International Journal of Social Robotics
https://doi.org/10.1007/s12369-025-01320-8

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

Delara Forghani[1,2] · Samira Rasouli[1] · Moojan Ghafurian[3] · Melanie Jouaiti[1,4] · Chrystopher L. Nehaniv[1,3] · Kerstin Dautenhahn[1]

Received: 13 May 2024 / Revised: 16 July 2025 / Accepted: 10 September 2025

Abstract

Presentation rehearsal is an essential activity that significantly impacts the quality of presentation delivery, especially for novice presenters, such as university students. However, rehearsals are often not done appropriately, or lack constructive feedback. To encourage effective presentation rehearsal, we devised a system involving a social humanoid robot acting as a public speaking coach, analyzing students' presentations and providing feedback. We monitored acoustic aspects of speech, speech prosodies, and eye contact maintenance during presentations. Our aim was to assess robot acceptance, participants' sense of interpersonal closeness with the robot, and perceived human nature attributes of the robot. This study presents the system development, followed by evaluations with 50 university students, as well as an evaluation by a public speaking coach. We found that the students, on average, gave high acceptance scores for the robot and reported moderate interpersonal closeness with the robot, and attributed human nature attributes to it. Additionally, an expert public speaking coach found the system to be able to provide reliable and relevant feedback to students considering their performance, and he also gave useful insights on potential improvements of the system.

Keywords Human-robot interaction · Public speaking training · Human factors · Social robots

✉ Delara Forghani
delara.forghani@uwaterloo.ca

Samira Rasouli
samira.rasouli@uwaterloo.ca

Moojan Ghafurian
moojan@uwaterloo.ca

Melanie Jouaiti
m.jouaiti@bham.ac.uk

Chrystopher L. Nehaniv
chrystopher.nehaniv@uwaterloo.ca

Kerstin Dautenhahn
kerstin.dautenhahn@uwaterloo.ca

[1] Department of Electrical and Computer Engineering, University of Waterloo, 200 University Ave., Waterloo, ON N2L 3G1, Canada

[2] Cheriton School of Computer Science, University of Waterloo, 200 University Ave W, ON N2L 3G1 Waterloo, Canada

[3] Department of System Design Engineering, University of Waterloo, 200 University Ave., Waterloo, ON N2L 3G1, Canada

[4] School of Computer Science, University of Birmingham, Birmingham B15 2TT, UK

1 Introduction

Social robots are "designed to interact with people in human-centric terms and to operate in human environments alongside people" [1]. Social robots have the potential to be used in different roles, especially in educational settings as coaches and instructors [2].

Mastering in public speaking is important in both academic and industry settings. People who work in academia typically need to present their work in conferences, lectures, workshops, etc. Moreover in client-facing roles in industry, one should possess excellent communication skills, e.g. to effectively present a product to a client or an administrative team. Therefore, practicing and rehearsing presentations during higher education is essential as it enhances the quality of the presentation [3–6]. Nevertheless, reasons such as lack of time for practice, absence of an audience, a coach, or knowledgeable person, absence of a practice site, public speaking anxiety, or lack of understanding of the importance of rehearsals can prevent someone from participating in rehearsals [3, 7]. A survey of 2,501 public interest professionals revealed that only 45% of them always or usually

Published online: 18 September 2025

rehearse a presentation, while 35% rarely or never rehearse [3, 8]. Public speaking anxiety (PSA) is one of the most common types of social anxiety. According to studies, approximately 15% to 30% of the total population experience PSA [9, 10]. In a survey involving 1,500 undergraduate and 300 graduate students, a total of 35% of students reported having PSA [7]. A survey conducted with students in the United Kingdom showed that 80% of students reported oral presentations as a source of social anxiety [4, 5]. Towards the end of the pandemic in 2020–2021, most university presentations transitioned from their online format back to in-person settings. However, there is little evidence of how this shift has impacted the fear of public speaking [4]. A recent study with 1054 undergraduate students at the University of Waterloo revealed that public speaking is the most anxiety-provoking situation for students among all situations common in academic contexts [11]. Concerns about making mistakes, lack of experience, and facing humiliation from the audience are among the negative concerns associated with PSA. People with PSA are more likely to express speech disfluencies and avoid making eye contact with the audience [12, 13]. A survey conducted with 46 undergraduate and graduate students identified six themes regarding fear of oral presentations: fear of being judged, the negative impact on the university experience, uncertainty about the topic, physical symptoms, concerns about practicing and preparation, and the need for more practical support [4].

To practice for presentations, receiving feedback from a source that aims to improve the presentation delivery, such as public speaking coaches is useful. Public speaking training can be especially helpful for those in the early stages of public speaking experience, such as beginners or students who want to improve their presentation speaking skills [14]. Also, continuous exposures to simulated presentations could contribute to reducing anxiety for a real presentation [6]. Human coaches often have limited availability and time constraints [7, 11, 15]. Additionally, hiring a human coach can involve significant costs, and the fear of negative evaluation by a human may adversely affect presenters [7, 11, 15]. Also, the feedback given by individuals can be biased based on their personal preferences [16]. Technology-based interventions, such as virtual audiences and conversational agents, can be a solution to mitigate these limitations and to assist participants in improving their presentations [6, 8, 10, 17–21].

Social robots with anthropomorphic characteristics and social embodiment can promote a "sense of presence" of another individual in participants [2]. Therefore they can appear more naturally and engagingly compared to other forms of interactive technology [22]. Studies have demonstrated the effectiveness of human-robot interactions, specifically for addressing stress and anxiety for individuals with or without social anxiety [23]. Notably, there is growing interest in the research community in adapting robots for coaching and instructional roles [2].

In a previous study, a robotic coach, RoboCop, was developed for public speaking training using an anthropomorphic robot head called Furhat [6]. In the role of a public speaking coach, RoboCop provided verbal feedback to speakers, considering both slide-level feedback and overall feedback on key aspects of the presentation, including speech quality, content coverage, and audience orientation. In a within-participants study with 30 students, the robot head was compared to visual feedback, as well as spoken feedback without a robot, and the robotic coach demonstrated significant improvements in the presenters. The focus of this study was to examine how this automated feedback could contribute to observable improvements in presentation quality over time. However, this study and other previous studies did not evaluate the feedback provided by the robot from either the speaker's point of view or an expert's perspective [6].

In terms of the feedback mechanism, studies have focused on the timing of the feedback [21], the feedback delivery method (which is related to the embodiment of the technology and the scope of its capabilities) [20, 24]. One way to deliver feedback is through verbal communication, which is appropriate for embodied agents. In our previous study, we compared a voice assistant agent and a social robot in public speaking rehearsal coaching roles, and the social robot turned out to be more likable [25].

