

Research Article

Delara Forghani*, Moojan Ghafurian, Samira Rasouli, Chrystopher L. Nehaniv, and Kerstin Dautenhahn

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

<https://doi.org/10.1515/pjbr-2024-0004>
received May 18, 2023; accepted August 13, 2024

Abstract: The use of interactive tools, such as voice assistants and social robots, holds promise as coaching aids during public speaking rehearsals. To create a coach that is both effective and likable, it is important to understand how people perceive these agents when they observe them during actual presentation sessions. Specifically, it is important to assess people's perceptions of the agents' physical embodiment and nonverbal social behaviour, taking into account both listening and feedback periods. To this end, we conducted an online study with 168 participants who watched videos of agents acting as public speaking coaches. The study had three conditions: two with a humanoid social robot in either (1) active listening mode, using nonverbal backchannelling, (2) passive listening mode, and (3) a voice assistant agent. The results showed that the social robot in both conditions was perceived more positively in terms of its human-like attributes, and likability than the voice assistant agent. The active listener robot was perceived as more satisfying, more engaging, more natural, and warmer than the voice assistant agent, but this difference was not seen between the passive listener robot and the voice assistant agent. Additionally, the active listener robot was found to be more natural than the passive listening robot. However, there were no significant differences in perceived intelligence, competence, discomfort, and helpfulness between the three agents. Finally, participants' gender and personality traits were found to affect their evaluations of the agents. The study offered insights into general attitudes towards using social robots and voice assistants as public

speaking coaches, which can guide the future design and use of these agents as coaches.

Keywords: human–robot interaction, public speaking training, social robots, voice assistant agents, non-verbal social behaviour, backchannels

1 Introduction

Our long-term goal is to develop a social robot as a public speaking coach. To achieve this goal, it is important to understand the influence of the agent's embodiment and non-verbal behaviour, and how this impacts participants' perception of the agent.

Public speaking is an invaluable skill that is essential for academic success and career advancement, specifically when requiring a client-facing role [1–3]. Studies have highlighted the value of communication skills, and oral presentation skills in graduate studies and their role in the workplace [3,4]. For example, a study that targeted engineering students identified the need to include work-force communication skills in engineering education to support students' career success [5].

Delivering an effective presentation requires practice and rehearsals [6,7]. Although rehearsal is crucial, according to survey studies, it is frequently disregarded [7]. During rehearsals, the feedback given by a tutor or coach plays an important role in enhancing the development of public speaking skills [8]. Although individuals may hire a public speaking coach to improve their public speaking skills, there are several reasons why someone may have difficulty practising with an individual (a human coach) who can provide feedback on their presentation delivery skills. Gaining access to or affording a human coach, and dealing with time constraints associated with presentation deadlines or human coaches can pose challenges for individuals. Moreover, practicing in front of a human can be difficult for individuals with public speaking anxiety, which has been shown to be one of the most prevalent forms of social anxiety [6,9]. The development of interactive technologies,

* **Corresponding author: Delara Forghani**, Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada, e-mail: delara.forghani@uwaterloo.ca

Moojan Ghafurian, Samira Rasouli, Kerstin Dautenhahn: Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada

Chrystopher L. Nehaniv: Departments of Systems Design Engineering and of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada

such as robots and voice assistant agents as public speaking coaches can be a valuable addition to existing support systems and can aid individuals who do not have access to or may not feel comfortable with a human coach, especially during the early stages of acquiring fundamental skills.

Social robots have been used in educational settings because they can behave within the expected behavioural norms, and they possess a visual appearance and social capabilities, as well as technical capabilities (motion sensing, voice recognition, and face tracking) that make them suitable for such applications [10]. They have been used in the role of teaching assistants, tutors, and peers [11–13]. One of the most widely used social robots in education is Nao, and recently the Pepper robot [13], developed by SoftBank Robotics (<https://www.softbankrobotics.com/>). Both robots have a friendly, cartoon-like, non-threatening appearance that helps avoiding the “uncanny valley” problem [14]. In addition to social robots, voice assistants have been used in educational settings, especially in the realm of language learning, due to their ability to speak and understand multiple languages [15,16].

A number of studies have demonstrated the effectiveness of interactive systems in providing feedback on the presenter’s verbal and/or nonverbal communication skills [2,6,17–21]. Several studies have primarily focused on evaluating the efficiency of a system in delivering feedback and how this can enhance public speaking training, with a focus on the feedback mechanism and timing of feedback delivery [17–20]. Intelligent interfaces, such as Google Glass [17], user interfaces that offer real-time visual feedback [19], or designs that combine various technologies to deliver multi-modal feedback, such as combining visual and haptic feedback [20], software embedded in Microsoft PowerPoint [18], and mobile applications [22] have been explored as approaches for providing feedback during public speaking training.

A study made use of conversational agents, such as smart speakers, to engage participants in cognitive reconstruction exercises to alleviate anxiety associated with public speaking [2]. These studies focused mainly on evaluating the effectiveness of training systems and did not include comparing different types of agents with varying behaviour or appearance.

Previous research has compared the use of a social robot and a voice assistant agent as game peers to evaluate their respective effects on users’ entertainment experiences [23]. Moore and Urakami utilized three voice user interfaces (VUIs) – a voice-only system, a smart speaker, and a social robot – to investigate their distracting impact on users’ cognitive performance, and defined their respective advantages and disadvantages [24]. The voice-only VUI

had the least amount of distraction. The physical embodied VUI (smart speaker) was perceived as the calmest due to its minimalist design, and users’ ability to associate the voice with a specific source, thus resulting in a comfortable user experience. The social embodied VUI (social robot) was found to be the most likable of the three VUIs [24].

Looije *et al.* conducted a study to compare the persuasive ability as being both empathetic and trustworthy between a physical robot, a virtual robot, and a text interface while assisting older adults with their health self-management program [25]. Two modes of social and non-social behaviour were implemented for the virtual and physical robots. The social character exhibits behaviours related to three empathy skills, which are being complimentary, attentive, and compassionate using three social cues: looking, understanding, and listening. The non-social character exhibits only the behaviour related to the complimentary skill. Based on the results, the social characters were found to be more empathetic than the text interface. On the contrary, non-social characters were perceived as less trustworthy and less appreciated than the text interface [25]. Also, participants perceived the non-social physical robot as annoying [25].

The incorporation of anthropomorphism and non-verbal social behaviour may not always be needed for agents [26]. Instead, the design of an agent’s characteristics and behaviours should account for the agent’s role and user expectations, resulting in a trade-off between an agent’s sociability and its task performance [26].

As such, it is crucial to investigate how participants would perceive and evaluate different agents with varying degrees of social embodiment, anthropomorphism, and social non-verbal behaviour when functioning as public speaking coaches.

In addition to the type of agent, we investigate the impact of non-verbal backchannelling, as a proxy for active listening, on people’s perception of a coach. Backchannels are verbal or non-verbal signals sent by a listener to a speaker in order to convey understanding, attention, agreement, and willingness to continue the conversation (e.g. nodding or “hmm” utterances) [27]. Smith *et al.* [28] highlighted the importance of backchannels in human–agent interactions. Bodie *et al.* showed that in a self-disclosure session between two humans, behaviours associated with active listening, both verbal and nonverbal, indicated higher emotional awareness and resulted in a better emotional recovery [29]. In a self-disclosure scenario, Eyssel *et al.* [30] suggested that the degree of engagement in the conversation by a robot (as a passive or active listener) may influence perceived human–robot interaction more than the verbal content exchanged between the robot and the human [30].

Previous research has utilized virtual audiences and conversational agents in public rehearsal training sessions to provide listening feedback on participants' presentations [21,31]. We anticipate that investigating the adoption of active listening behaviour in an agent as a coach, using backchannelling, can contribute to the creation of future intelligent agents as coaches that can, to some extent, foster the presenter's sense of being heard and cared for.

Therefore, to understand the factors that can facilitate the accomplishment of an agent as a public speaking coach, we compared three interactive agents. Two of the agents were depicted as social robots that were either active listening or passive listening, with and without non-verbal backchannelling, whereby backchannelling was displayed through head nods during the presentation. In the feedback phase, the robot, in both conditions, expressed the same gestures and arm movements. The third agent was a voice assistant agent (a smart speaker) that did not display any backchannelling during the listening phase and provided the same verbal feedback after the presentation.¹

The remainder of the article is organized as follows: Section 2 states our research questions along with our hypotheses. Section 3 provides a background overview related to our research. Section 4 explains the methods and procedures of our study. Section 5 outlines the results of the analyses, followed by a discussion of results in Section 6. Section 7 presents limitations and future work. Finally, Section 8 summarizes our findings.

2 Research questions and hypotheses

This study aims to answer the following Research Questions (RQs):

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 backchannelling, 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 aforementioned RQs, we investigate 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 backchannelling 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.

H1 was based on previous research on the influence of human likeness of an agent [6,26,32]. **H2** was based on the effect of backchannelling on user's having more positive interaction experience with an agent [21,33,34]. **H3** was based on a previous study on interpersonal similarity and interpersonal attraction, and strong correlations were shown between the Big-Five personality traits and the God-speed questionnaire [35].

¹ Note, we did not include a condition with a human coach since our long-term goal is to develop an assistive tool for university students to improve access to a public speaking coach. Moreover, if we had chosen a control condition with a human, then specific details such as age, gender, etc., would have influenced the results due to implicit biases.

3 Background

3.1 Public speaking training

Previous studies have developed mobile apps, intelligent interfaces, and virtual conversational agents to help users with practicing public speaking skills and receiving feedback on their verbal and/or non-verbal behaviour during presentations [17,18,20]. However, most of these studies have focused on how the systems provide feedback, rather than on comparing different types of agents, and how the agents listen to the presentations – which is the focus of our study.

Presentation trainer (PT) is a multimodal tool which uses Microsoft Kinect sensor V2 to track presenters' body joints and use of filler sounds through its speech recognition tools. It also tracks the presenter's voice volume and pauses to provide multimodal real-time feedback using a graphical user interface and a haptic wristband [20]. PT was perceived to be a better method of learning presentation skills than traditional classroom settings. The feedback provided by PT was found to have a significant effect on learning non-verbal skills in presentations (posture, gesticulation, voice volume, use of pauses, making use of phonetic pauses, and not shifting weight from one foot to another) [20].

PitchPerfect is a system which was embedded in Microsoft Powerpoint and focused on content coverage and time management [18]. In addition to improving presentation quality, PitchPerfect supported learning content, managing time, and boosting confidence when giving presentations [18].

Tanveer *et al.* [17] developed a Google Glass-based interface to monitor participants' speaking rate and volume during presentation practices. They found a sparse feedback strategy to be more satisfying and less distracting than a continuous feedback stream [17].

Wang *et al.* [2] used Amazon Alexa to help participants with reducing their public speaking anxiety through guided cognitive reconstruction exercises. In this study, the agent's level of sociability was manipulated. It was shown that the agent's sociability positively affected participants' satisfaction and willingness to use the tutor again [2]. Also, the higher levels of sociability were associated with a greater sense of interpersonal closeness, and it reduced pre-speech anxiety [2].

In educational settings, social robots can serve as classroom teachers, peers, and telepresence instructors [36,37]. In this study, in addition to a voice assistant agent, we employ a social robot as a public speaking coach. Hence, this study centres on the design implications of two distinct physically embodied agents. Rather than evaluating the

effectiveness of the interactive agents in enhancing presentation performance, the study evaluates people's attitudes and perceptions towards those agents as potential public speaking coaches.

3.2 Agent comparison

Several studies compared the effect of different types of agents and/or interfaces on participants' perceptions of the agent in different contexts/tasks.

Voice is a fundamental human communication method, which is important in socially interactive agents to improve interface likability and users' impressions [37,38]. Listeners naturally respond to speech, irrespective of its source. They interpret technology-based voices and decide on appropriate behaviour using the same mental rules employed in human interactions [36–39]. Voice as a social signal can encourage people to believe that a machine has multiple distinct personalities [40]. However, non-verbal behaviour is a key factor in social coordination and signalling [40]. A human-like agent is said to lead humans to expect a greater degree of social interaction [26]. Social robots were developed to interact with humans in a “natural” manner, often inspired by how humans interact with other humans. Frequently, anthropomorphic features were used in social robots to increase their perception as a “social entity.” The ability of social robots to express non-verbal social behaviour was shown to facilitate pragmatic communication [26].

Luria *et al.* [32] compared a social robot [41] with three different interfaces for smart home control: a voice-controlled, a wall-mounted, and a mobile application. Using a microscope metaphor, they designed a social robot named Vyo based on strategies of being engaging, unobtrusive, device-like, respectful, and reassuring. As a result of an exploratory study, the advantages of the robot were found to be higher levels of enjoyment, and engagement, as well as higher situation awareness. However, the voice interface offered the advantage of being hands-free and ubiquitous [32].

Moore and Urakami [24] evaluated voice user interfaces (VUIs) through four cognitive tasks to assess their respective distracting effects, perceived social presence and perceived calm. A voice user interface uses voice as the main medium of interaction. Three VUIs were compared including 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).

