They Won’t Happen to Yours*. Cause Communications, 2006.
- [48] Charles Goodwin. *Conversational Organization: Interaction between Speakers and Hearers*. Academic Press, London, 1981.
- [49] Nest Mini. still mini. even more mighty. https://store.google.com/product/google_nest_mini?hl=en-GB&pli=1. Accessed: 2023-11-08.
- [50] Samuel D Gosling, Peter J Rentfrow, and William B Swann. A very brief measure of the big-five personality domains. *Journal of Research in Personality*, 37(6):504–528, 2003.
- [51] Dan Grandstaff. *Speaking As A Professional: Enhance Your Therapy Or Coaching Practice Through Presentations*. WW Norton & Company, 2004.

- [52] Jonathan Gratch, Ning Wang, Anna Okhmatovskaia, Francois Lamothe, Mathieu Morales, R. J. van der Werf, and Louis-Philippe Morency. Can virtual humans be more engaging than real ones? In Julie A. Jacko, editor, *Human-Computer Interaction. HCI Intelligent Multimodal Interaction Environments*, pages 286–297, Berlin, Heidelberg, 2007. Springer Berlin Heidelberg.
- [53] Muriel A. Hagenaars and Agnes van Minnen. The effect of fear on paralinguistic aspects of speech in patients with panic disorder with agoraphobia. *Journal of Anxiety Disorders*, 19(5):521–537, 2005.
- [54] Fasih Haider, Loredana Cerrato, Nick Campbell, and Saturnino Luz. Presentation quality assessment using acoustic information and hand movements. In *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, page 2812–2816. IEEE Press, 2016.
- [55] Judith A. Hall and Sarah D. Gunnery. *21 Gender differences in nonverbal communication*, pages 639–670. De Gruyter Mouton, Berlin, Boston, 2013.
- [56] Nick Haslam, Stephen Loughnan, Yoshihisa Kashima, and Paul Bain. Attributing and denying humanness to others. *European Review of Social Psychology*, 19(1):55–85, 2008.
- [57] Marcel Heerink, Ben Kröse, Vanessa Evers, and Bob Wielinga. Assessing acceptance of assistive social agent technology by older adults: the almere model. *International Journal of Social Robotics*, 2(4):361–375, Dec 2010.
- [58] Julio Hernández-March, Mónica Martín del Peso, and Santiago Leguey. Graduates’ skills and higher education: The employers’ perspective. *Tertiary Education and Management*, 15(1):1–16, Mar 2009.
- [59] Rebecca Hincks. Measuring liveliness in presentation speech. In *Ninth European Conference on Speech Communication and Technology*, 2005.
- [60] Akihiro T. Sasaki Hiroaki Kawamichi, Kazufumi Yoshihara. Perceiving active listening activates the reward system and improves the impression of relevant experiences. *Social Neuroscience*, 10(1):16–26, 2015. PMID: 25188354.
- [61] Sherry Holladay and W. Coombs. Communicating visions: An exploration of the role of delivery in the creation of leader charisma. *Management Communication Quarterly - MANAG COMMUN Q*, 6:405–427, 05 1993.

- [62] Janet Holmes. Functions of you know in women’s and men’s speech. *Language in Society*, 15(1):1–21, 1986.
- [63] Debra A. Hope, Richard G. Heimberg, and Cynthia L. Turk. *Managing Social Anxiety, Therapist Guide: A Cognitive-Behavioral Therapy Approach*. Oxford University Press, 05 2010.
- [64] Mohammed (Ehsan) Hoque, Matthieu Courgeon, Jean-Claude Martin, Bilge Mutlu, and Rosalind W. Picard. Mach: My automated conversation coach. In *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, UbiComp ’13, page 697–706, New York, NY, USA, 2013. Association for Computing Machinery.
- [65] Nusrah Hussain, Engin Erzin, T. Metin Sezgin, and Yücel Yemez. Training socially engaging robots: Modeling backchannel behaviors with batch reinforcement learning. *IEEE Transactions on Affective Computing*, 13(4):1840–1853, 2022.
- [66] Nusrah Hussain, Engin Erzin, Tevfik Sezgin, and Yucel Yemez. Speech driven backchannel generation using deep q-network for enhancing engagement in human-robot interaction. *CoRR*, abs/1908.01618, 2019.
- [67] Ewa Jacewicz, Robert A Fox, Caitlin O’Neill, and Joseph Salmons. Articulation rate across dialect, age, and gender. *Lang Var Change*, 21(2):233–256, July 2009.
- [68] Patrik Jonell, Taras Kucherenko, Ilaria Torre, and Jonas Beskow. Can we trust on-line crowdworkers? comparing online and offline participants in a preference test of virtual agents. In *Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents*, IVA ’20, New York, NY, USA, 2020. Association for Computing Machinery.
- [69] Malte F. Jung, Jin Joo Lee, Nick DePalma, Sigurdur O. Adalgeirsson, Pamela J. Hinds, and Cynthia Breazeal. Engaging robots: Easing complex human-robot teamwork using backchanneling. In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work*, CSCW ’13, page 1555–1566, New York, NY, USA, 2013. Association for Computing Machinery.
- [70] Younbo Jung and Kwan Min Lee. Effects of physical embodiment on social presence of social robots. *Proceedings of PRESENCE*, 2004:80–87, 2004.
- [71] Everlyne Kimani, Timothy Bickmore, Ha Trinh, and Paola Pedrelli. You’ll be great: Virtual agent-based cognitive restructuring to reduce public speaking anxiety. In

2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII), pages 641–647, 2019.