In the present study, we decided to replicate part of the system development carried out in RoboCop (an anthropomorphic robot head) [6], using a different, expressive humanoid robot and test it with a larger population of university students. We used the Pepper robot, which has a humanoid embodiment and can display a wide range of non-verbal expressions and movements with its body. In future prototypes of our system human public speaking coaches could be involved during the whole development stage as part of a co-design process [26–29], which was not possible at this stage due to time limitations. Another salient difference between the Furhat robot (RoboCop) and Pepper is their different sizes. Among humans, studies have shown that taller people are perceived as more competent than shorter ones [30]. However, it is unclear how these findings would translate to HRI studies in which people manipulate a robot's height [31, 32]. Moreover, instead of investigating the effects of different feedback modalities on speakers' performance improvement, our study investigates the usability and acceptance of a social humanoid robot as a public speaking coach. We incorporated an expert evaluation of the system in addition to the user evaluation. This decision was based on findings indicating that video-based remote heuristic evaluations by expert evaluators are a valuable

approach for identifying usability problems in human-robot interaction design during an early development phase [33]. We aimed to evaluate the robot in its coaching role from two perspectives: the speakers' views of the robot's feedback, compared to their self-evaluations, as well as an expert's view when observing the robot's performance in assessing speakers' performances. The ultimate goal of this study is to propose a potential application for social robots as long-term robotic coaches for public speaking rehearsals at the beginner level for university students. Before moving towards this goal, in this study, we developed a humanoid social robot as a public speaking coach and explored how 50 participants perceive such a system, and their preferences for future use, with very encouraging results that will inform the future development of the system. The study received ethics approval from the University of Waterloo Human Research Ethics Board.

2 Background Review

In this section, we review the literature that motivated the design of our study in several related areas of research.

2.1 The Importance of Prosodic Aspects of Speech

Prosody is essential in defining the quality of speech. Public speaking guides suggest using dynamic vocal delivery, represented by variation in intonation, rhythm and volume [34]. Greater variation in fundamental frequency (F_0), and speech rate is correlated with higher perceptions of liveliness in speech [35, 36]. In terms of liveliness, pitch variety affects a speech's categorization as either monotone or lively, with the effect found to be stronger in male speakers than in females [35]. Other attributes contributing to the increased quality of a presentation include the speaker's ability to maintain eye contact with the audience and being aware of their non-verbal communication such as body language [37].

Studies on the charismatic effect of speakers showed that the perception of charisma in a speaker is not only about visual aspects but is also associated with what is being said (the content) and how it is being delivered [38, 39]. Studies have found that the factors related to the fundamental frequency or F_0, speech intensity and speech rate are positively correlated with perceived charisma [40]. In a study, a comparison was made based on samples of Steve Jobs's and Mark Zuckerberg's presentations [41, 42]. They chose Steve Jobs because he has been known as one of the most charismatic speakers [42]. According to Niebuhr (2016) [42], there have been many articles in newspapers and blogs aiming to describe his "presentation secrets" [43] and advise on how to emulate his presentation delivery style [44]. They chose Zuckerberg as the less charismatic speaker compared to Steve Jobs, according to a CNN article by Sutter (2011) [45] that compares the two CEOs on an impressionistic level. Extracting the two CEOs' acoustic features showed that Steve Jobs had a very high F_0 level and a larger F_0 range (equivalent to a higher pitch variety) than average reference speakers collected from a set of previous studies. Analysis of acoustic features in Zuckerberg's speeches showed that he had a slightly higher speech rate than the reference values. Moreover, his acoustic features (speaking rate, hesitation duration, hesitation count, F_0 level, F_0 range, etc.) were also higher than reference values but not as significantly as Steve Jobs's [41]. Indeed, charismatic speech is associated with a high F_0 level, a larger F_0 range, and a relatively fast, however not too fast, rate of speech [41].

2.2 Coaching Systems for Public Speaking

Existing literature has highlighted the potential of interactive technologies for rehearsing public presentations and supplementing the currently available support services in public speaking training. For example, Hoque et al. (2013) [18] developed a virtual conversational agent called MACH for mock interview sessions. The agent would ask interview questions, monitor facial expressions and speech prosodies, and would react with verbal and non-verbal behaviour in real-time. The system would give feedback on the total pausing time, speech rate, amount of fillers, pitch variation, intensity of participants' smiles, and voice loudness. The feedback included both a summarized and a focused version presented in a visual format. According to experts, students who practiced with MACH showed improvement in their overall performance throughout the interview, while the students in a control group did not.

Trinh et al. (2014) [8] developed a system, an add-in in Microsoft PowerPoint called PitchPerfect, to track content coverage and time management during a presentation. The findings showed that this system can significantly improve overall presentation quality and content coverage, and it helps facilitate mastery of content, presentation timing, and confidence.

Tanveer et al. (2015) [21] utilized Google Glass, a head-mounted wearable, as a tool to provide visual real-time feedback. The device monitored the presenter's voice loudness and speech rate and offered feedback accordingly. The feedback strategy was either continuous throughout the presentation or sparse, that is, at regular time intervals. The goal was to increase the utility while reducing the distraction. The results showed that participants were more satisfied with the sparse feedback strategy.

Schneider et al. (2015) [20] developed a multimodal system called the Presentation Trainer to rehearse for an elevator pitch and provide presenters with real-time feedback, on nonverbal aspects such as body posture, gestures, voice loudness, and using pauses as well as filler sounds or phonetic pauses. A mirrored image of the presenter would be displayed on a screen for the control group, and for the experimental group visual or haptic feedback would be also given using the screen and a wristband. The findings showed that participants preferred this method of learning over traditional classroom methods. Saukh et al. (2019) [46] developed a smartphone application named Quantle to measure fundamental acoustic aspects, including pace, pitch, and pause in speech. According to [46], Quantle makes it possible for the user to compare their presentation delivery to other speakers, and receive real-time feedback.

Wang et al. (2020) [16] developed a user interface as a voice coaching system for analyzing multiple prosodic aspects of speech for improvement. A recommendation module was developed and operated as a search engine, using a learning algorithm that retrieved relevant TED talk examples based on the speech content and the voice modulation inputs from the user. This matching process relied on the cosine similarity of high-dimensional data within the system. Participants found the system valuable for improving their skills and expressed satisfaction with their interactions with the system.

Trinh et al. (2015) [47] showed that using a virtual agent as a co-presenter with the presenter results in a significant reduction in PSA among presenters and boosts confidence, especially for non-native English presenters. Kimani et al. (2019) [19] used a virtual agent in cognitive restructuring exercises to reduce public speaking anxiety. They showed that in the treatment condition where the cognitive restructuring exercise was carried out, the presentation practice was more satisfying and enjoyable for presenters, and it led to a significant alleviation in speech anxiety and lower nervousness as self-reported by participants. Wang et al. (2020) [10] utilized the Amazon Alexa smart speaker to guide participants in cognitive reconstruction exercises and varied the sociability level of the voice agent. The results showed that sociability positively influences perceived interpersonal closeness, resulting in the reduction of pre-speech anxiety and boosting satisfaction levels and inclination to use the agent in the future [10].