During the cognitive tasks, researchers measured cognitive speed, working memory, visual memory, as well as process inhibition. Based on the results, the voice-only system had the least distracting effect on cognitive speed while having a slight negative impact on visual memory. The physical embodied system, followed by the social embodied system, caused the most distractions [24]. The most likeable VUI was the social embodied one, followed by the physical embodied VUI, and the least liked was the voice-only VUI [24]. The physical embodied VUI scored higher for perceived calm compared with the social embodied VUI. According to the authors, this may indicate that participants may feel more comfortable if designs are simplified to essential features [24].

Similarly, Kontogiorgos et al. examined humans' responses to different agents during guided cooking tasks [26]. Three devices were used: a smart speaker, a robotic head resembling a human face that interacted solely through speech, and the same robotic head that also interacted through gaze and head movements. The findings of the study suggest that it is not always beneficial for agents to adopt anthropomorphic characteristics and communicate using non-verbal cues [26].

Laban et al. demonstrated that in a disclosure setting with a human, a social robot, and a voice assistant, the agent's embodiment is more important than the topic of disclosure [42]. According to their findings, agents' embodiment influenced how people would perceive disclosures, how much information they disclosed, and how they would communicate their disclosures [42]. The number of words participants used when speaking to the human and their disclosure duration was significantly more than when speaking to the social robot or the voice assistant. Thus, when it comes to disclosure, people would prefer a human embodiment since they are more familiar with it, and it enables maximum disclosure [42]. While people perceived the differences between the two artificial agents (the social robot, and the voice assistant agent), the amount of information they disclosed, as well as the perception of the quality and quantity of their disclosures were not affected by the embodiment of the agents [42]. Similarly, Barfield showed a clear inclination among participants to share personal information with a human counsellor, irrespective of the nature of the information. However, when considering the content of self-disclosure, the data also revealed that participants were willing, to some degree, to disclose personal information to a robot with a friendly appearance or a female android, more so than to a robot perceived as lacking emotional responsiveness [43].

In the context of public speaking, Trinh et al. introduced RoboCOP as a coach to give verbal feedback in a presentation practice session on slides and the overall

performance of the presenter [6]. In a within-participants design, they used an anthropomorphic robotic head [44] to provide spoken feedback on three aspects of a presentation: content coverage, connecting to the audience, and speech quality (e.g. speaking rate, filler rate, and pitch variety). The robot was compared with visual-only feedback using a graphical user interface, and voice-only feedback without any interactive tool [6]. Results showed improvements in the presentation experience with the robot in comparison to the visual-only and voice-only feedback [6]. Participants who were afraid of public speaking found RoboCOP to be more comfortable than practising in front of a live audience [6].

Note, the study employed an anthropomorphic robot head that lacked the ability to make gestures during interactions. Also, voice-based feedback was provided without explicitly identifying a device as the source of the voice [6]. However, according to Lee et al., users who are unfamiliar with this type of interaction may experience discomfort and privacy concerns in the absence of a physical body [45].

3.3 Backchannels

Backchannels are verbal or non-verbal conduits for brief cues emitted by a listener without interrupting the turn of the speaker and indicate attentiveness and listenership to the speaker [46]. According to Flint [47], backchannels are not intended to answer questions, interrupt the speaker, or elicit a particular reaction in the speaker. They are normally used to show agreement [48], involvement [49], surprise [50], and understanding [51]. Dennis and Kinney [52] mentioned that backchannelling can perform four cognitive functions, including signalling understanding, signalling confusion, providing clarification for a message, and completion of the sentence. Nonverbal backchannels are mostly characterized by body language [47] notably nods, shoulder shrugs [53], smiles [54], and gaze [55]. Cathcart et al. [56] found that backchannels are usually preceded by a short pause in the speaker's utterance.

Gratch et al. [33] generated listening behaviour in a virtual agent based on the inputs received from a speaker, such as their head movements, gaze directions, and acoustic features. They found that the virtual agent that provided positive listening feedback (such as head nods) could provide more engagement and foster a stronger rapport than a human listener.

Sidner et al. [34] defined engagement as the process by which individuals involved in an interaction initiate, maintain, and terminate their perceived connection with each

other. They showed that using backchannelling behaviour in a robot made the interaction more engaging for participants in a collaborative task with the robot [34]. Jung *et al.* [57] evaluated the role of backchannelling and task complexity in human–robot teamwork for an Urban Search and Rescue (USAR) task. For this study, two upper-torso humanoid robots were used, Nexi and Maddox. In the backchannelling condition, the robot would turn towards the participant and nod in parallel with the participant's utterance in a smooth, quick motion. Based on the results of a between-participants study design with 96 participants, backchannelling during the complex task improved team performance and increased the perception of engagement in the robot. However, it led to a reduction in the perceived competence of the robot [57].

Head nods as a non-verbal backchannel convey attention and agreement; however, certain features such as their amplitude, frequency and pace may result in different interpretations and meanings [58,59]. Bousmalis [58] surveyed head nods as almost-universal signals of agreement, although they can have different meanings and functions depending on their amplitude, number of cycles, and duration, as well as the context in which they occur [58]. For instance, nods of greater amplitude and greater frequency indicated affirmation, while smaller ones indicated active participation in the conversation [60,61].

Oertel *et al.* [59] reported head nods generally conveyed more attentiveness than audio backchannels. By using a corpus of recorded human–human interactions, they identified 77 different head nods based on their duration, the number of oscillations, frequency, amplitude, maximum upward/downward speed oscillations, etc. [59]. Furthermore, different nodding forms could be perceived as more or less attentive and combining multiple modalities increased perceived attentiveness [59].

Recently, backchannel generation models have been tested to predict the timing of backchannels to increase the flow of the conversation with an agent [62–64].

Park *et al.* studied a rule-based backchannelling model in a storytelling scenario involving 4–6-year-old children [64]. Their backchannel opportunity prediction (BOP) model detects four speaker cues and generates backchannels. By using the BOP model robot, they found an increase in perceived attentiveness. Also, children preferred to tell stories to the BOP model robot [64].

Murali *et al.* [21] designed a virtual agent to support a student in oral presentations by expressing attentive listening through head nods, smiling, and raising eyebrows when the presenter looked at it and showing an “OK” gesture whenever the presenter moved to the next slide. Both

self-reports and physiological measures (heart rate (HR) and heart rate variability (HRV)) revealed significant reductions in public speaking anxiety among participants compared to a control condition with only a text display [21].

The aforementioned studies showed how backchannel cues can be used by social agents and how they are interpreted by people interacting with the agents. Nevertheless, in the context of public speaking, the perceptions of non-verbal backchannelling indicating active listening have, to the best of our knowledge, not been studied. The present study investigates how participants assess the robot coach responding through head nods during the presentation as a proxy for active listening, compared to a robot that does not use backchannelling and a voice assistant agent without specific non-verbal behaviour.

4 Method

We conducted an online video study using Amazon Mechanical Turk (MTurk). We used an online tool developed using HTML, JavaScript, SQL, and the Go programming languages that was used and tested in previous online studies in our laboratory, and modified for the purpose and contents of our study. To address our research questions, we compared three experimental conditions with different agents as public speaking coaches. In one, we used a voice assistant agent and in the other two we used a social robot. (a) **VoiceAgent**: The Google Nest Mini was used and gave verbal feedback after the presentation (Google Nest mini, https://store.google.com/product/google_nest_mini?hl=en-US&pli=1), (b) **ActiveListenRobot**: the Pepper robot was used, which showed non-verbal backchannelling during the presentation and gave verbal feedback accompanied by hand gestures after the presentation during the feedback phase (c) **PassiveListenRobot**: the Pepper robot was stationary during the presentation and provided verbal feedback accompanied by gestures after the presentation during the feedback phase. The experiment had a between-participant design and participants were randomly assigned to one of the three conditions.

4.1 Procedures and measures

After reading and consenting to the study information form, participants proceeded to follow these steps:

Step 1 – Demographics and pre-experimental questionnaires: Participants were asked about their age and gender. In addition, we inquired about any past attendance of public speaking classes and workshops, as well as their

perceived need to attend such events. Finally, participants were asked about their Big 5 personality traits [65]. Answering all demographic questions was optional for participants as required by our Human Research Ethics Board.

Step 2 – Watching the video: Participants were given instructions on watching a video and answering some questions about it. The video showed a person presenting on “History of Canada Day.” The content was taken from the official website of the Government of Canada (canada.ca). The presenter presented for approximately 1 min and 30 s (note: the agent showed different behaviours during this period according to the experimental condition, which will be described in Section 4.2). During the presentation, slides related to the presentation content appeared in the upper left corner of the video and were changed appropriately as the student spoke (Figure 1). As soon as the student finished their presentation, the agent provided feedback to the student for approximately 2 min.

The agent (as the coach) began the feedback by expressing appreciation for the presentation and how interesting it was. Feedback focused primarily on vocal modulation skills and acoustic features of the presentation (e.g., speech rate, volume, pitch variety, rising intonations during sentences, and falling intonations at the end of sentences). The

agent pointed out positive aspects of the presentation (e.g. “You demonstrated a good speech rate in your presentation! I liked the pitch variety in your speech!”) followed by suggestions for improving the speech (e.g. “Please avoid using fillers! Try not to use pauses too often or for too long! Try not to speak too fast or too slow!”). Finally, the agent expressed appreciation for the presenter’s efforts.

Once the agent’s feedback was finished, the video came to an end. After that, participants could choose to either replay the video or proceed to answer the questions. To ensure that participants did not engage in concurrent tasks while the video was being played, the video playback was coded to freeze if the current window or tab was minimized or changed, and to resume playing as soon as the window or tab was refocused.

The presenter’s and agent’s audios, the presentation content, and the feedback content were identical in all conditions. They were pre-recorded once and used while recording the video. For the agent’s voice, we utilized the Pepper robot’s voice. We modified Pepper’s pitch and frequency to make it sound more like an adult voice as it is originally designed to sound childlike.

The camera used to capture the video was positioned at the same location, behind the presenter’s head as depicted in Figure 1, throughout all experimental conditions, to ensure that the presenter and their age etc. remained unidentified to avoid any potential biases.

Step 3 – Post-experimental questionnaires: To assess participants’ perceptions of the agents as coaches, we used several questionnaires, some of which were adapted from standard questionnaires used in human–robot interaction (HRI) studies. Below is a detailed description of each of the questionnaires used in the study. Our aim was to evaluate users’ interaction experiences with the social robot and the voice assistant agent in different experimental conditions. Specifically, we assessed whether the agents could perform effectively as public speaking coaches (Questionnaire 1), how human-like and machine-like they behaved (Questionnaire 2), what social attributes can be attributed to them (Questionnaire 3), and the level of intelligence and likability they were perceived to have (Questionnaire 4). The questionnaires were presented with either their defined standard scale (Questionnaires 3 and 4) or a continuous scale ranging from 0 to 1,000 for those that did not have a defined standard scale (Questionnaires 1 and 2). This approach was taken because studies have shown that continuous scales have benefits over Likert-point scales in web-based research and online survey studies [66–68].

Questionnaire 1 – agent behaviour: Here, we specifically targeted participants’ attitudes about the agents as public speaking coaches and their behaviour while interacting with the student. The questionnaire consisted of 11

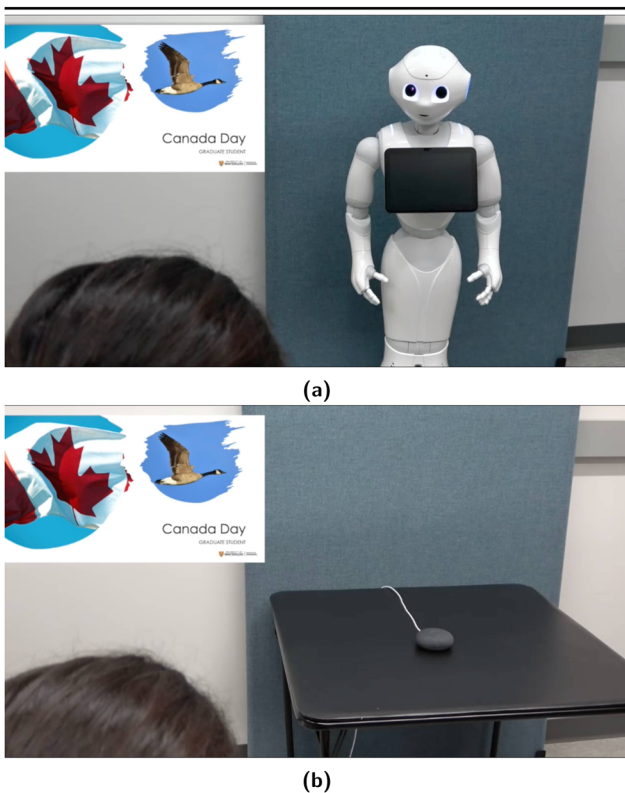


Figure 1: Picture of the experimental setup with (a) the social robot, and (b) the voice assistant agent.