- [72] Mitsuhiro Kimoto, Takamasa Iio, Masahiro Shiomi, Ivan Tanev, Katsunori Shimohara, and Norihiro Hagita. Relationship between personality and robots’ interaction strategies in object reference conversations. In *Proceedings of the Second International Conference on Electronics and Software Science (ICESS2016)*, Japan, pages 128–136, 2016.
- [73] Dimosthenis Kontogiorgos, Andre Pereira, Olle Andersson, Marco Koivisto, Elena Gonzalez Rabal, Ville Vartiainen, and Joakim Gustafson. The effects of anthropomorphism and non-verbal social behaviour in virtual assistants. In *Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents, IVA ’19*, page 133–140, New York, NY, USA, 2019. Association for Computing Machinery.
- [74] Guy Laban, Jean-Noël George, Val Morrison, and Emily S. Cross. Tell me more! assessing interactions with social robots from speech. *Paladyn, Journal of Behavioral Robotics*, 12(1):136–159, 2021.
- [75] J. Richard Landis and Gary G. Koch. The measurement of observer agreement for categorical data. *Biometrics*, 33(1):159–174, 1977.
- [76] Sunok Lee, Minji Cho, and Sangsu Lee. What if conversational agents became invisible? comparing users’ mental models according to physical entity of ai speaker. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 4(3), sep 2020.
- [77] Rosemarijn Looije, Mark A. Neerincx, and Fokke Cnossen. Persuasive robotic assistant for health self-management of older adults: Design and evaluation of social behaviors. *International Journal of Human-Computer Studies*, 68(6):386–397, 2010. Human-Computer Interaction for Medicine and Health care (HCI4MED): Towards making Information usable.
- [78] Nichola Lubold and Heather Pon-Barry. Acoustic-prosodic entrainment and rapport in collaborative learning dialogues. In *Proceedings of the 2014 ACM Workshop on Multimodal Learning Analytics Workshop and Grand Challenge*, MLA ’14, page 5–12, New York, NY, USA, 2014. Association for Computing Machinery.
- [79] Andrew Kwok-Fai Lui, Sin-Chun Ng, and Wing-Wah Wong. A novel mobile application for training oral presentation delivery skills. In Jeanne Lam, Kwan Keung

- Ng, Simon K.S. Cheung, Tak Lam Wong, Kam Cheong Li, and Fu Lee Wang, editors, *Technology in Education. Technology-Mediated Proactive Learning*, pages 79–89, Berlin, Heidelberg, 2015. Springer Berlin Heidelberg.
- [80] Michal Luria, Guy Hoffman, and Oren Zuckerman. Comparing social robot, screen and voice interfaces for smart-home control. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, page 580–628, New York, NY, USA, 2017. Association for Computing Machinery.
 - [81] James C McCroskey, Michael J Beatty, Patricia Kearney, and Timothy G Plax. The content validity of the prca-24 as a measure of communication apprehension across communication contexts. *Communication Quarterly*, 33(3):165–173, 1985.
 - [82] C. Meghan McMurtry, Patrick J. McGrath, Elissa Asp, and Christine T. Chambers. Parental reassurance and pediatric procedural pain: A linguistic description. *The Journal of Pain*, 8(2):95–101, 2007.
 - [83] Tzvi Michelson and Shmuel Peleg. Audio-visual evaluation of oratory skills. In *2021 Third International Conference on Transdisciplinary AI (TransAI)*, pages 103–106, 2021.
 - [84] Billie Akwa Moore and Jacqueline Urakami. The impact of the physical and social embodiment of voice user interfaces on user distraction. *International Journal of Human-Computer Studies*, 161:102784, 2022.
 - [85] Louis-Philippe Morency, Iwan de Kok, and Jonathan Gratch. Predicting listener backchannels: A probabilistic multimodal approach. In Helmut Prendinger, James Lester, and Mitsuru Ishizuka, editors, *Intelligent Virtual Agents*, pages 176–190, Berlin, Heidelberg, 2008. Springer Berlin Heidelberg.
 - [86] Nasim Motalebi, Eugene Cho, S. Shyam Sundar, and Saeed Abdullah. Can alexa be your therapist? how back-channeling transforms smart-speakers to be active listeners. In *Conference Companion Publication of the 2019 on Computer Supported Cooperative Work and Social Computing*, CSCW '19, page 309–313, New York, NY, USA, 2019. Association for Computing Machinery.
 - [87] Prasanth Murali, Ha Trinh, Lazlo Ring, and Timothy Bickmore. A friendly face in the crowd: Reducing public speaking anxiety with an emotional support agent in the audience. In *Proceedings of the 21st ACM International Conference on Intelligent Virtual Agents*, IVA '21, page 156–163, New York, NY, USA, 2021. Association for Computing Machinery.