2.3 Social Robots as Instructors

Many studies have employed social robots as pedagogical agents in educational context [48–51]. Belpaeme et al. identified two main roles, which social robots can take in educational settings: as tutors and as peer learners [52]. They stated that social robots could potentially provide a personalized learning experience tailored to the individual, offering support and challenges that go beyond what is possible in today's resource-constrained educational settings [52]. Pai et al. [53] indicated that younger students prefer a social robot over a private tutor as their learning companion. This preference was attributed to the robot's ability to engage interactively and vividly express emotions [53]. Research has demonstrated their positive impact on the "sense of presence" when using social robots compared to other interactive technologies. Studies have revealed that social robots, when employed as pedagogical agents, demonstrate the ability to provide support, such as positive feedback, thereby positively influencing student motivation, learning, and conformity with the robot's requests [54, 55]. Lemaignan et al. [56] showed that the presence of the Pepper robot integrated into the educational settings of a Special Educational Needs school for children with Autism Spectrum Disorder could improve the well-being of students and the school eco-system [56]. The robot was effective in engaging children through various sensory interactions and offering psycho-social support. According to studies co-located robots are preferred over virtual/animated agents or remote robots in terms of participants' preference, perceived trust, enjoyment, and participant engagement [57–61].

Trinh et al. (2017) [6] employed an anthropomorphic robotic head (Furhat) to analyze five aspects of a presentation such as content coverage, audience orientation, speaking rate, filler rate, and pitch variety. The robot provided participants with automated verbal feedback. The study compared Furhat with voice-only feedback without the robot and with visual-based feedback using a screen. Results indicated that Furhat enhanced presentations compared to the other two forms of feedback. Participants expressed high satisfaction with Furhat and showed a willingness to rehearse with it in the future. While the study primarily focused on participants' rehearsal experiences and worked on the overall improvement in the presentation quality, it did not evaluate how participants and a human public speaking coach would assess the accuracy of the automated feedback generated by the robot, nor explorepotential ways to enhance the feedback.

Furthermore, when creating a natural-language interaction for a social robot, it is crucial to consider more than just the conversation design itself. The effectiveness of a human-robot interaction can be notably influenced by seemingly minor factors, including the robot's physical appearance and non-verbal behaviour [62]. In terms of social robots, social humanoid robots have been shown in other studies to have more favourable impacts on participants. For example, regarding the comparison between Furhat and the Pepper robot, older adults [63], showed a negative attitude toward

and negative social acceptance of Furhat. On the other hand [62] found that Furhat was perceived as displaying emotions better, more intelligent, and more trustworthy in comparison to the Pepper robot, however, both of them were perceived as equally friendly. As mentioned above, the Pepper robot has also been utilized effectively in a therapy context [56]. Considering Pepper is a humanoid full-body robot with a cartoon-like appearance capable of complementing speech with displays of various non-verbal cues and body movements and gestures, we opted for the Pepper robot as the public speaking coach in our study.

3 Research Questions and Hypotheses

Developing the full-bodied humanoid Pepper as a public speaking coach we sought to answer the following research questions (RQs):

RQ1–1 What are acceptance rates and intentions to use the robot as a public speaking coach?

RQ1–2 How much interpersonal closeness do participants perceive with the robot when using it as a public speaking coach?

RQ1–3 To what extent do participants perceive human nature attributes in the robot?

RQ1–4 How do participants evaluate the performance/feedback provided by the robot?

RQ2 What are the views and opinions of a professional public speaking coach regarding the feedback provided by the robot?

The following hypotheses can be proposed to answer the aforementioned research questions in light of the literature, although the research is exploratory in nature.

H1–1 Participants are expected to exhibit a positive attitude towards the Pepper robot and accept it as an appropriate option for presentation rehearsals. The robot's ability to provide sufficient feedback without causing participants to feel nervous about potential negative judgments is anticipated to contribute to this positive attitude [4].

H1–2 While behavioural closeness commonly arises in long-term relationships, it can also manifest in short-term interactions [64]. Therefore, in this study, we interpret interpersonal closeness as a transient sense of connection between the self and another person [64]. Participants are expected to experience subjective interpersonal closeness with the Pepper robot as a coach. This assumption is based on the friendly face of the robot, the active listening behaviour, and the sandwich feedback.

H1–3 We sought human nature attribution for the robot based on studies in counselling and self-disclosure, which indicate that participants tend to be more comfortable with a human due to their familiar experiences interacting with humans. Consequently, the more an agent resembles a human, the greater the comfort people feel interacting with it [65]. Based on the anthropomorphic characteristics of the Pepper robot, and its potential to express non-verbal behaviour, we expect that people will anthropomorphize and ascribe human nature attributes to it that will distinguish it from machine-like behaviour [66, 67].

H1–4 RQ1–4 asks for participants' evaluation of the robot's performance and its feedback as a coach. Essentially, we aimed to confirm the functionality of the speech-processing modules from the presenters' points of view. We anticipate that participants will largely agree on the robot's competence and the accuracy of its feedback.

Note, that we did not hold any particular expectations concerning the insights gathered from the evaluation session with the human public speaking coach regarding the developed system.[1]

4 Methods

We conducted an in-person user study with university students and staff to test our system during presentation rehearsals and gather participants' opinions about our system. The experiment consisted of one rehearsal session with the Pepper robot and involved delivering a short presentation in the presence of the Pepper robot, which played the role of a public speaking coach. The chosen topic for the presentation was the "History of Canada Day". The primary reference for the presentation material was the official website of Immigration, Refugees and Citizenship Canada.[2] The presentation covered the chain of events over subsequent years that led to the celebration of Canada Day on July 1st. We selected this topic to avoid overwhelming participants with a cognitively demanding subject. However, we also

[1] The human public speaking coach was affiliated with the Renison University College, Waterloo, Canada and holds lectures for university students on delivering public academic presentations. The expert coach was contacted after one of the researchers attended one of their classes.

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

aimed to ensure the topic wasn't too easy, incorporating some political and historical events in Canada's history, such as "The British North-America Act," "Constitution Act", and "Anniversary of Confederation."

After the conclusion of the presentation, the robot autonomously provided feedback to the presenter. Each rehearsal session lasted less than five minutes for each participant, with the presentation lasting around two to three minutes and the feedback taking less than two minutes. The feedback encompassed an analysis of the presenter's speech prosody, speech disfluencies, and audience orientation.

4.1 System Implementation

To analyze participants' presentations, we considered multiple metrics in the prosody and vocal aspects of speech, as well as participants' audience orientation. All analysis of acoustic and prosodic aspects of speech was conducted using the "Praat" software [68] and its Python library called "Parselmouth" [69] as it provides us with useful functionalities for speech analysis.