Table 1: Questions designed to assess participants' perceptions of the agent's behaviour as a public speaking coach. Answers were given on a continuous scale. Wording used in at the both ends of the scales are 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)

questions which are listed in Table 1. Answers were provided on a continuous scale (with input data collected ranging from 0 to 1,000)².

Questionnaire 2 – human nature (HN) attributes: Haslam *et al.* [69] suggested that there are two distinct perceptions of humanness: Human Nature (HN) and Human Uniqueness (HU). There are specific attributes associated with each of them, and the absence of those attributes indicates dehumanization or denial of humanness. Human uniqueness attributes distinguish humans from non-human animals and describe what makes us human. They are civility, refinement, moral sensibility, higher cognition, and maturity. The human nature attributes are emotionality, warmth, openness, agency (liveliness), and depth. The human nature features do not distinguish humans from non-human animals, rather they are essential to humans, that is when something is denied in human nature, it appears machine-like [69]. These items have been previously used to assess people's perception of HN and HU attributes of a virtual agent [70]. We used the same questions as in [70] for evaluating the HN dimensions of the agents in this article and asked participants to rate the agent as a coach on four dimensions of HN, i.e., emotionality, warmth, openness, and agency (liveliness). Similar to the previous questionnaire, answers were given on a continuous scale (the values obtained ranged from 0 to 1,000).

Questionnaire 3 – the robotic social attributes scale (RoSAS): RoSAS is a standardized HRI metric that was developed based on the Godspeed questionnaires [71] and social psychological research on social perceptions of robots [72]. It encompasses three social dimensions, namely, warmth³, competence, and discomfort, measured by 18 attributes. A

significant advantage of RoSAS is that it is designed for a wide range of robots. Also, it is not intended to replace any previous HRI measurements, such as Godspeed [72], since many items are counted in Godspeed that are not included in RoSAS [72]. In this study, RoSAS was used to measure participants' perceptions of the agent as a coach on these social attributes, i.e. warmth, competence, and discomfort.

RoSAS does not suggest a specific scale; however, the authors recommended having a neutral value, such as an uneven number of possible responses to Likert-type items. In this study, we used a seven-point Likert scale. Studies [73,74] also utilized RoSAS to assess how participants perceived the robot's social behaviour in terms of warmth, competence, and discomfort.

Questionnaire 4 – perceived intelligence and likeability (Godspeed): The Godspeed questionnaire assesses anthropomorphism, likeability, animacy, perceived intelligence, and perceived safety [71]. It can be used as a complement to RoSAS when specific aspects cannot be fully measured by RoSAS [71,72]. In this study, we only measured perceived intelligence and likeability of agents. Safety, anthropomorphism, and animacy aspects were not related to our research questions and thus considered not relevant for this study. Note, in one of our conditions, the agent was a voice assistant that did not move in a goal-directed manner and did not have anthropomorphic features. Furthermore, previous research on the perception of animacy suggested that animacy is influenced by physical interactions with robots rather than observing/watching HRI scenarios [75]. Consequently,

² Since we did not find a standard questionnaire that addressed these items exactly, we created this questionnaire ourselves.

³ In RoSAS, warmth is indirectly assessed based on six attributes (organic, sociable, emotional, compassionate, happy, feeling); however in HN, see above, warmth is directly rated on a pre-determined scale.

we considered only the subscales for perceived intelligence and likeability.

To make the questionnaires more readable, each of the questionnaires described earlier was placed on a separate page in the online interface and participants completed each questionnaire before moving on to the next. At the end of each questionnaire, an attention check was included (e.g., “The presentation topic was about a city in Canada: True – False”). All attention checks were True-False questions. Participants' remunerations were not affected by their performance in the attention checks. However, in our analysis, we excluded data from those who had more than two incorrect answers out of five attention checks.

4.2 Experimental conditions

For the two robot conditions, the Pepper robot was placed in front of the student while orienting its head and torso towards the student (Figure 1(a)). For the voice assistant condition, a Google Nest Mini was placed on a table in front of the student (Figure 1(b)). As discussed before, we controlled for different factors that could influence participants' perceptions of the agents, such as using identical audio in all three conditions and positioning the camera in a way so that the “student presenter” in the video was not identifiable.

While the verbal behaviour of the agents were identical, depending on the experimental condition, the non-verbal behaviour of the agents varied as follows:

Condition 1 – ActiveListenRobot: During the presentation, the Pepper robot employed non-verbal backchanneling as a means of demonstrating active listening by head nodding. During the creation of the video, using the Wizard of Oz method, one of the researchers moderated the timing of the head nods of the Pepper robot. The researcher used the end of utterances and small pauses during the presenter's speech as a basis for making the robot nod. Note that the study did not examine how the timing of the nods might have been perceived, and they were only added based on researchers' discretion.

According to Oertel et al. [59], different amplitudes and frequencies can reflect different meanings for head nods. To reduce the chance that a specific amplitude or frequency may affect the outcome, we applied two modes of nodding in a random order as backchanneling in the robot. One type involved a slow vertical movement of the robot's head, while another involved two sequential vertical movements with a relatively faster speed and smaller amplitude of the robot's head joint.

In the video, after the presentation and during the feedback phase, the robot used non-verbal gestures (identical for the two robot conditions). We designed the gestures and arm movements of the Pepper robot using the [Choregraphe] (<http://doc.aldebaran.com/2-5/software/choregraphe/index.html>) software and adjusted their timing in accordance with the verbal content. The gestures were added to make use of the expressive capabilities of the robot using its actuators and to distinguish between the physical embodiment of the voice assistant agent and the social robot.

Condition 2 – PassiveListenRobot: During the presentation, the Pepper robot maintained a fixed head position, stayed still, and did not show any backchanneling behaviour. The robot's position and orientation with regards to the “student presenter” were the same as in Condition 1.

After the presentation and during the feedback phase, hand gestures, arm movements, and their timings were identical to Condition 1.

Condition 3 – VoiceAgent: The participants saw a Google Nest Mini device placed on a table in front of the student. During the presentation, Google Nest was silent and in listening mode. Google Nest turns on four lights when it is in listening mode. After the presentation, these lights were turned off and Google Nest began providing the same verbal feedback as in the previous two conditions. The lights were kept to make the interactions with the voice assistant similar to the common behaviour of Google Nest (which could affect those who were familiar with the device). In the video, the lights were not too bright to cause unnecessary distractions (Figure 2), but we hoped that they were noticeable to those who might have paid attention.

4.3 Participants

A total number of 184 participants were recruited using the MTurk platform. A power analysis for our setup with the effect size of 0.25, the significance level of $\alpha = 0.05$ and the statistical power of ($\beta = 0.95$) suggested the sample size of 132. As part of our inclusion criteria for MTurk workers, we only recruited from Canada and the United States and considered an approval rate of at least 80% based on at least 50 HITs. There were 130 participants from Canada and 54 participants from the United States. Among 184 participants, 16 records were discarded due to having more than two failed attention checks out of five. As a result, there were a total of 168 participants, 38 of whom were from the United States, and 130 from Canada (66 self-identified as female, 100 as male, and 2 preferred not to share; age range: 19–67 years). There were 58 participants in the ActiveListenRobot condition (28 self-identified as female, 28



Figure 2: A display of the lights of the Google Nest in the video.

as male, and 2 did not share; age range: 20–67 years, mean age: 36.24), 56 participants in the PassiveListenRobot condition (18 self-identified as female, 38 as male; age range: 23–67 years, mean age: 35.32), and 54 in the VoiceAgent condition (20 self-identified as female, 34 as male; age range: 21–59 years old, mean age: 35.20). Figure 3 shows the age distributions of the participants in the three experimental conditions.

The study was expected to take around 20 min and participants received 4 USD for completing the study.

Participants were restricted to using one of three browsers (Chrome, Safari, or Firefox) that had been thoroughly tested with our interface. Furthermore, we restricted participants to using laptops or desktop computers to interact with the system, since we wanted them to pay attention to videos and questions on larger screens rather than on small smartphone screens. All elements on

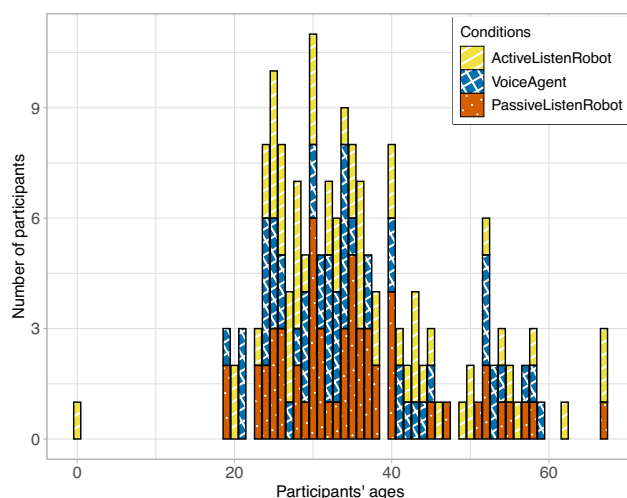


Figure 3: The histogram plot depicts participants' age distributions across different conditions.

each page were always adapted to the size of the participant's screen.

Table 2 indicates the number of participants in each condition who had attended previous classes on public speaking. It also indicates whether they felt the need to attend public speaking classes or workshops. The remaining participants chose not to disclose any information.

4.4 Statistical analysis

We used Generalized Linear Models (GLMs) [76] considering a Gaussian family to investigate the significant effects of experimental conditions and other factors, as well as to account for the potential confounding factors (e.g., those collected during the pre-experimental phase) on the dependent variables measured through the questionnaires. Even though some questionnaires possess an ordinal nature, the response or dependent variables were derived by averaging the scores of multiple question items in the questionnaire. As a result, we were able to treat them as continuous variables, making them eligible to be used in a GLM. The GLMs were simplified step-wise to use a subset of the total factors to create a model that minimized Akaike Information Criterion (AIC) [77]. Please note that we confirmed our p -values with a more conservative Bonferroni–Holm post-hoc analysis for controlling the family-wise error rate (FWER) [78]. The Bonferroni–Holm correction decreases the risk of finding significance by chance due to the multiple dependent measures associated with our Hypotheses [79]. Most of the p -values were significant after correcting for the FWER. For the few cases where the p -value exceeded the significance threshold of $\alpha = 0.05$ after the Bonferroni–Holm adjustment, we reported both pre- and post-adjustment p -values.

Table 2: Participants' answers regarding previous public speaking classes and workshop experiences. The first number indicates the number of participants, and the second one indicates the percentage of participants

Experimental condition	Previous experience		Sense of need	
	Yes (%)	No (%)	Yes (%)	No (%)
VoiceAgent	16–29.63	36–66.66	20–37.04	34–62.96
PassiveListenRobot	14–25	40–71.43	23–41.07	32–57.14
ActiveListenRobot	18–31.03	37–63.79	25–43.10	34–58.62

5 Results

This section explains the results of the analysis of participants' responses to questionnaires, and we will address the RQs while explaining the statistical analyses.

5.1 Behaviour evaluation of the agent as a coach

Participants' evaluations of the agents' behaviours as coaches (addressing RQ1-1 and RQ2-1) are shown in Figure 4 (the corresponding questions are shown in Table 1). Table 3 shows the results of a generalized linear model (GLM) predicting the ratings on each item according to the experimental condition.

Participants rated the ActiveListenRobot as more satisfying ($se = 42.47, t = 2.642, p < 0.05$) than the VoiceAgent. According to the results before applying the Bonferroni-Holm correction, the ActiveListenRobot was perceived as more satisfying than the PassiveListenRobot ($se = 41.99, t = 2.041, p < 0.05$); however, the post hoc Bonferroni-Holm correction revealed the difference exceeded the significance threshold of ($\alpha = 0.05, p = 0.08$). Participants perceived the ActiveListenRobot as more engaging ($se = 43.41, t = 2.535, p < 0.05$) than the VoiceAgent. In terms of

naturalness, participants perceived the ActiveListenRobot as more natural than the VoiceAgent ($se = 46.98, t = 3.535, p < 0.001$). The post hoc Bonferroni-Holm correction confirmed that the difference was significant ($p < 0.01$). Also, the ActiveListenRobot was perceived as more natural than the PassiveListenRobot ($se = 46.62, t = 2.466, p < 0.05$) (Table 3). Moreover, participants perceived the ActiveListenRobot as more humane than the VoiceAgent ($se = 52.363, t = 1.986, p < 0.05$). A Bonferroni-Holm post hoc analysis revealed that perceived humaneness was not significant in the ActiveListenRobot condition compared to the VoiceAgent condition. Please note that we reported the pre-adjustment p -value for humaneness in Table 3.

We did not find any significant difference between the three agents considering other items in the Questionnaire 1.

The results of the study provide partial support in favour of H1-1 and H2-1. Specifically, participants rated the ActiveListenRobot more positively than the VoiceAgent in terms of participants' feeling of satisfaction, agent's engagement, and agent's naturalness, thereby supporting H1-1. Participants also perceived the ActiveListenRobot as a more natural coach than the PassiveListenRobot, thus partially supporting H2-1.