- [88] Iain R. Murray and John L. Arnott. Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion. *Journal of the Acoustical Society of America*, 93(2):1097–1108, 1993. Medline is the source for the MeSH terms of this document.
- [89] John Ashworth Nelder and Robert WM Wedderburn. Generalized linear models. *Journal of the Royal Statistical Society. Series A (General)*, 135(3):370–384, 2023/01/15/ 1972. Full publication date: 1972.
- [90] Oliver Niebuhr, Alexander Brem, Eszter Novák-Tót, and Jana Voße. *Charisma in business speeches: A contrastive acoustic-prosodic analysis of Steve Jobs and Mark Zuckerberg*, page 79. Speech Prosody Special Interest Group, May 2016. 8th Speech Prosody Conference ; Conference date: 31-05-2016 Through 03-06-2016.
- [91] Tatsuya Nomura, Takayuki Kanda, Tomohiro Suzuki, and K. Kato. Psychology in human-robot communication: an attempt through investigation of negative attitudes and anxiety toward robots. In *RO-MAN 2004. 13th IEEE International Workshop on Robot and Human Interactive Communication (IEEE Catalog No.04TH8759)*, pages 35–40, 2004.
- [92] Tatsuya Nomura, Takayuki Kanda, Tomohiro Suzuki, and Kensuke Kato. Prediction of human behavior in human–robot interaction using psychological scales for anxiety and negative attitudes toward robots. *IEEE Transactions on Robotics*, 24(2):442–451, 2008.
- [93] Judith Norback and Jill Hardin. Integrating workforce communication into senior design. *IEEE Transactions on Professional Communication*, 48(4):413–426, 2005.
- [94] Catharine Oertel, José Lopes, Yu Yu, Kenneth A. Funes Mora, Joakim Gustafson, Alan W. Black, and Jean-Marc Odobez. Towards building an attentive artificial listener: On the perception of attentiveness in audio-visual feedback tokens. In *Proceedings of the 18th ACM International Conference on Multimodal Interaction, ICMI ’16*, page 21–28, New York, NY, USA, 2016. Association for Computing Machinery.
- [95] Hae Won Park, Mirko Gelsomini, Jin Joo Lee, and Cynthia Breazeal. Telling stories to robots: The effect of backchanneling on a child’s storytelling. In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction, HRI ’17*, page 100–108, New York, NY, USA, 2017. Association for Computing Machinery.

- [96] Sung Park and Mincheol Whang. Empathy in human–robot interaction: Designing for social robots. *International journal of environmental research and public health*, 19(3):1889, 2022.
- [97] Maria Teresa Parreira, Sarah Gillet, Katie Winkle, and Iolanda Leite. How did we miss this? a case study on unintended biases in robot social behavior. In *Companion of the 2023 ACM/IEEE International Conference on Human-Robot Interaction, HRI '23*, page 11–20, New York, NY, USA, 2023. Association for Computing Machinery.
- [98] Paul A. Pavlou. Consumer acceptance of electronic commerce: Integrating trust and risk with the technology acceptance model. *International Journal of Electronic Commerce*, 7(3):101–134, 2003.
- [99] Miranda A.G. Peeters, Harrie F.J.M. van Tuijl, Christel G. Rutte, and Isabelle Reymen. Personality and team performance: a meta analysis. *European journal of personality*, 20(5):377–396, 2006.
- [100] Kathrin Pollmann, Christopher Ruff, Kevin Vetter, and Gottfried Zimmermann. Robot vs. voice assistant: Is playing with pepper more fun than playing with alexa? In *Companion of the 2020 ACM/IEEE international conference on human-robot interaction*, pages 395–397, 2020.
- [101] Praat: Doing phonetics by computer. <https://www.fon.hum.uva.nl/praat/>. Accessed: 2023-11-08.
- [102] Jakub Prochazka, Martin Ovcari, and Michal Durinik. Sandwich feedback: The empirical evidence of its effectiveness. *Learning and Motivation*, 71:101649, 2020.
- [103] Qualtrics xm. <https://www.qualtrics.com/>. Accessed: 2023-11-08.
- [104] Hugo Quené. Multilevel modeling of between-speaker and within-speaker variation in spontaneous speech tempo. *The Journal of the Acoustical Society of America*, 123(2):1104–1113, 2008.
- [105] Natalia Reich-Stiebert and Friederike Eyssel. (ir)relevance of gender? on the influence of gender stereotypes on learning with a robot. In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction, HRI '17*, page 166–176, New York, NY, USA, 2017. Association for Computing Machinery.
- [106] Edison Research. The smart audio report from npr and edison research, spring 2018, Jul 2018.