4.1.1 Robot's Non-Verbal Behaviour

Listening Behaviour, and Robot's Engagement: Pepper expressed active listening during the participant's presentation through head nodding. We decided for the nodding movement to occur every six seconds, after some exploratory pilot trials. Based on tests by the research team during implementation, as well as feedback from four pilot participants, we decided that nodding every six seconds feels natural. The vertical head movements alternated between two random patterns. One involved a slow vertical movement, mimicking comprehension and the process of grasping information, while the other comprised two consecutive vertical movements with a relatively faster pace and smaller amplitude, resembling subtle nods [70]. This variation was meant to make the listening behaviour appear as natural as possible, imitating human non-verbal back-channelling through head nods, and to prevent participants from perceiving a repetitive pattern in the robot's behaviour.

Every forty seconds, the robot would lean forward with its torso to further mimic attention. The 40-second period was decided in an exploratory manner, as we did for the nodding, considering that more frequent behaviours might be perceived as distracting, and less frequent behaviours might be perceived as the robot being not attentive enough. Additionally, Pepper's face-tracking module was activated, allowing it to track participants' faces throughout their presentations.

4.1.2 Presentation Analysis

Articulation Rate: In phonetics, the terms "speech rate" and "articulation rate" are often used interchangeably, however, they differ in their definitions. While both represent "the number of output units per unit of time", speech rate includes silent pauses, whereas the articulation rate omits silent pauses in speech [71].

The articulation rate is defined as the pace at which the speaker pronounces speech segments disregarding nuances that the speaker may use while conveying the information, including filled pauses, hesitations, and so forth [71]. In contrast, speech rate considers larger speech attributes, such as the frequency of pauses, laughter, and filler words for instance "you know", "I mean", etc. These attributes normally affect speech fluency, and define a unique communication style associated with the speaker [71]. For this study, articulation rate provides us with a more relevant assessment of cross-dialectal disparities in speech tempo by eliminating additional within-speaker variables [72].

To compute the articulation rate, we utilized counting the syllable nuclei method [73]. To this aim, we used an open-source script for detecting syllable nuclei which are peaks preceded and followed by dips in intensity. The algorithm employs the intensity to find peaks in energy contour to detect and count the centers of syllables or nuclei [73]. Syllable nuclei or the vowels within a syllable (the syllable nucleus) have higher energy than the surrounding sounds. Then the intensity contour is used to ensure that the intensity between the current peak and the preceding peak is sufficiently low [73]. Praat considers a silence threshold of −25 dB or −20 dB. According to the reference, the segments that are at least 2 dB above the median intensity measured across the total sound file were considered as potential syllables [73]. Any segment with an intensity below this threshold was classified as an unvoiced (silence) segment and excluded from the analysis, leaving only the voiced segments [74]. Finally, we calculated the number of syllable nuclei per phonation time (speaking time minus the time of silent pauses).

Pitch Variety: The pitch contour is derived using autocorrelation in Praat. For this study, we considered the pitch floor of 75 Hz, and the pitch ceiling of 500 Hz [75]. According to previous studies, the overall pitch variation is significantly associated with speech quality [76]. To calculate the pitch variety, we computed the difference between the 95th quantile and the 5th quantile of the pitch contour in Hertz [6].

Voice Intensity: To compute voice intensity, we utilized the intensity contour and assessed the mean intensity throughout the entire presentation duration. The threshold values for voice intensity were determined empirically (by

conducting some initial tests inside the experimental room). Minimum and maximum intensity thresholds of 52 dB and 70 dB in Praat, respectively, were established to define the audio signal's quietness and loudness.

Filler Rate: Filler words such as "um" and "uh" were detected using the AssemblyAI speech-to-text transcription functionality. Participants' recorded audio signals were transmitted to the API using its token and an HTTP request. The response contained the transcribed text along with metadata associated with the text. It can be configured to include the number of filler words presented in the audio. The specific filler words considered by the API could be set in the HTTP request. Finally, we divided the number of filler words by the phonation time (speaking time excluding silent pauses) in minutes to calculate the filler rate. The APIs response, which depended on the audio signal's duration, was received with a short delay. In our study, where most presentations were around 3 minutes, the response time for calculating the filler rate was approximately 20 seconds. The AssemblyAI API automatically deletes uploaded files from their servers either once the transcription process concludes or after 24 hours from the initial file upload.

Audience Orientation: Audience orientation was intended to evaluate participants' maintenance of eye contact with the audience. Since presenters were not standing close to the robot, monitoring their gaze would be prone to errors. Therefore, we estimated eye contact using the head pose of participants throughout their presentations. For this purpose, we attached a webcam to the top of Pepper's tablet to capture and track participants' upper body movements. The real-time head pose estimation algorithm utilized the perspective-n-point (PnP) solution [77] to estimate the position of one's head in the three-dimensional space from the two-dimensional input image. The PnP solution produced transitional as well as rotational matrices. The rotational matrix included a three-dimensional angle in space. The angle concerning the y-axis represented head yaw and provided a reliable estimation of the time during which the presenter was looking towards the robot or the simulated audience. We chose an angle of 13° as the threshold for head orientation. If the presenter's head yaw exceeded 13° or fell below −13° it indicated a shift from looking towards the audience.

4.1.3 Robot's Verbal Feedback

All speech parameters and the previously discussed audience orientation (towards the robot) are calculated. As the presentation reaches its conclusion, the results of these calculations are compared to empirically defined thresholds, cf. [6]. We considered these values as appropriate benchmarks, based on insights gathered from literature on successful speech and information collected from experts [6, 21, 35]. We used the thresholds reported in [6]. In their study, eight participants were recruited to subjectively rate the speech measures and eye contact of 20 presentation samples. The presentations were randomly selected from a corpus of 696 samples from different speakers, each 20 seconds long. They calculated the cutoff thresholds based on the positive or negative feedback given by participants. As identifying appropriate ranges for speech measures in public speaking can only be defined based on empirical evidence, we found the aforementioned method acceptable and adopted the thresholds. The robot provides positive or negative feedback based on whether the results of the calculations for each parameter fall within the accepted range or not. Table 1 represents the accepted range for all variables we calculated for each presentation based on a previous study [6].

We utilized the sandwich feedback method for the robot to provide feedback [79]. This method involves sandwiching corrective feedback between two general positive statements. Research has demonstrated its positive effects on the performance of individuals [79]. It assumes that people are more receptive to negative feedback when it follows a compliment, therefore reducing discomfort associated with criticism. Although there have been some recent doubts about the effectiveness of this feedback strategy [80], we decided to adopt the same feedback method as it is widely used [81–83].

The feedback structure is as follows:

1. A brief opening statement conveying appreciation for how interesting the presentation was.
2. Positive feedback consists of acknowledging aspects of the presentation that were successfully conducted.
3. Constructive feedback consists of aspects that could have been improved along with corresponding suggestions for improvement.