As supported by the results shown in Tables 3 and 4, and as a response to RQ3, we found a positive correlation between the agreeableness personality trait with perceptions

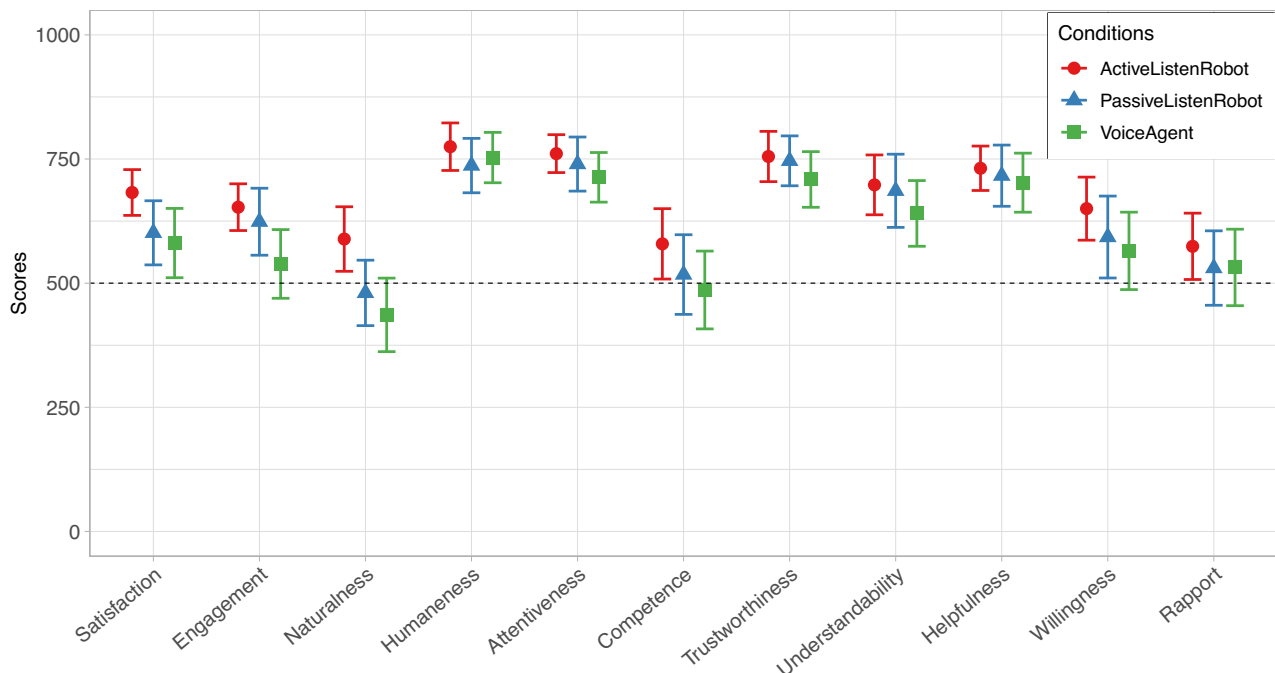


Figure 4: Generalized linear model predicting participants' scores regarding evaluations of agents. The dashed line specifies the neutral choice at the centre of the continuous scale (data obtained ranging from 0 to 1,000). Error bars indicate 95% confidence intervals.

of naturalness ($se = 18.501$, $t = 2.044$, $p < 0.05$), humaneness ($se = 19.816$, $t = 2.719$, $p < 0.01$), attentiveness ($se = 13.14$, $t = 2.148$, $p < 0.05$), and competence ($se = 12.27$, $t = 2.084$, $p < 0.05$) in the agent as a coach. Participant's willingness to use the agent in the future ($se = 19.76$, $t = 2.073$, $p < 0.05$), and the level of rapport they reported feeling with the agent ($se = 19.04$, $t = 2.186$, $p < 0.05$) had positive correlations with their agreeableness personality trait.

The conscientiousness personality trait had a significant negative effect on the perceptions of naturalness ($se = 16.861$, $t = -3.478$, $p < 0.001$) and humaneness ($se = 18.177$, $t = -2.996$, $p < 0.01$) of the agent as a coach. It was also significantly and negatively correlated with participants' satisfaction with the agent as a coach ($se = 15.016$, $t = -2.534$, $p < 0.05$), their willingness to use it in the future ($se = 17.89$, $t = -2.801$, $p < 0.01$), and the level of rapport reported to feel with the agent ($se = 17.05$, $t = -3.233$, $p < 0.01$).

Furthermore, participants' emotional stability had a positive correlation with perception of satisfaction ($se = 13.506$, $t = 2.648$, $p < 0.01$).

The results regarding correlations between ratings and Big Five personality traits do not depend on the experimental conditions, as the two-way interaction between Big Five personality traits and the experimental conditions was not significant. These findings support H3.

Participants' ages appears to have negative correlations with perceptions of naturalness ($se = 1.733$, $t = -2.134$, $p < 0.05$).

We found correlations between participants' genders with experimental conditions. These results were based on analyzing two-way interactions between the gender factor and the experimental conditions factor, and are not shown in a Table. Male participants found the VoiceAgent significantly less engaging ($se = 43.41$, $t = -2.535$, $p < 0.05$) than the other two agents. Male participants' willingness to use the ActiveListenRobot was significantly higher compared to the PassiveListenRobot ($se = 105.33$, $t = 2.178$, $p < 0.05$). Also, participants who self-identified as male felt higher rapport towards the ActiveListenRobot in comparison to the PassiveListenRobot ($se = 100.95$, $t = 2.802$, $p < 0.001$). Regardless of experimental conditions, male participants tended to find the agent's feedback significantly less trustworthy than those who self-identified as female ($se = 30.59$, $t = -2.266$, $p < 0.05$). Female participants found the agents to be significantly more attentive ($se = 30.30$, $t = 2.042$, $p < 0.05$), and were willing to use it in the future significantly more ($se = 43.89$, $t = 2.330$, $p < 0.05$). They also reported significantly higher levels of rapport with the agents ($se = 42.29$, $t = 2.052$, $p < 0.05$).

5.2 Perception of human nature attributes of the agents

We examined the four human nature (HN) attributes, namely emotionality, warmth, openness, and agency (liveliness) for each agent.

According to the results (addressing RQ2-1 and RQ2-2), perceptions of HN attributes were significantly influenced by the presence of the social robot (ActiveListenRobot and PassiveListenRobot conditions). As shown in Table 5, and Figure 5. In comparison to VoiceAgent, both ActiveListenRobot and PassiveListenRobot were rated significantly higher in the four HN attributes. According to Table 5 compared to the VoiceAgent, the ActiveListenRobot and PassiveListenRobot were rated higher on emotionality (ActiveListenRobot: $se = 46.19$, $t = 3.193$, $p < 0.01$; PassiveListenRobot: $se = 46.12$, $t = 2.308$, $p < 0.05$), warmth (ActiveListenRobot: $se = 46.47$, $t = 3.701$, $p < 0.001$; PassiveListenRobot: $se = 46.39$, $t = 2.434$, $p < 0.05$), openness (ActiveListenRobot: $se = 49.59$, $t = 2.939$, $p < 0.05$; PassiveListenRobot: $se = 49.51$, $t = 2.199$, $p < 0.05$), and agency (ActiveListenRobot: $se = 44.15$, $t = 2.784$, $p < 0.05$; PassiveListenRobot: $se = 44.08$, $t = 2.283$, $p < 0.05$). The post hoc Bonferroni-Holm analyses confirmed these significant effects.

The results support H1-2, i.e. that the social robots in the ActiveListenRobot and PassiveListenRobot conditions were perceived as being more human-like based on the Human Nature attributes, compared to the voice assistant agent (VoiceAgent). H2-2 was not supported as we did not find a significant difference between ActiveListenRobot and PassiveListenRobot on perceptions of HN attributes.

Regardless of the experimental condition, there was a gender bias in perceptions of HN. Participants who self-identified as male tended to perceive the agent significantly less emotional ($se = 39.50$, $t = -2.790$, $p < 0.01$), less warm ($se = 39.74$, $t = -3.055$, $p < 0.01$), and less open ($se = 43.24$, $t = -3.051$, $p < 0.01$), compared to participants who self-identified as female.

In response to RQ3 and in support of H3, both agreeableness and conscientiousness show significant correlations with perceptions of HN attributes, with agreeableness having a positive correlation (emotionality: $se = 17.60$, $t = 3.321$, $p < 0.01$, warmth: $se = 17.70$, $t = 3.005$, $p < 0.01$, agency: $se = 16.82$, $t = 2.888$, $p < 0.01$) and conscientiousness having a negative correlation (emotionality: $se = 15.88$, $t = -3.038$, $p < 0.01$, warmth: $se = 15.98$, $t = -2.185$, $p < 0.05$, agency: $se = 15.18$, $t = -1.979$, $p < 0.05$).

Table 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 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)

Covariate	Satisfaction				Engagement			
	Estimate	SE	<i>t</i>	<i>p</i>	Estimate	SE	<i>t</i>	<i>p</i>
Condition								
ActiveListenRobot	112.197	42.47	2.642	<0.05*	110.04	43.41	2.535	<0.05*
PassiveListenRobot	26.492	42.276	0.627	0.532	78.78	43.89	1.795	0.074
Age								
Age	-2.184	1.557	-1.403	0.162				
Gender								
Male	-68.647	35.836	-1.916	0.057				
TIPI								
Agreeableness					23.18	15.77	1.470	0.143
Conscientiousness	-38.048	15.016	-2.534	<0.05*				
Emotional stability	35.771	13.506	2.648	<0.01**				
Condition								
ActiveListenRobot	166.080	46.98	3.535	<0.01**	103.999	52.363	1.986	<0.05*
PassiveListenRobot	51.135	46.867	1.091	0.277	46.471	53.065	0.876	0.382
Age								
Age	-3.697	1.733	-2.134	<0.05*	-2.929	1.906	-1.537	0.126
Gender								
Male	-75.161	41.014	-1.833	0.069				
TIPI								
Extraversion					21.558	13.575	1.588	0.114
Agreeableness	37.818	18.501	2.044	<0.05*	53.885	19.816	2.719	<0.01**
Conscientiousness	-58.65	16.861	-3.478	<0.001***	-54.461	18.177	-2.996	<0.01**
Emotional stability	27.792	15.302	1.816	0.071				

* = $p < 0.05$; ** = $p < 0.01$; *** = $p < 0.001$.

5.3 Perceived social attributes of the agent

In response to RQ3-1 and RQ3-2, we trained a GLM on the ratings of the RoSAS questionnaire. Results are shown in Figure 6 and Table 6.

According to the results, the ActiveListenRobot ($se = 0.248$, $t = 2.796$, $p < 0.05$) was perceived as warmer than the VoiceAgent. The PassiveListenRobot was rated higher on the warmth dimension than the VoiceAgent ($se = 0.247$, $t = 2.045$, $p < 0.05$); however, a post hoc Bonferroni-Holm correction revealed that perceived warmth between the PassiveListenRobot and the VoiceAgent was not significant ($se = 0.247$, $t = 2.045$, $p < 0.1$) (please note that we reported the pre-adjustment p -value for warmth in Table 6). Please note that warmth is one of the HN attributes that is explicitly measured as one of the four attributes; however, in RoSAS warmth is a social dimension that is calculated based on six attributes in the

questionnaire (organic, sociable, emotional, compassionate, happy, feeling). We did not find a significant difference between the VoiceAgent and the social robot in the other two conditions in terms of competence (ActiveListenRobot: $se = 0.221$, $t = 0.792$, $p = 0.429$; PassiveListenRobot $se = 0.223$, $t = 0.649$, $p = 0.517$) and discomfort (ActiveListenRobot: $se = 0.195$, $t = 0.711$, $p = 0.478$; PassiveListenRobot: $se = 0.197$, $t = 0.967$, $p = 0.335$). Therefore, the results partially support H1-3.

However, H2-3 was not supported as we did not find any significant difference between the ActiveListenRobot and PassiveListenRobot in terms of the social dimensions of RoSAS.

Responding to RQ3, the two personality traits agreeableness and conscientiousness had significant effects on participants' assessments of the agent, partially supporting H3-3. Agreeableness had a positive significant effect on perceived social warmth ($se = 0.097$, $t = 3.387$, $p < 0.001$),

Table 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 AIC

Covariate	Attentiveness				Competence			
	Estimate	SE	<i>t</i>	<i>p</i>	Estimate	SE	<i>t</i>	<i>p</i>
Gender								
Male	−61.87	30.30	−2.042	<0.05*	−55.43	28.30	−1.959	.052
TIPI								
Agreeableness	28.22	13.14	2.148	<0.05*	25.58	12.27	2.084	<0.05*
Covariate	Willingness				Rapport			
	Estimate	SE	<i>t</i>	<i>p</i>	Estimate	SE	<i>t</i>	<i>p</i>
Gender								
Male	−102.26	43.89	−2.330	<0.05*	−86.80	42.29	−2.052	<0.05*
TIPI								
Agreeableness	40.96	19.76	2.073	<0.05*	41.62	19.04	2.186	<0.05*
Conscientiousness	−50.11	17.89	−2.801	<0.01**	−55.12	17.05	−3.233	<0.01**
Openness to Experiences	28.64	16.63	1.722	0.087				

* = $p < 0.05$; ** = $p < 0.01$; *** = $p < 0.001$.