- [107] Peter Roach. On the distinction between ‘stress-timed’ and ‘syllable-timed’ languages. *Linguistic controversies*, 73:79, 1982.
- [108] Andrew Rosenberg and Julia Hirschberg. Acoustic/prosodic and lexical correlates of charismatic speech. In *Interspeech*, 2005.
- [109] Howard M Rosenfeld and Margaret Hancks. The nonverbal context of verbal listener responses. *The relationship of verbal and nonverbal communication*, 25, 1980.
- [110] Hanan Salam, Oya Çeliktutan, Isabelle Hupont, Hatice Gunes, and Mohamed Chetouani. Fully automatic analysis of engagement and its relationship to personality in human-robot interactions. *IEEE Access*, 5:705–721, 2017.
- [111] Olga Saukh and Balz Maag. Quantle: Fair and honest presentation coach in your pocket. In *2019 18th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*, pages 253–264, 2019.
- [112] Klaus R. Scherer. Expression of emotion in voice and music. *Journal of Voice*, 9(3):235–248, 1995.
- [113] Paul Schermerhorn, Matthias Scheutz, and Charles R. Crowell. Robot social presence and gender: Do females view robots differently than males? In *Proceedings of the 3rd ACM/IEEE International Conference on Human Robot Interaction, HRI ’08*, page 263–270, New York, NY, USA, 2008. Association for Computing Machinery.
- [114] Jan Schneider, Dirk Börner, Peter van Rosmalen, and Marcus Specht. Presentation trainer, your public speaking multimodal coach. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction, ICMI ’15*, page 539–546, New York, NY, USA, 2015. Association for Computing Machinery.
- [115] Candace L. Sidner, Christopher Lee, Cory D. Kidd, Neal Lesh, and Charles Rich. Explorations in engagement for humans and robots. *Artificial Intelligence*, 166(1):140–164, 2005.
- [116] Softbank robotics. <https://www.softbankrobotics.com/>. Accessed: 2023-11-08.
- [117] Micol Spitale, Minja Axelsson, and Hatice Gunes. Robotic mental well-being coaches for the workplace: An in-the-wild study on form. In *Proceedings of the 2023 ACM/IEEE International Conference on Human-Robot Interaction, HRI ’23*, page 301–310, New York, NY, USA, 2023. Association for Computing Machinery.

- [118] Anna-Brita Stenstrom. *An introduction to spoken interaction*. Longman, London, 1994.
- [119] John D. Sutter. When it comes to presentation, mark zuckerberg is no steve jobs, Jul 2011.
- [120] Dag Sverre Syrdal, Kheng Lee Koay, Michael L. Walters, and Kerstin Dautenhahn. A personalized robot companion? - the role of individual differences on spatial preferences in hri scenarios. In *RO-MAN 2007 - The 16th IEEE International Symposium on Robot and Human Interactive Communication*, pages 1143–1148, 2007.
- [121] M. Iftekhhar Tanveer, Emy Lin, and Mohammed (Ehsan) Hoque. Rhema: A real-time in-situ intelligent interface to help people with public speaking. In *Proceedings of the 20th International Conference on Intelligent User Interfaces, IUI '15*, page 286–295, New York, NY, USA, 2015. Association for Computing Machinery.
- [122] Kevin Tasa, Greg J. Sears, and Aaron C. H. Schat. Personality and teamwork behavior in context: The cross-level moderating role of collective efficacy. *Journal of Organizational Behavior*, 32(1):65–85, 2011.
- [123] Hartmut Traunmüller and Anders Eriksson. The perceptual evaluation of f0 excursions in speech as evidenced in liveliness estimations. *The Journal of the Acoustical Society of America*, 97:1905–15, 04 1995.
- [124] Ha Trinh, Reza Asadi, Darren Edge, and Timothy Bickmore. Robocop: A robotic coach for oral presentations. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 1(2), jun 2017.
- [125] Ha Trinh, Lazlo Ring, and Timothy Bickmore. Dynamicduo: Co-presenting with virtual agents. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, CHI '15*, page 1739–1748, New York, NY, USA, 2015. Association for Computing Machinery.
- [126] Ha Trinh, Koji Yatani, and Darren Edge. Pitchperfect: Integrated rehearsal environment for structured presentation preparation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '14*, page 1571–1580, New York, NY, USA, 2014. Association for Computing Machinery.
- [127] Jacqueline Urakami and Sujitra Sutthithatip. Building a collaborative relationship between human and robot through verbal and non-verbal interaction. In *Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction, HRI*