Table 1 Defined threshold values for all the variables calculated in a presentation [6]

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

4. A concluding expression of appreciation for the presenter.

Furthermore, we generated corresponding written versions of each positive or corrective feedback. The written notes were displayed on Pepper's tablet, synchronized with the timing of the verbal feedback. The accompanying written versions of feedback were intended to help participants have a concise visual representation of the verbal feedback, facilitating feedback recall for later. Positive feedback was showcased on a green-coloured screen, while corrective feedback appeared on a red-coloured screen.

4.2 Experimental Setup

The experiment room was designed to simulate a presentation rehearsal class, including a lectern with a laptop for the participants to use, presentations projected onto a wall, and a 'simulated' audience. The Pepper robot was positioned on one side of the room, and a podium with a laptop was placed within a distance, directly in front of the robot. The podium was set at an angle, requiring participants to turn their heads away from the robot if they wanted to view the slides on the laptop. A projector was installed in the room, connected to the laptop on the podium, projecting the presentation slides onto the wall behind the participant. To provide contextual information, copyright-free images were incorporated to represent the audience [84]. Figure 1 illustrates the arrangement of audience images surrounding the robot and attached to the wall behind it only as contextual information. It was not possible during the study to have a real live audience (i.e. a group of people who come to each session, behaving in a consistent manner for each participant). We initially considered having pre-recorded videos of an audience, but we excluded this option since the behaviour of that audience would be non-contingent with the presentation and the presenter, and thus might influence the speaker [85–87] as contingent behaviour is crucial in human-human interaction [88]. Finally, instead of participants speaking to a blank white wall, we decided to display pictures of a diverse "potential audience" just behind/around the robot, so that the participants' gaze towards the robot coach and "audience" were close.

A laptop containing post-study questionnaires was positioned on a separate desk for participants to complete the post-experimental questionnaires. Two cameras recorded

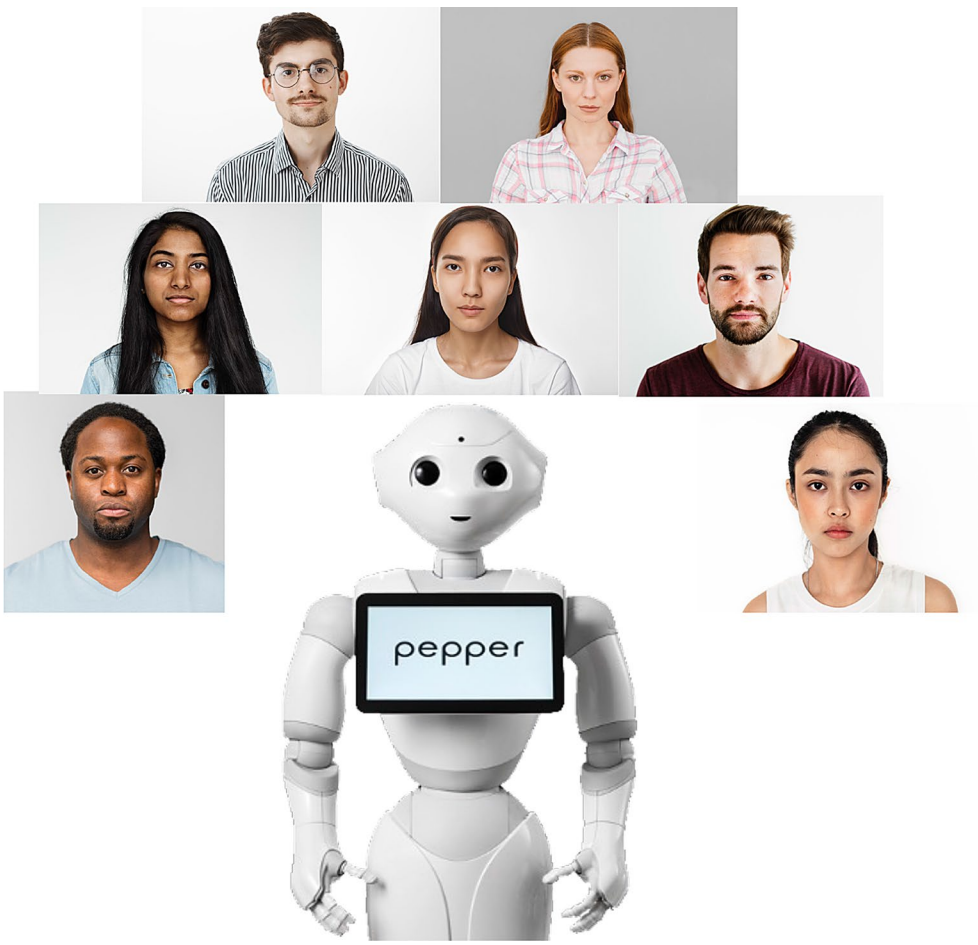

Fig. 1 The image shows the simulated audience using copy-right free photos that were attached to the wall behind and around the robot's location [84]

each presentation session, with one capturing participants from the front view and the other recording the entire experiment room from a side view. The researcher initiated the recording on both cameras just before the participant entered the experiment room. Figure 2 depicts the experiment setup from the perspectives of both cameras.

4.3 Procedure and Measures

Most participants were university students, while a few were staff working at the university. Posters across the campus were used to inform potential participants of the study. Each participant initially received a code, which would be their identification code linked to the data collected from them.

The researcher shared the presentation slides and notes with each participant a few days before their sessions. The notes, meant to help participants to prepare for their presentations, explained a sequence of events that led to the celebration of Canada Day based on Canada.ca. The presentation, which can be delivered in about two minutes, comprises less than one page of notes and consists of only four slides. The first slide shows the topic of the presentation, and the last one includes an appreciation sentence (cf. details in Appendix A). Participants were requested to review the material before the experiment. Additionally, participants were asked to review the consent form and the information sheet, sign them, and complete the pre-study questionnaire.

On the experiment day, we invited participants to our lab and requested them to stand next to the podium, approximately one and a half meters away from where the Pepper robot was positioned. Participants were instructed to stand in front of the Pepper robot as much as possible without walking around. The podium was angled relative to the participant's position so that when they wanted to look at the laptop screen, their head yaw would exceed the orientation towards the robot and the audience pictures. Participants could shift from orienting towards the robot to looking at the projected slides on the wall behind them during the presentation, which is a common situation when giving presentations to an audience.

Participants were provided with brief instructions about the experiment after they entered the experiment room. The experimenter would ask them not to move the podium or the laptop. Also, the researcher would make sure that the webcam attached to the robot would properly capture participants' upper body. Participants were also given a wireless microphone to attach to their shirts. Then the researcher would exit the experiment room, start recording the participant's audio, and ask participants to start their presentation.