Table 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 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)

Covariate	Emotionality				Warmth			
	Estimate	SE	<i>t</i>	<i>p</i>	Estimate	SE	<i>t</i>	<i>p</i>
Condition								
ActiveListenRobot	147.49	46.19	3.193	<0.01**	172.01	46.47	3.701	<0.001***
PassiveListenRobot	106.43	46.12	2.308	<0.05*	112.91	46.39	2.434	<0.05*
Gender								
Male	−110.23	39.50	−2.790	<0.01**	−121.42	39.74	−3.055	<0.01**
TIPI								
Agreeableness	58.43	17.60	3.321	<0.01**	53.20	17.70	3.005	<0.01**
Conscientiousness	−48.26	15.88	−3.038	<0.01**	−34.91	15.98	−2.185	<0.05*
Covariate	Openness				Agency			
	Estimate	SE	<i>t</i>	<i>p</i>	Estimate	SE	<i>t</i>	<i>p</i>
Condition								
ActiveListenRobot	145.74	49.59	2.939	<0.05*	122.93	44.15	2.784	<0.05*
PassiveListenRobot	108.86	49.51	2.199	<0.05*	100.63	44.08	2.283	<0.05*
Gender								
Male	−131.91	43.24	−3.051	<0.01**	−67.31	37.76	−1.783	0.076
TIPI								
Agreeableness	34.69	19.43	1.785	0.076	48.57	16.82	2.888	<0.01**
Conscientiousness	−28.26	17.82	−1.586	0.115	−30.04	15.18	−1.979	<0.05*
Emotional stability	29.45	16.01	1.840	0.068				

* = $p < 0.05$; ** = $p < 0.01$; *** = $p < 0.001$.

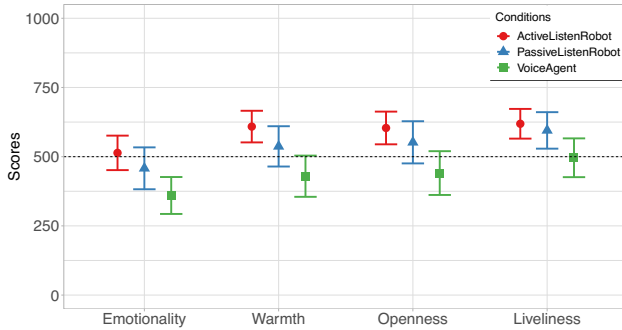


Figure 5: Human Nature scores of the agent in different conditions. The dashed lines specify the neutral choice at the centre of the continuous scale (ranging from 0 to 1,000). Error bars indicate 95% confidence intervals.

and on perceived social competence ($se = 0.077$, $t = 2.583$, $p < 0.05$). Conscientiousness had a significant negative effect on perceived social warmth ($se = 0.089$, $t = -3.541$, $p < 0.001$). Emotional stability had a significant negative effect on perceived social discomfort ($se = 0.06$, $t = -3.315$, $p < 0.05$).

Participants who identified as male rated on the agent's social warmth significantly lower than those who identified as female ($se = 0.216$, $t = -2.954$, $p < 0.01$), and on perceived social discomfort males scored significantly higher than females ($se = 0.166$, $t = 2.083$, $p < 0.05$).

5.4 Perceived intelligence and likeability of the agent

The results on Godspeed perceived intelligence and Godspeed likeability are shown in Figure 7.

Addressing RQ1-4, and RQ2-4, we could not support H1-4 and H2-4 our data did not show significant differences between agents on the perceived intelligence. Having the VoiceAgent as the baseline the predicted results of the GLM model for the ActiveListenRobot were ($se = 0.180$, $t = 1.093$, $p = 0.276$), and for the PassiveListenRobot were ($se = 0.182$, $t = 0.844$, $p = 0.4$). According to Table 7, the ActiveListenRobot and PassiveListenRobot were found significantly more likeable than the VoiceAgent (ActiveListenRobot: $s = 0.176$, $t = 2.887$, $p < 0.05$; PassiveListenRobot: $se = 0.178$, $t = 2.685$, $p < 0.05$). A post hoc Bonferroni-Holm correction confirmed these results. Thus, H1-5 was supported, but H2-5 was not supported as there was no significant difference found on likeability scores between the ActiveListenRobot and the PassiveListenRobot.

Furthermore, in response to RQ3, participants with the agreeableness personality trait had positive correlations with likeability scores ($se = 0.064$, $t = 2.142$, $p < 0.05$).

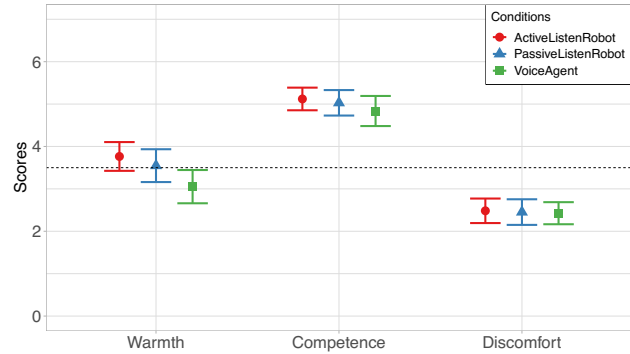


Figure 6: The scores for the three social dimensions (warmth, competence, discomfort) defined in RoSAS across the three experimental conditions. The dashed line indicates the neutral choice at the center of the seven-point Likert scale. The error bars indicate 95% confidence intervals.

6 Discussion

In this online study, participants were requested to evaluate three agents functioning as public speaking coaches. To investigate how participants perceived different agents as public speaking coaches, we analyzed two types of agents that varied in their physical embodiment (a voice assistant agent and a social robot). In addition, for one of the agents (the social robot), we altered its non-verbal behaviour by including head nods as a means of active listening.

6.1 Perceptions of agents as public speaking coaches

We found that the active listener robot agent, i.e., a robot showing backchannelling behaviour, was perceived as more natural than the passive listener robot (i.e., a robot without backchannelling behaviour) and the voice assistant agent. The active listener robot was perceived as more satisfying than the voice assistant agent, while the results before Bonferroni-Holm correction indicated that the active listener robot was perceived as more satisfying than the passive listener robot as well. There was no difference between the passive listener robot and the voice assistant agent regarding perceived satisfaction and naturalness. This finding supports the effectiveness of active listening behaviour in the robot and is consistent with earlier research, namely, “active engagement” that can lead to an increase in relational satisfaction [80,81]. According to Bodie [81] when one feels that they are being heard and understood, they tend to be more content with their

Table 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 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)

Covariate	Warmth				Competence				Discomfort			
	Estimate	SE	<i>t</i>	<i>p</i>	Estimate	SE	<i>t</i>	<i>p</i>	Estimate	SE	<i>t</i>	<i>p</i>
Condition												
ActiveListenRobot	0.692	0.248	2.796	<0.05*								
PassiveListenRobot	0.505	0.247	2.045	<0.05*								
Age												
Age									0.013	0.007	1.779	0.077
Gender												
Male	−0.638	0.216	−2.954	<0.01**					0.346	0.166	2.083	<0.05*
TIPI												
Extraversion												
Agreeableness	0.328	0.097	3.387	<0.001***	0.2	0.077	2.583	<0.05*	0.079	0.049	1.606	0.110
Conscientiousness	−0.315	0.089	−3.541	<0.001***					−0.121	0.075	−1.625	0.106
Emotional stability	0.155	0.08	1.943	0.054					−0.201	0.06	−3.315	<0.01**

* = $p < 0.05$; ** = $p < 0.01$; *** = $p < 0.001$.

relationships, and this can have a significant impact on their physical health and well-being.

Several studies have proposed probabilistic and reinforcement learning models to predict the appropriate timing for generating backchannels during dyadic conversations between humans and agents [82–84]. In our research, we utilized a Wizard-of-Oz approach to initiate backchannels without considering appropriate timing. Yet, we found that simply the act of head nodding while listening, even given that we did not pay attention to when the proper time is to make the robot nod, can impact the perception of the robot's naturalness as a coach. A related study [85], investigated the use of random verbal backchannels in a voice assistant agent and found that this approach led to prolonged interactions and greater user engagement.

Regarding agent engagement, participants perceived the active listener robot as having a more engaging interaction with the student compared to the voice agent. However, we found no significant difference in terms of engagement between the active listener and the passive listener robots. This result may be due to the novelty effect of the social robot, but it cannot be solely attributed to novelty because there was also no significant difference in the engagement factor between the passive listener robot and the voice agent. This aligns with our **H1**, however, it solely shows the advantages of the active listener robot in comparison to the voice agent. Our results also support previous research on active listening and engagement in human–robot interactions [34]. June et al (2013) found that incorporating backchannelling in robots resulted in an improved perception of engagement in the robots [57].

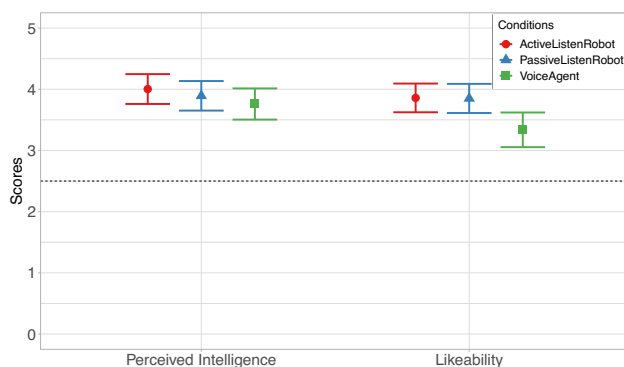


Figure 7: The average of the Godspeed scores on the dimensions of perceived intelligence and likeability across all experimental conditions. The dashed line indicates the neutral choice at the center of the five-point Likert scale. The error bars indicate 95% confidence intervals.

Table 7: Generalized linear models predicting the Godspeed likeability. Factors included in the model for each dependent variable were selected based on the AIC. The estimates were according to comparisons to the baseline (VoiceAgent)

Covariate	Likeability			
	Estimate	SE	<i>t</i>	<i>p</i>
Condition				
ActiveListenRobot	0.508	0.176	2.887	<0.05*
PassiveListenRobot	0.477	0.178	2.685	<0.05*
Age				
Age	−0.011	0.006	−1.715	.088
TIPI				
Agreeableness	0.138	0.064	2.142	<0.05*

* = $p < 0.05$; ** = $p < 0.01$; *** = $p < 0.001$.

It is noteworthy that the active listener robot was perceived as significantly more satisfying, engaging, natural, and humane than the voice agent, while we did not find the passive listener robot to differ significantly from the voice agent on these measures. This is interesting because both the active listener and passive listener robots used nonverbal social cues in the form of hand gestures and arm movements while giving feedback, but only the agent that displayed active listening behaviour was perceived significantly better than the voice agent on the aforementioned metrics. The study did not find any significant differences between the agents in terms of other metrics that were more relevant to their task performance as a coach, including helpfulness, attentiveness, competence, and trustworthiness. This is also in accordance with participants' agreement score for having the agents as a coach which was relatively high in all conditions. (For more information please refer to [86].) One possible explanation might be that the quality and accuracy of the agent's feedback might have had a more salient effect on peoples' evaluations of it as a coach than its social behaviour. Furthermore, the study did not find any significant differences between the feeling of rapport toward the agents and the participants' self-reported intention to use the agents in the future. This may be due to participants' higher familiarity with smart speakers, their ease of use, and their prevalence of them that causes them to be popular [87]. In addition, it could be that individuals who were presented with the voice assistant agent were unable to envision engaging with a social robot or integrating social nonverbal signals into their feedback. As a result, they deemed the voice-only feedback received from the voice assistant agent to be sufficient.

6.2 Perceptions of the human nature and social attributes of agents

Having a social robot as the coach in both *ActiveListenRobot* and *PassiveListenRobot* conditions led to significantly higher ratings of HN attributes [69] than the *VoiceAgent* condition. Therefore, it seems that the type of the agent, specifically the presence of anthropomorphic features and non-verbal social behaviours during feedback, might have positively affected the perception of HN attributes. Although we did not find strong evidence supporting that active listening behaviour affected the perception of the HN attributes, our implementations and the amount of non-verbal behaviour used in both robotic conditions could have affected our results. In the study by Kontogiorgos et al. [26], the anthropomorphic robotic head was perceived as more co-present in the room compared to the non-anthropomorphic voice assistant agent, even in the

absence of non-verbal behaviours such as gaze and head movements. Similarly, Looije et al. [25] compared anthropomorphic and non-anthropomorphic agents and found that social characters were perceived to be more empathetic compared to a text interface, and elicited more conversational behaviour. Also, in a game entertainment scenario, both animated and non-animated Pepper robots were perceived to have higher entertainment value and hedonic quality compared to a voice assistant agent [23].