- '21 Companion, page 257–261, New York, NY, USA, 2021. Association for Computing Machinery.
- [128] Steven Vicinanza. Information technology as a discipline: What employers want in an it graduate. *ACM Inroads*, 4(4):59–63, dec 2013.
 - [129] Sebastian Wallkötter, Rebecca Stower, Arvid Kappas, and Ginevra Castellano. A robot by any other frame: Framing and behaviour influence mind perception in virtual but not real-world environments. In *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction, HRI '20*, page 609–618, New York, NY, USA, 2020. Association for Computing Machinery.
 - [130] Jinping Wang, Hyun Yang, Ruosi Shao, Saeed Abdullah, and S. Shyam Sundar. Alexa as coach: Leveraging smart speakers to build social agents that reduce public speaking anxiety. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, CHI '20*, page 1–13, New York, NY, USA, 2020. Association for Computing Machinery.
 - [131] Xingbo Wang, Haipeng Zeng, Yong Wang, Aoyu Wu, Zhida Sun, Xiaojuan Ma, and Huamin Qu. Voicecoach: Interactive evidence-based training for voice modulation skills in public speaking. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, CHI '20*, page 1–12, New York, NY, USA, 2020. Association for Computing Machinery.
 - [132] Paul Warren. *Uptalk: The phenomenon of rising intonation*. Cambridge University Press, 2016.
 - [133] Torsten Wörtwein, Louis-Philippe Morency, and Stefan Scherer. Automatic assessment and analysis of public speaking anxiety: A virtual audience case study. In *2015 International Conference on Affective Computing and Intelligent Interaction (ACII)*, pages 187–193, 2015.
 - [134] Victor H. Yngve. On getting a word in edgewise. In *CLS-70*, pages 567–577. University of Chicago, 1970.
 - [135] Jakub Złotowski, Ewald Strasser, and Christoph Bartneck. Dimensions of anthropomorphism: from humanness to humanlikeness. In *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction*, pages 66–73, 2014.

APPENDICES

Appendix A

Ethics Clearance Certificates

UNIVERSITY OF WATERLOO

Notification of Ethics Clearance to Conduct Research with Human Participants

Principal Investigator: Kerstin Dautenhahn (Electrical and Computer Engineering)

Student investigator: Delara Forghani (Electrical and Computer Engineering)

Co-Principal Investigator: Chrystopher Nehaniv (Electrical and Computer Engineering)

Co-Investigator: Moojan Ghafurian (Electrical and Computer Engineering)

Collaborator: Samira Rasouli (Electrical and Computer Engineering)

File #: 43950

Title: Evaluating people's perceptions of an agent as a public speaking coach

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

Initial Approval Date: 09/13/22 (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 (TCPS, 2nd edition), International Conference on Harmonization: Good Clinical Practice (ICH-GCP), the Ontario Personal Health Information Protection Act (PHIPA), the applicable laws and regulations of the province of Ontario. 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).

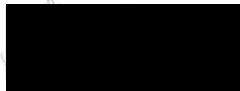
This study is to be conducted in accordance with the submitted application and the most recently approved versions of all supporting materials.

Expiry Date: 09/14/23 (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



Asima Naeem, Ethics Advisor, a33naeem@uwaterloo.ca, 519-888-4567, ext. 41067

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: EndMessage_v1_202207.pdf

file: ConsentForm_v1_202207.pdf

file: OnlineStudyCanadaDay.pptx

file: Procedures_v1_202207.pdf

file: RecruitmentMaterial_v1_202207.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.

UNIVERSITY OF WATERLOO

Notification of Ethics Clearance to Conduct Research with Human Participants

Principal Investigator: Kerstin Dautenhahn (Electrical and Computer Engineering)

Student investigator: Delara Forghani (Electrical and Computer Engineering)

Collaborator: Moojan Ghafurian (Electrical and Computer Engineering)

Co-Investigator: Chrystopher Nehaniv (Electrical and Computer Engineering)

Collaborator: Samira Rasouli (Electrical and Computer Engineering)

Collaborator: Melanie Jouaiti (Imperial College London)

File #: 44790

Title: Evaluating a Social Robot as a Public Speaking Coach

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

Initial Approval Date: 01/03/23 (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 (TCPS, 2nd edition), International Conference on Harmonization: Good Clinical Practice (ICH-GCP), the Ontario Personal Health Information Protection Act (PHIPA), the applicable laws and regulations of the province of Ontario. 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).

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

Expiry Date: 01/04/24 (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: Procedure_v1_202212.pdf

file: ExperimentalSetup.jpg

file: CanadaDaySlides.pptx

file: ConsentForm_v1_202212.pdf

file: InformationSheet_v1_202212.pdf

file: StudyFlyer_v1_202212.pdf

file: AppreciationLetter_v1_202212.pdf

file: PepperUsersManual_v1_202212.pdf

file: PEPPER_UserGuide_EN_2019 07 05_1.pdf

Approved Protocol Version 3 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 - Pre and Post-Study Questionnaires

B.1 Participants' Demographics

Demographic Information Form

Gender (write N/A if you prefer not to share):

Age:

Next

B.2 Pre-experimental Questions

Pre-experimental Questions

Have you taken public speaking classes/workshops before?