Fig. 2 Images of the experimental setup showing the viewpoints that the two cameras recorded during the sessions

The presentation typically lasted around two minutes, after which the robot provided verbal feedback. Before delivering the feedback, the robot would request a brief moment ("Please take a short break, while I am reviewing your presentation.") to review the presentation due to the short delay of the AssemblyAI speech-to-text transcription module. Finally, the Pepper robot asked participants to answer post-study questions using the laptop on the desk. This approach aimed to minimize interactions between the experimenter and the participants until the end of the post-experimental phase. The entire process of data monitoring, analysis, and robot feedback was nearly autonomous, and the experimenter did not interfere during the session. When the presentation ended, the researcher would only stop the listening phase and start the feedback phase, as the system cannot identify the end of the presentation.

4.3.1 Pre-study questionnaire

The pre-study and post-study questionnaires were created using Qualtrics.[3] The pre-study questionnaire covered demographics and participants' prior experience with interacting with robots and robot programming. Participants were asked whether English was their second language or if they were native English speakers. We asked this question to gather additional information about participants' demographics and to determine if there are any differences in ratings given to the robot between English as a Second Language and native speakers. Additionally, they were asked to specify the languages they spoke at home ("What languages do you speak at home?"). The questionnaire also queried about the number of presentations participants had delivered in the past two years, whether they had taken any public speaking classes, and whether they felt the need for such classes. Following that, participants were presented with the two questionnaires below. (Please note that participants were given the option to not answer any of the pre-study questions if they wished, in compliance with the ethics guidelines at University of Waterloo).

Personal Report of Communication Apprehension, Public Speaking Sub-scale (PRCA-PS): The public speaking sub-scale of the PRCA [89] consists of six items, each rated on a 5-point Likert Scale ranging from 1 (strongly agree) to 5 (strongly disagree). This results in a minimum score of 6 and a maximum score of 30, where higher scores indicate greater anxiety about public speaking. Scores falling between 13.75 and 20.75 represent moderate levels of anxiety, while scores above 20.75 indicate high levels of anxiety. Previous studies have demonstrated good reliability and internal consistency of these scores based on Cronbach's alpha [90].

Ten-Item Personality Inventory questionnaire (TIPI): TIPI is a concise method of assessing the Big-Five personality traits with only ten items, including extraversion, agreeableness, conscientiousness, emotional stability, and openness to experiences. The items are on a 7-point Likert scale from 1 (Disagree strongly) to 7 (Agree strongly) [91].

4.3.2 Post-study questionnaires

In this section, we describe the post-study questionnaires that were asked of participants right after their presentations:

Acceptance and Intention to Use The Almere model is a continuation of the Unified Theory of Acceptance and Use of Technology (UTAUT) questionnaire, primarily developed to assess the acceptance of social assistive robots among older adults [92]. Previous studies demonstrated the Almere constructs are solid enough to be used for older adults in various settings [92]. We made minor adjustments to the Intention To Use construct so that instead of "during the next few days," we have "in the future" (since participants could not expect to see the robot again). For the Attitude towards technology, we added "to rehearse presentations" to the sentence "It's good to make use of this robot," and "as a public speaking coach" to the end of "I think it's a good idea to use the robot" to adapt it to the current application (public speaking). For the same reason, in the "Perceived Usefulness" instead of "I think the robot can help me with many things," we asked, "I think this robot can help me with my presentations."

The questionnaire consists of 13 constructs: Anxiety, Attitude towards technology, Facilitating Conditions, Intention to Use, Perceived Adaptiveness, Perceived Enjoyment, Perceived Ease of Use, Perceived Sociability, Perceived Usefulness, Social Influence, Social Presence, Trust, and Use. The Attitude towards technology is an essential factor for Intention to Use. Each construct includes some question items, and the responses are on a 5-point Likert scale from "Totally disagree" to "Totally agree" [92]. According to [92] constructs have interrelations, and some of these constructs are significant in predicting the Intention to Use, such as Perceived Usefulness, Perceived ease of Use, Perceived Enjoyment, Trust, and Attitude towards technology. On the contrary, Anxiety, Perceived Sociability, Perceived Social Presence and Perceived Adaptiveness do not appear to be directly associated with the Intention to Use [92].

Perceived Closeness: The Inclusion of Other in the Self (IOS) Scale is a method used to assess connection or interpersonal closeness with another entity or a group, employing Venn-like diagrams in the form of circles with varying degrees of overlap [93]. The IOS scale has been shown as a

[3] https://www.qualtrics.com.

Table 2 Questions on perceived active listening behaviour - these questions were posed to participants to gauge their perception of the robot's active listening behaviour

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

psychometric suitable measure for closeness which can be applied in different scenarios [93]. In our specific context, we inquired about participants' interpersonal closeness with the Pepper robot. The scoring system is based on assigning a numerical value based on the level of overlap among circles: 1 = no overlap; 2 = little overlap; 3 = some overlap; 4 = equal overlap; 5 = strong overlap; 6 = very strong overlap; 7 = most overlap.

Perceived Human Nature (HN): According to Haslam et al. [94] aspects of human nature can be summarized into emotional responsiveness, interpersonal warmth, cognitive openness, agency, and depth. When any of these aspects is denied, it leads to entering the dehumanization area or a mechanistic perspective, causing the entity to be perceived as inert, cold, rigid, passive, and superficial [94]. In the current study, we evaluated the Pepper robot in terms of its HN attributes using a continuous scale ranging from 0 to 1000 (for all questions in this study we used a continuous scale [95, 96], participants would see sliders with attributes on each end, not showing the recorded values (0–1000), that were recorded according to their choices).

Perceived Listening Behaviour: We asked for participants' perceptions of the Pepper robot's active listening behaviour, considering previous studies that indicate the importance of such behaviour in conveying attentiveness, fostering engagement, and building rapport [70, 97–100]. Participants' perception of the robot's active listening behaviour was crucial, as the robot, in the role of a coach, may need to demonstrate a certain level of comprehension in the presentation rather than mechanistically calculating the parameters. To inquire about the perceived listening behaviour of the robot, we asked questions in Table 2 on a continuous scale.

General Evaluation of the Robot's Behaviour and its Feedback: To assess participants' evaluations of the Pepper robot in its role as a coach, we inquired about the robot's competence, the accuracy of its feedback, and participants' willingness to use the Pepper robot for future rehearsals. To evaluate feedback we considered the comprehensibility and relevance of the feedback with respect to participants' self-evaluations. Responses were provided on a continuous scale. The question items are presented in Table 3.

The Negative Attitudes towards Robots Scale: The Negative Attitudes Towards Robots Scale (NARS) [78] is used to measure anxiety or negative emotions towards robots, which may prevent someone from having dyadic interactions with robots [78]. This questionnaire consists of 14 items divided into three subscales: S1 - Negative Attitude Toward Situations of Interaction with Robots (6 items), S2 - Negative Attitude Toward Social Influence of Robots (5 items), and S3 - Negative Attitude Toward Emotions in Interaction with Robots (3 items). Responses are recorded on a 5-point Likert scale from "Strongly disagree" to "Strongly agree." Due to the way the items are phrased, they can potentially introduce negative biases in participants; therefore, participants answered this questionnaire last.