The results of the RoSAS questionnaire suggested that across three dimensions – warmth, competence, and discomfort – the active listener robot was perceived as more socially “warm” than the voice agent. Note, before applying the Bonferroni-Holm correction, the passive listener robot was perceived as having more warmth than the voice assistant agent; however, this was not confirmed by the Bonferroni-Holm post hoc analysis. The higher perceptions of warmth and friendliness in the active listener robot (and maybe the passive listener robot) compared to the voice assistant agent can be attributed to its social capacities. According to Kontogiorgos [26], humans tend to have a cognitive preference for social activities when exposed to subtle social cues and anthropomorphic features. Familiar channels of communication are particularly relevant for tasks that require face-to-face collaboration [26]. In accordance with our findings, [25] found that participants rated a robotic head with a human-like face as more sociable than a voice assistant agent, indicating that the presence of anthropomorphic features in agents may evoke a sense of familiarity akin to natural communication.

6.3 Perceived intelligence and likeability of agents

The findings indicate that there were no significant differences among the agents in terms of perceived intelligence, potentially due to the more prominent impact of the feedback phase, and its content, on participants' perception of the agent's intelligence. One possible explanation for why we did not observe significant differences in perceived intelligence and other factors, such as attentiveness and competence among the agents, could be that the effect of verbal feedback may have overshadowed the impact of listening behaviour and that participants mainly evaluated the agents as coaches based on the provided feedback. It is possible that participants in our study may have started to forget about the phase when the robot was actively listening to the student and how it behaved during that

time. This could be attributed to the fact that participants were presented with surveys and questionnaires immediately after the video was completed, which brought them closer to the moment when they observed the coach's feedback. However, the results demonstrated that the active listener and passive listener robots were more likable than the voice agent, which is similar to a study carried out by Moore and Urakami on different voice user interfaces and their distracting effects, which found that the scores of likeability were higher for the social embodied VUI (the social robot) than the physical embodied VUI (the smart speaker) [24].

6.4 Potential gender differences in perception of agents

The outcomes of our analysis indicated a significant effect of gender on the perceptions of the agent as a coach. Participants who self-identified as female rated the agents significantly more positively than those who self-identified as male in terms of attentiveness and social warmth, as well as on the HN attributes of the agents (emotionality, warmth, openness). This is in accordance with previous research on the gender difference in non-verbal communication, which suggested that females tend to utilize non-verbal cues more than males and that females can judge emotions and personalities based on nonverbal cues more accurately than males [88]. Nevertheless, as per Schermerhorn *et al.* [89], males typically perceive robots as more human-like than females, which contradicts our findings.

Males provided significantly lower ratings than females for intention to use and feeling of rapport with the agents. They also rated the feedback provided by the agents as less trustworthy compared to females. These results, particularly for the two conditions with a social robot, contradict previous studies that have demonstrated that males typically have a stronger preference for robots than females [89,90]. Note that we only had participants who identified themselves as male or female (with two participants who preferred not to share their gender), therefore, we could not compare these results for gender identities other than male and female.

It should be noted that despite the differences in ratings between genders, only males were actually more inclined to use the active listener robot than the passive listener robot and reported a greater sense of rapport with it.

6.5 Participants' personality traits and differences in perceptions

A meta-analysis examining the impact of the Big Five personality traits on robot acceptance [91] showed that agreeableness and conscientiousness were the least frequently assessed variables. Agreeableness pertains to the extent to which an individual is warm and amiable [92]. Conscientiousness, on the other hand, refers to the degree to which a person is diligent, careful, and aware of their behaviour and its consequences [93].

Our findings indicate that individuals with high levels of agreeableness were significantly more inclined to utilize the agents as coaches in the future and experienced a greater sense of rapport with them. Furthermore, participants with high levels of agreeableness perceived the agent as a natural coach and attributed HN attributes to it. Agreeableness has been reported to have positive correlations with robot acceptance in several studies [94,95]. According to Bernotat and Eyssel [96], the reason stems from the fact that people who are more agreeable tend to trust more. Trust can reduce perceived risk and increase the acceptance of new technologies [97].

On the contrary, conscientiousness negatively correlated with evaluations of the agents as coaches. Participants who scored higher on conscientiousness perceived the agent as less warm and less emotional. It was less likely that they would use such a system in the future, and they felt less rapport with the agent as a coach. Similarly, Looije *et al.* found that the more conscientious a participant is, the less they liked a social robot [25]. Kimoto *et al.* found negative correlations between the conscientiousness personality trait and robot acceptance [94]. Also, Syrdal *et al.* found that people with high conscientiousness were less likely to allow a robot to approach them closer [98].

Our findings support hypothesis **H3**, which predicted that there is a relationship between the evaluations of the agents and the Big Five personality traits.

6.6 Participants' comments about the agents

We also posed open-ended questions to participants regarding their views on the extent to which the agents could function as public speaking coaches, as well as whether they preferred the same agent they had observed, a different type of agent, or a human as a coach (see [99] for more details). According to a thematic analysis (TA) [86] of the participants'

responses, we found a high level of agreement on the use of social robots and voice assistant agents as public speaking coaches. The TA results suggested informative pros and cons for each type of agent, which could be used to enhance the design and implementation of social robots and other agents as public speaking coaches. The thematic analysis and details of the results go beyond the scope of this article and are reported elsewhere [99].

6.7 Summary of findings

The main takeaways from this study are as follows:

- The active listener robot was perceived as more satisfying, natural, and engaging than the voice assistant agent. This difference was not observed between the passive listener robot and the voice assistant agent.
- The active listener robot was perceived as more natural than the passive listener robot, but no more significant differences were observed between the active listener robot and the passive listener robot.
- Both the active listener robot and the passive listener robot were perceived as more human-like than the voice assistant agent.
- The active listener robot was perceived as warmer during the interaction than the voice assistant agent, but no significant difference was seen between the passive listener robot and the voice assistant agent in this regard.
- Both the active listener robot and the passive listener robot were more likeable than the voice assistant agent, according to participants.

7 Limitations and future work

The experiment was conducted virtually, which came with its set of advantages and disadvantages. By using Amazon Mechanical Turk, we were able to recruit a large pool of participants. The prevalence of online studies has increased since the COVID-19 pandemic due to its safety in collecting data in the field of HRI [100]. However, using crowdsourcing platforms such as MTurk comes with limitations regarding the quality control of responses. This is due to the population of MTurk workers being composed of individuals with diverse and unknown abilities, technical resources, personal goals, the remote nature of participants' involvement in the study, etc. [101,102].

In our study, we took great care to adjust our virtual interface in a way that it would minimize the potential

influence of confounding factors, such as agent's voice and feedback. We applied attention checks to identify and exclude participants who did not seem to engage with the task at hand.

Our study had some limitations that were noted by participants. One such limitation was the quality of the voice used by the agents, which some participants found to be robotic and difficult to understand. Please note that the Pepper robot's voice was used in all conditions, and we tried to test a range of pitches and frequencies within its capabilities to make it as "adult" and "comprehensible" as possible. In future studies, we could consider e.g. using a recorded human voice or the voice of the Google Nest as alternative approaches. In fact, using Pepper's voice on Google Nest could have affected the perception of the agent by those who were familiar with Google Nest's voice. However, due to the trade-off and potential impacts of having different voices (which can affect speech rate, pitch, etc.), we decided to use the identical voice and recording for all agents.

As previously mentioned, the feedback provided by the robot may have taken precedence over the robot's behaviour during the listening phase. This could have resulted in participants forgetting how the robot behaved while the student was presenting, affecting comparisons between the robot showing backchannelling and the one that remained still during the presentation. In future studies, we could address this limitation by separating the evaluation of the robot's behaviour during the listening phase from the feedback phase. This would allow us to emphasize people's judgment 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 in some cases been shown to produce results that are comparable to those obtained through in-person studies [103,104]. However, our results may differ from those obtained in a live human-robot interaction (HRI) scenario. Participants who interact with a social agent in person are able to pay attention to many more aspects of the agent and recognize them better [105], particularly when it comes to detecting backchannels and head nods, which are more evident when participants are physically co-present with the robot (e.g., can also be affected by the noise from the robot's actuators). In-person interaction allows participants to experience the situation more vividly holistically, and they could be more likely to pay closer attention to the agent's behaviour and assess its capabilities. Therefore, for future studies, we intend to investigate the effectiveness of a similar scenario in an in-person interaction with the robot.

For the future implementation of an intelligent agent as a coach, one suggestion noted by some participants was

the format of the feedback provided by the robot. Some participants indicated that they would have preferred to receive written feedback in the form of notes as well, which would have made it easier for them to track and remember the information presented. In future studies, one could consider providing both verbal and written feedback through an external monitor or screen or through Pepper's tablet to accommodate participants' preferences.

A previous study [24] showed that a social embodied agent was more distracting in a cognitive task than a physical embodied agent, and a voice-only system. However, we did not focus our study on the distraction effect of the agent as a coach. Nevertheless, whether a coaching agent may distract participants in a public speaking task is a critical issue that needs to be studied in more detail.

In addition, future research may use a different robot and incorporate a variety of other non-verbal backchanneling cues, including smiling, raising eyebrows, leaning forward to indicate attention, as well as possibly incorporating verbal backchannels.

8 Conclusion

The main focus of our study was to explore people's perceptions of an agent's physical embodiment and behavioural capabilities as a public speaking coach, based on feedback from a large range of participants of different ages. Participants evaluated the agent's performance taking into account two phases, a listening and a feedback phase. Rather than testing the effectiveness of one specific agent, we compared people's perceptions of different agents as public speaking rehearsal coaches.

Our findings suggest that an agent's embodiment had a positive impact on the perception of human nature attributes and likeability. However, we did not observe any notable differences between the voice agent and the social robot in terms of perceived intelligence. Perhaps the feedback phase and the content of the verbal feedback had a more significant effect on perceptions of the agent's intelligence than the agent's non-verbal behaviour.

Our study results also showed that social non-verbal behaviour, particularly during listening, plays a significant role in creating a natural interaction. However, it does not necessarily affect how participants evaluate an agent's coaching ability throughout the rehearsal session. Perhaps the effect of active listening was not as salient as we originally expected, or it was "overshadowed" by the robot's extended verbal feedback.

In general, participants in our sample were receptive to using interactive agents as public speaking coaches. Specifically, individuals with higher levels of agreeableness expressed more positive attitudes towards the agents, while those with higher levels of conscientiousness expressed more negative attitudes towards using socially interactive agents as public speaking coaches.

The study provided information on general attitudes toward using social robots and voice assistants as public speaking coaches, which can inform future design and utilization of such agents as coaches.

Acknowledgement: The authors are grateful for the reviewers' valuable comments that improved the manuscript. The authors would like to express their appreciation to Pourya Aliasghari and Sahand Shaghaghi for assisting with setting up the online framework and analyzing the data throughout this study.

Funding information: This research was undertaken, in part, thanks to funding from the Canada 150 Research Chairs Program.

Author contributions: All authors have accepted responsibility for the entire content of this manuscript and consented to its submission to the journal, reviewed all the results, and approved the final version of the manuscript. All co-authors jointly designed the study which was implemented and carried out by Delara Forghani (DF), under the guidance of all co-authors. DF analyzed the data with guidance from the co-authors. DF wrote the first complete draft of the manuscript which was reviewed by all co-authors.

Conflict of interest: The authors state no conflict of interest.

Ethical approval: The research study was approved by the University of Waterloo Human Research Ethics Board.

Informed consent: Informed consent has been obtained from all individuals included in this study.

Data availability statement: Research data are not publicly available.