- ☐ Yes
- ☐ No
- ☐ Do not wish to share

Do you feel the need to take public speaking classes or workshops? (Please explain briefly, or write "Do not want to share" if you do not wish to answer)

Please select an option next to each statement to indicate the extent to which you agree or disagree with that trait about yourself.

I see myself as	Disagree strongly	Disagree moderately	Disagree a little	Neither agree nor disagree	Agree a little	Agree moderately	Agree strongly	Don't want to answer
Extraverted, Enthusiastic	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7	☐
Critical, quarrelsome	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7	☐
Dependable, self-disciplined	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7	☐
Anxious, easily upset	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7	☐
Open to new experiences, complex	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7	☐
Reserved, quiet	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7	☐
Sympathetic, warm	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7	☐
Disorganized, careless	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7	☐
Calm, emotionally stable	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7	☐
Conventional, uncreative	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7	☐

Next

B.3 Evaluation of the Agent as a Public Speaking Coach

How do you rate the agent as a coach?

(Scroll down to see all the questions)

I found the agent as a public speaking coach:

Not satisfying at all  Very satisfying

I found the interaction with the agent as a public speaking coach:

Not engaging at all  Very engaging

I found the behaviour of the agent as a public speaking coach:

Not natural at all  Very natural

I found the agent as a public speaking coach:

Not attentive at all  Very attentive

I found the agent as a public speaking coach:

Not competent at all  Very competent

I found the agent as a public speaking coach:

Not humane at all  Very humane

How would you rate the coach's feedback?

Not trustworthy at all  Very trustworthy

How well did you get the feedback provided by the coach?

Very difficult to understand  Very easy to understand

How helpful did you find this coach for students to improve their presentation skills?

Very unhelpful

Very helpful

How likely are you to rehearse with this coach for future presentations?

Not at all

Very much

How well do you think the coach can build rapport with a human partner?

Not at all

Very much

How familiar are you with interacting with voice assistant agents (such as Alexa, Google Nest, Siri)?

Not familiar at all

Very familiar

How familiar are you with interacting with social robots?

Not familiar not all

Very familiar

The topic of the presentation was about a city in Canada:

☐ True

☐ False

Next

B.4 Evaluation of the Human Nature Attributes of the Agent

I thought the agent was:

Unemotional	<input type="range"/>	Emotional
Cold	<input type="range"/>	Warm
Rigid	<input type="range"/>	Open
Passive	<input type="range"/>	Lively

The agent was nodding while the student was giving presentation:

☐ True

☐ False

Next

B.5 The Robotic Social Attributes Scale (RoSAS)

Using the scale provided, how closely are the following attributes associated with the agent?

	Not at all			A moderate amount			Very much so
	1	2	3	4	5	6	7
Dangerous	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7
Emotional	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7
Compassionate	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7
Aggressive	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7
Capable	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7
Happy	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7
Strange	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7
Organic	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7
Scary	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7
Responsive	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7
Awkward	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7
Knowledgeable	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7
Sociable	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7
Feeling	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7
Competent	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7
Reliable	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7
Awful	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7
Interactive	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7

The agent was expressing social cues (e.g. hand gestures) while giving feedback:

☐ True

☐ False

Next

B.6 Godspeed Perceived Intelligence, and Godspeed Likeability

Please rate your impression of the agent on these scales:

	1	2	3	4	5	
Ignorant	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	Knowledgeable
Awful	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	Nice
Unkind	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	Kind
Unintelligent	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	Intelligent
Dislike	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	Like
Foolish	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	Sensible
Incompetent	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	Competent
Unpleasant	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	Pleasant
Unfriendly	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	Friendly
Irresponsible	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	Responsible

I saw the student's face:

☐ True

☐ False

Next

B.7 Participants' Overall Evaluations

How do you feel about giving public presentations?

Not nervous at all



Very nervous

When I give a presentation, I feel:

Not comfortable at all



Very comfortable

Robots can be good coaches:

Strongly disagree



Strongly agree

How do you feel about having robots as public speaking coaches? Please explain your answer.

Please answer the following question and explain why you chose this answer:

I prefer:

- ☐ A robot coach like the one in the video
- ☐ A robot coach different from the one in the video
- ☐ An agent different from a robot, for example a voice assistant agent, such as the Google Nest
- ☐ A human coach

Appendix C

Study 2 - Pre and Post-Study Questionnaires

C.1 Participants' Demographics

Gender (If you prefer not to answer, please type "NA".)

Age (If you prefer not to answer, please type "NA".)

Education

☐ I am an undergraduate student

☐ I am a master's student

☐ I am a Ph.D. student

☐ other (please specify)

☐ Prefer not to answer

Robot

Have you ever programmed/interacted with a robot?