4.3.3 Expert Evaluation

After completing the study with university students, we invited a professional public speaking coach to evaluate our system, specifically focusing on the accuracy of the feedback provided by the robot. Nine videos, encompassing both the presentation and feedback phases, were randomly selected for assessment. The human coach attended the laboratory in-person and watched the videos on a television screen. To ensure audio clarity and minimize background noise, we provided the human coach with a headphone. Immediately after watching each video, the human coach would answer three questions, with two of them rated on a continuous scale, and the last one recorded as an audio response (with their approval). The questions to be rated by the human coach aimed to gauge the human coach's perspective on the relevance and instructive aspects of the robot's feedback (The question items are shown in Table 4).

Table 3 Questions asked on robot evaluation and feedback assessment

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

Table 4 Evaluation of Robot's feedback

	Questions	Response Scal
After watching each video	What do you think about the robot's feedback, given the presenter's performance?	Very irrelevant Very relevant
	What do you think about the content of the feedback provided by the robot?	Not instructive at all Very instructive
After watching all videos	How would you evaluate the performance of the robot as a coach?	Very unhelpful Very helpful
	Do you think the robot is useful to be used in the University of Waterloo for presentation rehearsals?	Not at all Very much

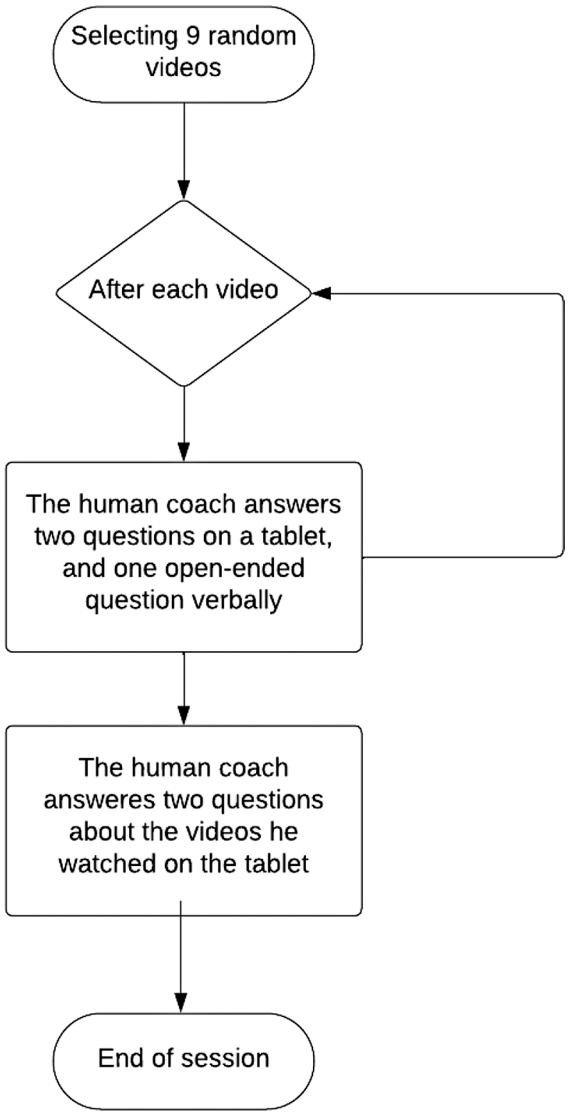

Fig. 3 The diagram shows the procedure we followed for evaluating the system by a human public speaking coach

Then the coach was asked an open-ended question, and the response was recorded on audio: What specific aspects of the feedback could be improved?

After watching all nine videos, we presented two additional questions to the human coach on the tablet to gather an overall assessment of the robot's performance. Figure 3 depicts the procedure we followed as well as the order of questions we asked the human coach.

4.4 Participants

We recruited a total of 48 university students and 2 university staff members for this study by distributing flyers around the campus. The age range was [97, 101], with a mean age of 20.82 years. Among them, 30 self-identified as female, 18 as male, and 2 chose not to disclose their gender. Regarding academic status, 23 were undergraduate students, 21 were master's students, 4 were Ph.D. students, and 2 chose not to disclose their academic status. We acknowledge that the participant sample is biased towards undergraduate students, who usually have less presentation experience than graduate students. Out of the 50 participants, 33 reported having previous interactions with robots, while 17 did not. Additionally, 11 had prior training in public speaking skills, and 28 expressed a need for training in public speaking skills. Figure 4 (a) displays the distribution of the languages participants reported as their primary spoken languages at home. Each participant received 10 CAD remuneration for their participation in this study. Figure 4 (b) represents the probability density function of the PRCA-PS scores.

5 Results

The findings of our evaluation study are presented in this section. The evaluation consisted of one rehearsal session with the Pepper robot.

5.1 Acceptance and Intention to Use

We used the 10 most relevant constructs in the questionnaire, including Anxiety (ANX), Attitude towards technology (ATT), Intention to Use (ITU), Perceived Adaptiveness (PAD), Perceive Enjoyment (PENJ), Perceived Ease of Use (PEOU), Perceived Sociability (PS), Perceived Usefulness (PU), Social Presence (SP), Trust. Figure 5 shows the results of the average scores for each of the constructs accompanied by the 95% confidence interval.

According to Fig. 5, the mean score and the standard deviation of anxious reactions when interacting with the robot was ($M = 1.74, sd = .68$), which shows that participants did not feel anxiety during interactions with the