References

- [1] A. K.-F. Lui, S.-C. Ng, and W.-W. Wong, "A novel mobile application for training oral presentation delivery skills," in: *Technology in education. Technology-mediated proactive learning* J. Lam, K. K. Ng,

- S. K. Cheung, T. L. Wong K. C. Li, and F. L. Wang, (Eds). Berlin, Heidelberg: Springer Berlin Heidelberg, 2015, pp. 79–89.
- [2] J. Wang, H. Yang, R. Shao, S. Abdullah, and S. S. 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*, ser. CHI '20. New York, NY, USA: Association for Computing Machinery, 2020, p. 1–13. doi: <https://doi.org/10.1145/3313831.3376561>.
- [3] J. Hernández-March, M. Martín del Peso, and S. Leguey, “Graduates’ skills and higher education: The employers’ perspective,” *Tertiary Education Manag.*, vol. 15, no. 1, pp. 1–16, Mar 2009. doi: <https://doi.org/10.1080/13583880802699978>.
- [4] S. Vicinanza, “Information technology as a discipline: What employers want in an IT graduate,” *ACM Inroads*, vol. 4, no. 4, p. 59–63, Dec 2013, doi: <https://doi.org/10.1145/2537753.2537774>.
- [5] J. Norback and J. Hardin, “Integrating workforce communication into senior design,” *IEEE Trans. Professional Commun.*, vol. 48, no. 4, pp. 413–426, 2005.
- [6] H. Trinh, R. Asadi, D. Edge, and T. Bickmore, “Robocop: A robotic coach for oral presentations,” *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* vol. 1, no. 2, Jun 2017. doi: <https://doi.org/10.1145/3090092>.
- [7] A. Goodman, and C. Communications, “Why bad presentations happen to good causes: And how to ensure they won’t happen to yours,” *Cause Commun.*, 2006. <https://books.google.ca/books?id=tcsPAQAAMAAJ>.
- [8] D. Kerby and J. Romine, “Develop oral presentation skills through accounting curriculum design and course-embedded assessment,” *J. Education Business*, vol. 85, no. 3, pp. 172–179, 2009. doi: <https://doi.org/10.1080/08832320903252389>.
- [9] A. W. Blöte, M. J. W. Kint, A. C. Miers, and P. M. Westenberg, “The relation between public speaking anxiety and social anxiety: a review,” *J. Anxiety Disord.*, vol. 23, no. 3, pp. 305–313, Apr. 2009.
- [10] J. Guggemos, S. Seufert, S. Sonderegger, and M. Burkhard, *Social robotics in education: Conceptual overview and case study of use*. Cham: Springer International Publishing, 2022, pp. 173–195. doi: https://doi.org/10.1007/978-3-030-90944-4_10.
- [11] T. Belpaeme, J. Kennedy, A. Ramachandran, B. Scassellati, and F. Tanaka, “Social robots for education: A review,” *Sci. Robot.*, vol. 3, no. 21, p. eaat5954, 2018. <https://www.science.org/doi/abs/10.1126/scirobotics.aat5954>.
- [12] O. Mubin, C. Stevens, S. Shahid, A. Mahmud, and J.-J. Dong, “A review of the applicability of robots in education,” *Technol. Educ. Learn.*, vol. 1, 2013.
- [13] H. Woo, G. K. LeTendre, T. Pham-Shouse, and Y. Xiong, “The use of social robots in classrooms: A review of field-based studies,” *Educ. Res. Rev.*, vol. 33, p. 100388, 2021. <https://www.sciencedirect.com/science/article/pii/S1747938X21000117>.
- [14] D. Li, P. P. Rau, and Y. Li, “A cross-cultural study: Effect of robot appearance and task,” *Int. J. Soc. Robot.*, vol. 2, pp. 175–186, 2010.
- [15] G. Terzopoulos and M. Satratzemi, “Voice assistants and smart speakers in everyday life and in education,” *Inform. Educ.*, vol. 19, pp. 473–490, 2020.
- [16] S. Sandeep, “Learning English language through Amazon Alexa for Indian students,” *J. Gujarat Res. Soc.*, vol. 21, no. 10, pp. 619–622, 2019.
- [17] M. I. Tanveer, E. Lin, and M. E. 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*, ser. IUI '15. New York, NY, USA: Association for Computing Machinery, 2015, p. 286–295. doi: <https://doi.org/10.1145/2678025.2701386>.
- [18] H. Trinh, K. Yatani, and D. Edge, “Pitchperfect: Integrated rehearsal environment for structured presentation preparation,” in: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '14. New York, NY, USA: Association for Computing Machinery, 2014, p. 1571–1580. doi: <https://doi.org/10.1145/2556288.2557286>.
- [19] X. Wang, H. Zeng, Y. Wang, A. Wu, Z. Sun, and X. Ma, “Voicecoach: Interactive evidence-based training for voice modulation skills in public speaking,” *CoRR*, vol. abs/200107876, 2020. <https://arxiv.org/abs/2001.07876>.
- [20] J. Schneider, D. Börner, P. van Rosmalen, and M. Specht, “Presentation trainer, your public speaking multimodal coach,” in: *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*, ser. ICMI '15. New York, NY, USA: Association for Computing Machinery, 2015, p. 539–546. doi: <https://doi.org/10.1145/2818346.2830603>.
- [21] P. Murali, H. Trinh, L. Ring, and T. 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*, ser. IVA '21. New York, NY, USA: Association for Computing Machinery, 2021, p. 156–163. doi: <https://doi.org/10.1145/3472306.3478364>.
- [22] O. Saukh and B. Maag, “Quantle: fair and honest presentation coach in your pocket,” in: *IPSN*, M. R. Eskicioglu, L. Mottola, and B. Priyantha, (Eds). ACM, 2019, pp. 253–264. <http://dblp.uni-trier.de/db/conf/ipsn/ipsn2019.html#SaukhM19>.
- [23] K. Pollmann, C. Ruff, K. Vetter, and G. 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*, ser. HRI '20. New York, NY, USA: Association for Computing Machinery, 2020, p. 395–397. doi: <https://doi.org/10.1145/3371382.3378251>.
- [24] B. A. Moore and J. Urakami, “The impact of the physical and social embodiment of voice user interfaces on user distraction,” *Int. J. Human-Comput. Stud.*, vol. 161, p. 102784, 2022. <https://www.sciencedirect.com/science/article/pii/S1071581922000131>.
- [25] R. Looije, M. A. Neerinx, and F. Cnossen, “Persuasive robotic assistant for health self-management of older adults: Design and evaluation of social behaviors,” *Int. J. Human-Comput. Stud.*, vol. 68, no. 6, pp. 386–397, 2010, human-Computer Interaction for Medicine and Health care (HCI4MED): Towards making Information usable. <https://www.sciencedirect.com/science/article/pii/S1071581909000113X>.
- [26] D. Kontogiorgos, A. Pereira, O. Andersson, M. Koivisto, E. Gonzalez Rabal, and V. Vartiainen, “The effects of anthropomorphism and non-verbal social behaviour in virtual assistants,” in: *Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents*, ser. IVA '19. New York, NY, USA: Association for Computing Machinery, 2019, p. 133–140. doi: <https://doi.org/10.1145/3308532.3329466>.
- [27] E. Ayedoun, Y. Hayashi, and K. Seta, “Communication strategies and affective backchannels for conversational agents to enhance learners’ willingness to communicate in a second language,” in: *Artificial intelligence in education*, E. André, R. Baker, X. Hu, M. M. T. Rodrigo, and B. du Boulay, Eds. Cham: Springer International Publishing, 2017, pp. 459–462.
- [28] C. Smith, N. Crook, J. Boye, D. Charlton, S. Dobnik, and D. Pizzi, “Interaction strategies for an affective conversational agent,” in:

- Intelligent virtual agents J. Allbeck, N. Badler, T. Bickmore, C. Pelachaud, and A. Safonova, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 301–314.
- [29] G. Bodie, A. Vickery, K. Cannava, and S. Jones, “The role of “active listening” in informal helping conversations: Impact on perceptions of listener helpfulness, sensitivity, and supportiveness and discloser emotional improvement,” *Western J. Commun.*, vol. 79, no. 2, pp. 151–173, Mar. 2015.
- [30] F. Eyssel, R. Wullenkord, and V. Nitsch, “The role of self-disclosure in human–robot interaction,” in: *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 2017, pp. 922–927.
- [31] L. Batrinca, G. Stratou, A. Shapiro, L.-P. Morency, and S. Scherer, “Cicero – towards a multimodal virtual audience platform for public speaking training,” in: *Intelligent virtual agents*, R. Aylett, B. Krenn, C. Pelachaud, and H. Shimodaira, (Eds). Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 116–128.
- [32] M. Luria, G. Hoffman, and O. 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*, ser. CHI '17. New York, NY, USA: Association for Computing Machinery, 2017, p. 580–628. doi: <https://doi.org/10.1145/3025453.3025786>.
- [33] J. Gratch, N. Wang, A. Okhmatovskaia, F. Lamothe, M. Morales, and R. J. van der Werf, “Can virtual humans be more engaging than real ones?,” in: *Human-computer interaction. HCI intelligent multimodal interaction environments*, J. A. Jacko, (Ed). Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 286–297.
- [34] C. L. Sidner, C. Lee, C. D. Kidd, N. Lesh, and C. Rich, “Explorations in engagement for humans and robots,” *Artif. Intell.*, vol. 166, no. 1, pp. 140–164, 2005. <https://www.sciencedirect.com/science/article/pii/S0004370205000512>.
- [35] B. Craenen, A. Deshmukh, M. E. Foster, and A. 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)*, 2018, pp. 626–631.
- [36] A. Edwards, C. Edwards, P. R. Spence, C. Harris, and A. Gambino, “Robots in the classroom: Differences in students’ perceptions of credibility and learning between “teacher as robot” and “robot as teacher,” *Comput. Human Behav.*, vol. 65, pp. 627–634, 2016. <https://www.sciencedirect.com/science/article/pii/S0747563216304332>.
- [37] C. Edwards, A. Edwards, B. Stoll, X. Lin, and N. Massey, “Evaluations of an artificial intelligence instructor’s voice: Social identity theory in human–robot interactions,” *Comput. Human Behav.*, vol. 90, pp. 357–362, 2019. <https://www.sciencedirect.com/science/article/pii/S0747563218304023>.
- [38] C. I. Nass and S. Brave, *Wired for speech: How voice activates and advances the human-computer relationship*. Cambridge: MIT Press, 2005.
- [39] C. E. Patric R. Spence, D. Westerman, and A. Edwards, “Welcoming our robot overlords: Initial expectations about interaction with a robot,” *Commun. Res. Reports*, vol. 31, no. 3, pp. 272–280, 2014. doi: <https://doi.org/10.1080/08824096.2014.924337>.
- [40] C. NASS, and J. Steuer, “Voices, boxes, and sources of messages,” *J. STEUER, Voices*, vol. 19, no. 4, pp. 504–527, 1993. <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1468-2958.1993.tb00311.x>.
- [41] H. Mahdi, S. A. Akgun, S. Saleh, and K. Dautenhahn, *A survey on the design and evolution of social robots – past, present and future Robot. Autonomous Syst.*, vol. 156, p. 104193, 2022. <https://www.sciencedirect.com/science/article/pii/S0921889022001117>.
- [42] G. Laban, J.-N. George, V. Morrison, and E. S. Cross, “Tell me more! assessing interactions with social robots from speech,” *Paladyn J. Behav. Robot.*, vol. 12, no. 1, pp. 136–159, 2021. doi: <https://doi.org/10.1515/pjbr-2021-0011>.
- [43] J. K. Barfield, “Self-disclosure of personal information, robot appearance, and robot trustworthiness,” in: *2021 30th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN)*, 2021, pp. 67–72.
- [44] S. Al Moubayed, J. Beskow, G. Skantze, and B. Granström, “Furhat: A back-projected human-like robot head for multiparty human-machine interaction,” in: *Cognitive behavioural systems*, A. Esposito, A. M. Esposito, A. Vinciarelli, R. Hoffmann, and V. C. Müller, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 114–130.
- [45] S. Lee, M. Cho, and S. Lee, “What if conversational agents became invisible? comparing users’ mental models according to physical entity of ai speaker,” *Proc. ACM Interact. Mob. Wearable Ubiquitous Technology*, vol. 4, no. 3, Sep 2020. doi: <https://doi.org/10.1145/3411840>.
- [46] V. H. Yngve, “On getting a word in edgewise,” in *CLS-70*, Chicago: University of Chicago, 1970, pp. 567–577.
- [47] A. Flint, *The effects of interlocutor backchannels and I1 backchannel norms on the speech of I2 english learners*, Ph.D. dissertation, University of Oxford, Oxford, United Kingdom, 2016.
- [48] R. Gardner, “Between speaking and listening: The vocalisation of understandings,” *Applied Linguistics*, vol. 19, pp. 204–224, 1998.
- [49] A.-B. Stenstrom, *An introduction to spoken interaction*. London: Longman, 1994.
- [50] R. Gardner, *When listeners talk: Response tokens and listener stance*. Amsterdam: John Benjamins, 2001. <https://www.jbe-platform.com/content/books/9789027297426>.
- [51] J. Algeo, *British or American English?: A Handbook of Word and Grammar Patterns*, ser. *Studies in English Language*. Cambridge: Cambridge University Press, 2006.
- [52] A. Dennis and S. Kinney, “Testing media richness theory in the new media: The effects of cues, feedback, and task equivocality,” *Inform. Syst. Res.*, vol. 9, pp. 256–274, 1998.
- [53] A. T. Dittmann and L. G. Llewellyn, “Relationship between vocalizations and head nods as listener responses,” *J. Personality Soc. Psychol.*, vol. 9, no. 1, pp. 79–84, 1968. doi: <https://doi.org/10.1037/h0025722>.
- [54] L. J. Brunner, *Smiles can be back channels.: Semantic scholar*, Jan 1979. <https://www.semanticscholar.org/paper/Smiles-can-be-back-channels.-Brunner/dd71404cfffbeb373cfccf209f47143a37f572e>.
- [55] C. Goodwin, *Conversational organization: Interaction between speakers and hearers*, London: Academic Press, 1981.
- [56] N. Cathcart, J. Carletta, and E. Klein, “A shallow model of back-channel continuers in spoken dialogue,” in: *Proceedings of the Tenth Conference on European Chapter of the Association for – Volume 1*, ser. *EACL '03*. USA: Association for Computational Linguistics, 2003, p. 51–58. doi: <https://doi.org/10.3115/1067807.1067816>.
- [57] M. F. Jung, J. J. Lee, N. DePalma, S. O. Adalgeirsson, P. J. Hinds, and C. Breazeal, “Engaging robots: Easing complex human–robot teamwork using back channeling,” in: *Proceedings of the 2013*