☐ Yes

☐ No

What did the robot/robots look like? (Select all that apply)

☐ Human-like robot

☐ Animal-like robot

☐ Other types of robots that are not human-like or animal-like

☐ Other robots (Please specify)

English

Are you a native English speaker or do you speak it as a second language?

☐ I am a native English speaker

☐ I speak English as a second language

☐ Prefer not to answer

How fluent do you think you are?

Not fluent
at all

Slightly
fluent

Moderately
fluent

Very fluent

Extremely
fluent

Prefer not
to answer

What languages do you speak at home? (If you prefer not to answer, please type "NA".)

How many presentations do you estimate you have given in the last two years? (If you prefer not to answer, please type "NA".)

Have you taken public speaking classes/workshops before?

- ☐ Yes
- ☐ No
- ☐ Prefer not to answer

Do you feel the need to take public speaking classes or workshops?

- ☐ Yes
- ☐ No
- ☐ Not sure (Please explain)

C.2 PRCA-PS

Please rate these statements based on how much you agree/disagree with them

	Strongly disagree	Disagree	Undecided	Agree	Strongly agree
I have no fear of giving a presentation.	☐	☐	☐	☐	☐
Certain parts of my body feel very tense and rigid while giving a presentation.	☐	☐	☐	☐	☐
I feel relaxed while giving a presentation.	☐	☐	☐	☐	☐
My thoughts become confused and jumbled when I am giving a presentation.	☐	☐	☐	☐	☐
I face the prospect of giving a presentation with confidence.	☐	☐	☐	☐	☐

C.3 TIPI

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	☐	☐	☐	☐	☐	☐	☐

C.4 Almere

Based on your interaction with the robot, please indicate to what extent you agree or disagree with each statement.

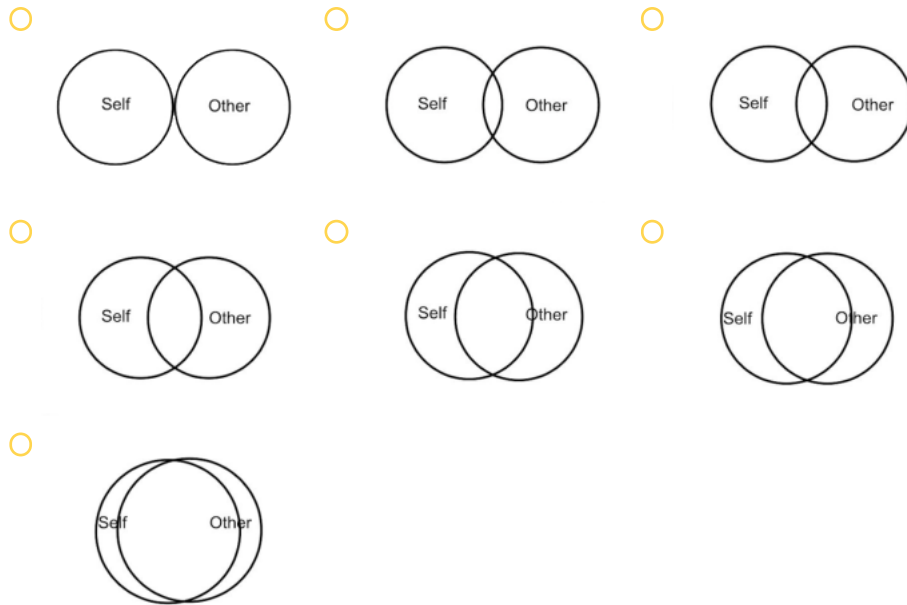
	Totally disagree	Disagree	Don't know	Agree	Totally agree
If I should use this robot, I would be afraid to make mistakes with it	☐	☐	☐	☐	☐
If I should use this robot, I would be afraid to break something	☐	☐	☐	☐	☐
I find this robot scary	☐	☐	☐	☐	☐
I find this robot intimidating	☐	☐	☐	☐	☐
I think it's a good idea to use this robot as a public speaking coach	☐	☐	☐	☐	☐
This robot would make life more interesting	☐	☐	☐	☐	☐
It's good to make the use of this robot to rehearse presentations	☐	☐	☐	☐	☐
I think I'll use this robot in the future	☐	☐	☐	☐	☐
	Totally disagree	Disagree	Don't know	Agree	Totally agree
I'm certain to use this robot in the future	☐	☐	☐	☐	☐
I plan to use this robot in the future	☐	☐	☐	☐	☐
I think this robot can be adaptive to what I need	☐	☐	☐	☐	☐
I think this robot will only do what I need at that particular moment	☐	☐	☐	☐	☐
I think this robot will help me when I consider it to be necessary	☐	☐	☐	☐	☐
I enjoy this robot talking to me	☐	☐	☐	☐	☐
I enjoy doing things with this robot	☐	☐	☐	☐	☐
I find this robot enjoyable	☐	☐	☐	☐	☐