- Conference on Computer Supported Cooperative Work, ser. CSCW '13*. New York, NY, USA: Association for Computing Machinery, 2013, p. 1555–1566. doi: <https://doi.org/10.1145/2441776.2441954>.
- [58] K. Bousmalis, M. Mehu, and M. Pantic, “Towards the automatic detection of spontaneous agreement and disagreement based on nonverbal behaviour: A survey of related cues, databases, and tools,” *Image Vision Comput.*, vol. 31, no. 2, pp. 203–221, 2013, affect Analysis In Continuous Input. <https://www.sciencedirect.com/science/article/pii/S0262885612001059>.
- [59] C. Oertel, J. Lopes, Y. Yu, K. A. F. Mora, J. Gustafson, and A. W. Black, “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, ser. ICMi '16*. New York, NY, USA: Association for Computing Machinery, 2016, p. 21–28. doi: <https://doi.org/10.1145/2993148.2993188>.
- [60] M. Argyle, *Bodily communication*, Routledge, England, UK, 2013.
- [61] H. M. Rosenfeld, and M. Hancks, “The nonverbal context of verbal listener responses,” *Relationship Verbal Nonverbal Commun.*, vol. 25, pp. 193–206, 1980.
- [62] A. Andriella, R. Huertas-García, S. Forgas-Coll, C. Torras, and G. 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)*, 2020, pp. 752–759.
- [63] Z. Ding, J. Kang, T. O. T. Ho, K. H. Wong, H. H. Fung, and H. Meng, *Talktive: A conversational agent using backchannels to engage older adults in neurocognitive disorders screening*, 2022. <https://arxiv.org/abs/2202.08216>.
- [64] H. W. Park, M. Gelsomini, J. J. Lee, and C. Breazeal, “Telling stories to robots: The effect of backchannelling on a child’s storytelling,” in: *Proceedings of the 2017 ACM/IEEE International Conference on human–robot Interaction, ser. HRI '17*. New York, NY, USA: Association for Computing Machinery, 2017, p. 100–108. doi: <https://doi.org/10.1145/2909824.3020245>.
- [65] S. D. Gosling, P. J. Rentfrow, and W. B. Swann, “A very brief measure of the big-five personality domains,” *J. Res. Personality*, vol. 37, no. 6, pp. 504–528, 2003. <https://www.sciencedirect.com/science/article/pii/S0092656603000461>.
- [66] F. Funke and U.-D. Reips, *Why semantic differentials in web-based research should be made from visual analogue scales and not from 5-point scales*, Field Methods, UK, 2012.
- [67] J. Matejka, M. Glueck, T. Grossman, and G. Fitzmaurice, “The effect of visual appearance on the performance of continuous sliders and visual analogue scales,” in: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, ser. CHI '16*, New York, NY, USA: Association for Computing Machinery, 2016, p. 5421–5432. doi: <https://doi.org/10.1145/2858036.2858063>.
- [68] H. Treiblmaier and P. Filzmoser, *Benefits from using continuous rating scales in online survey research*, International Conference on Interaction Sciences, Busan, Korea, 2011.
- [69] N. Haslam, S. Loughnan, Y. Kashima, and P. Bain, “Attributing and denying humanness to others,” *European Rev. Soc. Psychol.*, vol. 19, no. 1, pp. 55–85, 2008. doi: <https://doi.org/10.1080/10463280801981645>.
- [70] M. Ghafurian, N. Budnarain, and J. Hoey, “Improving humanness of virtual agents and users’ cooperation through emotions,” *IEEE Trans. Affective Comput.*, pp. 1461–1471, 2021.
- [71] C. Bartneck, D. Kulić, E. Croft, and S. Zoghbi, “Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots,” *Int. J. Soc. Robot.*, vol. 1, no. 1, pp. 71–81, Jan 2009. doi: <https://doi.org/10.1007/s12369-008-0001-3>.
- [72] C. M. Carpinella, A. B. Wyman, M. A. Perez, and S. J. Stroessner, “The robotic social attributes scale (ROSAS): Development and validation,” in: *2017 12th ACM/IEEE International Conference on human–robot Interaction (HRI)*, 2017, pp. 254–262.
- [73] M. Spitale, M. Axelsson, and H. 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, ser. HRI '23*. New York, NY, USA: Association for Computing Machinery, 2023, p. 301–310, doi: <https://doi.org/10.1145/3568162.3577003>.
- [74] M. T. Parreira, S. Gillet, K. Winkle, and I. 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, ser. HRI '23*. New York, NY, USA: Association for Computing Machinery, 2023, p. 11–20. doi: <https://doi.org/10.1145/3568294.3580032>.
- [75] H. Fukuda and K. Ueda, “Interaction with a moving object affects one’s perception of its animacy,” *Int. J. Soc. Robot.*, vol. 2, no. 2, pp. 187–193, Jun 2010. doi: <https://doi.org/10.1007/s12369-010-0045-z>.
- [76] J. A. Nelder and R. W. M. Wedderburn, “Generalized linear models,” *J. R. Stat. Soc. Ser. A (General)*, vol. 135, no. 3, pp. 370–384, 2023/01/15/1972, full publication date: 1972. doi: <https://doi.org/10.2307/2344614>.
- [77] H. Bozdoğan, “Model selection and Akaike’s information criterion (AIC): The general theory and its analytical extensions,” *Psychometrika*, vol. 52, no. 3, pp. 345–370, Sep 1987. doi: <https://doi.org/10.1007/BF02294361>.
- [78] S. Holm, “A simple sequentially rejective multiple test procedure,” *Scandinavian J. Stat.*, vol. 6, no. 2, pp. 65–70, 1979. <http://www.jstor.org/stable/4615733>.
- [79] T. L. Chen, C.-H. King, A. L. Thomaz, and C. C. Kemp, “Touched by a robot: an investigation of subjective responses to robot-initiated touch,” in: *Proceedings of the 6th International Conference on human–robot Interaction, ser. HRI '11*. New York, NY, USA: Association for Computing Machinery, 2011, p. 457–464. doi: <https://doi.org/10.1145/1957656.1957818>.
- [80] T. Fassaert, S. van Dulmen, F. Schellevis, and J. Bensing, “Active listening in medical consultations: development of the active listening observation scale (ALOS-global),” *Patient Educ Couns*, vol. 68, no. 3, pp. 258–264, Aug. 2007.
- [81] G. D. Bodie, “Listening as positive communication,” *Positive Side Interpersonal Commun.*, pp. 109–125, 2012.
- [82] N. Hussain, E. Erzín, T. M. Sezgin, and Y. Yemez, “Training socially engaging robots: Modeling backchannel behaviors with batch reinforcement learning,” *IEEE Trans. Affective Comput.*, vol. 13, no. 4, pp. 1840–1853, 2022.
- [83] N. Hussain, E. Erzín, T. Sezgin, and Y. Yemez, “Speech driven backchannel generation using deep q-network for enhancing engagement in human–robot interaction,” *CoRR*, vol. abs/1908.01618, 2019. <http://arxiv.org/abs/1908.01618>.
- [84] L.-P. Morency, I. de Kok, and J. Gratch, “Predicting listener backchannels: A probabilistic multimodal approach,” in: *Intelligent virtual agents*, H. Prendinger, J. Lester, and M. Ishizuka, (Eds), Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 176–190.
- [85] N. Motalebi, E. Cho, S. S. Sundar, and S. 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, ser. CSCW '19. New York, NY, USA: Association for Computing Machinery, 2019, p. 309–313. doi: <https://doi.org/10.1145/3311957.3359502>.
- [86] V. Braun and V. Clarke, “Using thematic analysis in psychology,” *Qualitative Res. Psychol.*, vol. 3, pp. 77–101, Jan 2006.
- [87] E. Research, *The smart audio report from npr and edison research, spring 2018*, Jul 2018. <https://www.edisonresearch.com/the-smart-audio-report-from-npr-and-edison-research-spring-2018/>.
- [88] J. A. Hall, S. Seufert, S. Sonderegger, and S. D. Gunnery, *21 Gender differences in nonverbal communication*. Berlin, Boston: De Gruyter Mouton, 2023, pp. 639–670. doi: https://doi.org/10.1007/978-3-030-90944-4_10.
- [89] P. Schermerhorn, M. Scheutz, and C. 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*, ser. HRI '08. New York, NY, USA: Association for Computing Machinery, 2008, p. 263–270. doi: <https://doi.org/10.1145/1349822.1349857>.
- [90] N. Reich-Stiebert and F. 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*, ser. HRI '17. New York, NY, USA: Association for Computing Machinery, 2017, p. 166–176. doi: <https://doi.org/10.1145/2909824.3020242>.
- [91] C. Esterwood, K. Essenmacher, H. Yang, F. Zeng, and L. P. 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*, ser. CHI '21. New York, NY, USA: Association for Computing Machinery, 2021. doi: <https://doi.org/10.1145/3411764.3445542>.
- [92] M. Peeters, H. van Tuijl, C. Rutte, and I. Reymen, “Personality and team performance: a meta analysis,” *Europ. J. Personality*, vol. 20, no. 5, pp. 377–396, 2006.
- [93] K. Tasa, G. J. Sears, and A. C. H. Schat, “Personality and teamwork behavior in context: The cross-level moderating role of collective efficacy,” *J. Organiz. Behav.*, vol. 32, no. 1, pp. 65–85, 2011. doi: <https://onlinelibrary.wiley.com/doi/abs/10.1002/job.680>.
- [94] M. Kimoto, T. Iio, M. Shiomi, I. Tanev, K. Shimohara, and N. Hagita, “Relationship between personality and robots- interaction strategies in object reference conversations,” in: *Proceedings of the Second International Conference on Electronics and Software Science (ICES2016)*, Japan, 2016, pp. 128–136.
- [95] H. Salam, O. Çeliktutan, I. Hupont, H. Gunes, and M. Chetouani, “Fully automatic analysis of engagement and its relationship to personality in human–robot interactions,” *IEEE Access*, vol. 5, pp. 705–721, 2017.
- [96] J. Bernotat and F. 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)*, 2017, pp. 641–646.
- [97] P. A. Pavlou, “Consumer acceptance of electronic commerce: Integrating trust and risk with the technology acceptance model,” *Int. J. Electronic Commerce*, vol. 7, no. 3, pp. 101–134, 2003. doi: <https://doi.org/10.1080/10864415.2003.11044275>.
- [98] D. S. Syrdal, K. Lee Koay, M. L. Walters, and K. 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*, 2007, pp. 1143–1148.
- [99] D. Forghani, M. Ghafurian, S. Rasouli, C. L. Nehaniv, and K. Dautenhahn, “What do people think of social robots and voice agents as public speaking coaches?,” in: *2023 32nd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, 2023, pp. 996–1003.
- [100] D. Feil-Seifer, K. S. Haring, S. Rossi, A. R. Wagner, and T. Williams, “Where to Next? The Impact of COVID-19 on Human-Robot Interaction Research” *J. Hum.-Robot Interact.*, vol. 10, no. 1, Jun 2020. doi: <https://doi.org/10.1145/3405450>.
- [101] F. Daniel, P. Kucherbaev, C. Cappiello, B. Benatallah, and M. Allahbakhsh, “Quality control in crowdsourcing: A survey of quality attributes, assessment techniques, and assurance actions,” *ACM Comput. Surv.*, vol. 51, no. 1, Jan 2018. doi: <https://doi.org/10.1145/3148148>.
- [102] M. Joosse, M. Lohse, and V. Evers, “Crowdsourcing culture in HRI: lessons learned from quantitative and qualitative data collections,” in: *3rd International Workshop on Culture Aware Robotics at ICSR*, vol. 15, 2015.
- [103] C. Bartneck, A. Duenser, E. Moltchanova, and K. Zawieska, “Comparing the similarity of responses received from studies in Amazon’s mechanical turk to studies conducted online and with direct recruitment,” *PLOS ONE*, vol. 10, no. 4, pp. 1–23, Apr 2015. doi: <https://doi.org/10.1371/journal.pone.0121595>.
- [104] P. Jonell, T. Kucherenko, I. Torre, and J. Beskow, “Can we trust online 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*, ser. IVA '20. New York, NY, USA: Association for Computing Machinery, 2020. doi: <https://doi.org/10.1145/3383652.3423860>.
- [105] S. Wallkötter, R. Stower, A. Kappas, and G. 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*, ser. HRI '20, 2020, p. 609–618, doi: <https://doi.org/10.1145/3319502.3374800>.