	Totally disagree	Disagree	Don't know	Agree	Totally agree
I find this robot fascinating	☐	☐	☐	☐	☐
I find this robot boring	☐	☐	☐	☐	☐
I think I will know quickly how to use this robot	☐	☐	☐	☐	☐
I find this robot easy to use	☐	☐	☐	☐	☐
I think I can use this robot without any help	☐	☐	☐	☐	☐
I think I can use this robot when there is someone around to help me	☐	☐	☐	☐	☐
I think I can use this robot when I have a good manual	☐	☐	☐	☐	☐
I consider this robot a pleasant conversational partner	☐	☐	☐	☐	☐
	Totally disagree	Disagree	Don't know	Agree	Totally agree
I find this robot pleasant to interact with	☐	☐	☐	☐	☐
I feel this robot understands me	☐	☐	☐	☐	☐
I think this robot is nice	☐	☐	☐	☐	☐
I think this robot is useful to me	☐	☐	☐	☐	☐
It would be convenient for me to have this robot	☐	☐	☐	☐	☐

	Totally disagree	Disagree	Don't know	Agree	Totally agree
I think this robot can help me with my presentations	☐	☐	☐	☐	☐
When interacting with this robot I feel like I am talking to a real person	☐	☐	☐	☐	☐
It sometimes felt as if this robot was really looking at me	☐	☐	☐	☐	☐
	Totally disagree	Disagree	Don't know	Agree	Totally agree
I can imagine this robot to be a living creature	☐	☐	☐	☐	☐
I often think this robot is not a real person	☐	☐	☐	☐	☐
Sometimes this robot seems to have real feelings	☐	☐	☐	☐	☐
I would trust this robot if it gave me advice	☐	☐	☐	☐	☐
I would follow the advice this robot gives me	☐	☐	☐	☐	☐
I would not trust this robot if it gave me advice	☐	☐	☐	☐	☐

C.5 Inclusion of Other in the Self (IOS) Scale

Please circle the picture below that best describes your relationship with the social robot:
(Please note “Other” refers to the social robot you interacted with in the session)



C.6 Perceived Listening Behaviour and Perceived Usefulness

While I was presenting the robot was:

Not listening at all

Strongly Listening

While I was presenting the robot was:

Not Attentive at all

Very attentive

While I was presenting the robot was:

Not engaged at all

Very engaged

I found the robot as a public speaking coach:

Not competent at all

Very competent

I found the robot as a public speaking coach:

Not helpful at all

Very helpful

How well did you find the feedback provided by the robot?

Not easy to understand at all

Very easy to understand

C.7 Open-ended Questions

I think this robot can be a good coach as a public speaking coach.

Strongly disagree

Strongly agree

A horizontal slider bar with a central dot, indicating a neutral position between 'Strongly disagree' and 'Strongly agree'.

Please explain your answer.

A large rectangular text input box with a thin border and a small diagonal line in the bottom right corner.

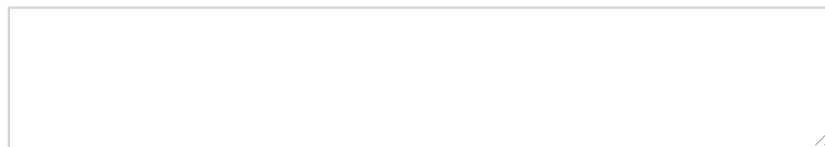
How relevant was the robot's feedback to you?

Not relevant at all

Very relevant

A horizontal slider bar with a central dot, indicating a neutral position between 'Not relevant at all' and 'Very relevant'.

Please explain.

A large rectangular text input box with a thin border and a small diagonal line in the bottom right corner.

C.8 Negative Attitudes Towards Robot Scale (NARS)

Please rate the following statements according to your opinion.

	Strongly disagree	Disagree	Undecided	Agree	Strongly agree
I would feel uneasy if robots really had emotions	☐	☐	☐	☐	☐
Something bad might happen if robots developed into living beings	☐	☐	☐	☐	☐
I would feel relaxed talking with robots	☐	☐	☐	☐	☐
I would feel uneasy if I was given a job where I had to use robots	☐	☐	☐	☐	☐
If robots had emotions, I would be able to make friends with them	☐	☐	☐	☐	☐
I feel comforted being with robots that have emotions	☐	☐	☐	☐	☐
The word “robot” means nothing to me	☐	☐	☐	☐	☐
I would feel nervous operating a robot in front of other people	☐	☐	☐	☐	☐
I would hate the idea that robots or artificial intelligences were making judgments about things	☐	☐	☐	☐	☐
I would feel very nervous just standing in front of a robot	☐	☐	☐	☐	☐
I feel that if I depend on robots too much, something bad might happen	☐	☐	☐	☐	☐
I would feel paranoid talking with a robot	☐	☐	☐	☐	☐
I am concerned that robots would be a bad influence on children	☐	☐	☐	☐	☐
I feel that in the future society will be dominated by robots	☐	☐	☐	☐	☐