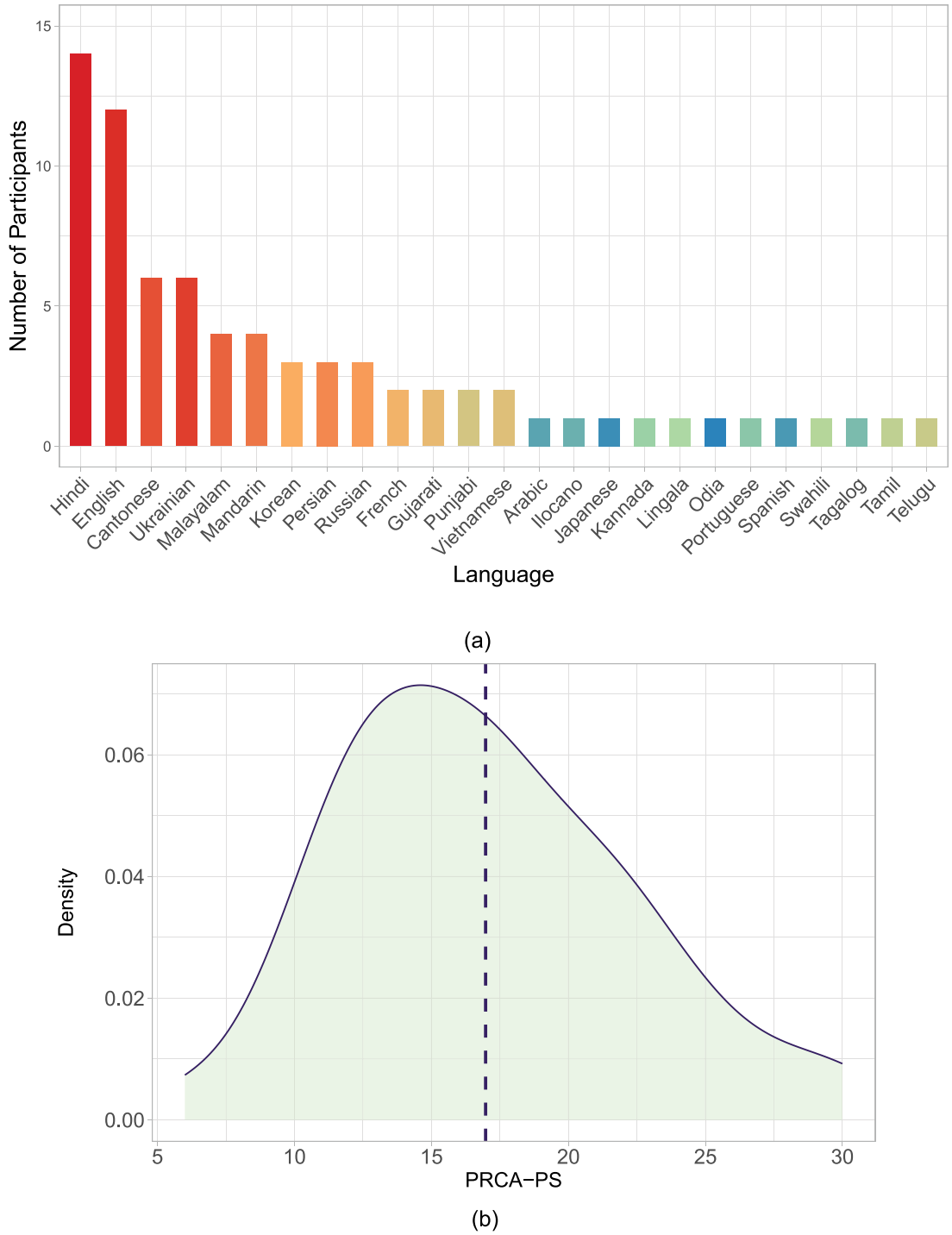

Fig. 4 (**a**) Distribution of languages participants speak at home, showing the number of participants for each language. (**b**) Probability density distribution plot of participants' PRCA-PS scores. The vertical line is the mean of the PRCA-PS scores among participants

Pepper robot. The results for Attitude towards technology ($M = 4.33, sd = .75$) showed positive rather than negative feelings towards appliance of the technology. Perceived Sociability ($M = 4.03, sd = .74$) shows a high perception of sociable behaviour in the robot, Perceived Usefulness ($M = 4.2, sd = .76$) also received a high average score represents the perception of assistive potential in the robot. Finally, the Trust ($M = 4.21, sd = .65$) score was also relatively high and above 4 which indicates that participants assessed the robot's performance as reliable.

The average score for Perceived Adaptiveness ($M = 3.9, sd = .65$), and Perceived Ease

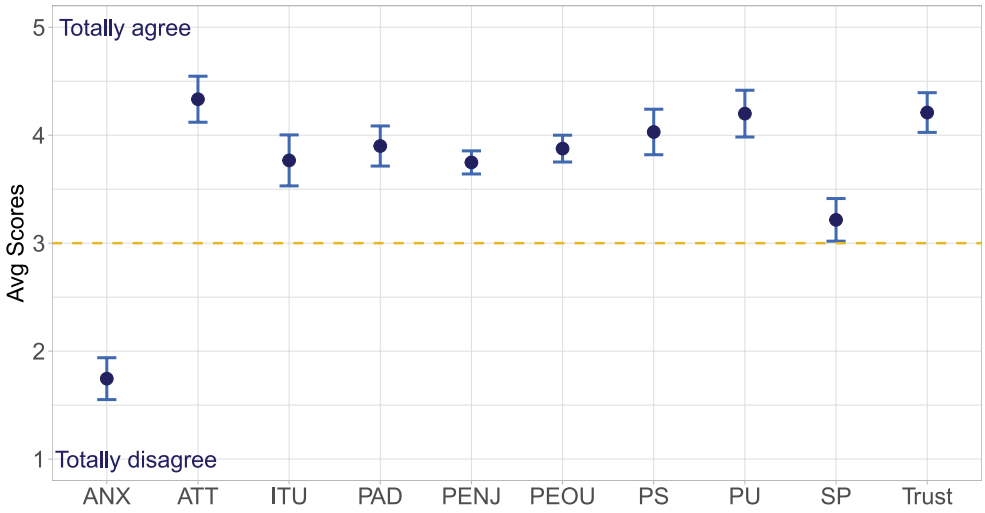

Fig. 5 Almere model constructs scores. Dark blue points show the mean score for each construct and error bars indicate 95% confidence intervals. The yellow line represents the neutral choice

Table 5 Number of participants choosing different diagrams in the inclusion of other in the self (IOS) scale

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

of Use ($M = 3.88, sd = .44$), Intention to Use ($M = 3.76, sd = .83$), and Perceived Enjoyment ($M = 3.75, sd = .38$) were also above the neutral choice. Perceived Social Presence was the only construct that was reported lower than other constructs, and slightly above the neutral choice ($M = 3.22, sd = .69$).

Further analysis: To explore the potential confounding effects of other factors, such as NARS and PRCA-PS scores, on participants' ratings of the robot, we conducted regression analysis using linear models. We ensured the appropriateness of linear models through tests like normality of residuals. The factors considered in the linear models were chosen to minimize multicollinearity and minimize the Akaike Information Criterion (AIC) [102]. We included 43 participants in the analysis since 7 records were missing from the PRCA-PS survey.

According to the results, the anxiety towards the robot (*se*=.006, *t*=4.324, $p < .001$) was positively correlated with the NARS score, and the attitude towards the robot (*se*=.032, *t*=-3.042, $p < .01$), as well as the intention to use the robot (*se*=.036, *t*=-2.862, $p < .01$) was negatively correlated with the NARS score. Moreover, the intention to use the system was negatively correlated with the PRCA-PS scores (se=.053, t=-2.693, p < .05), and previous attendance in public speaking skill classes (*se*=.711, *t*=-2.548, $p < .05$). Participants' perceived usefulness (*se*=.055, *t*=-2.721, $p < .01$), and the trust towards robot's feedback (*se*=.033, *t*=-2.231, $p < .05$) were negatively correlated with the PRCA-PS scores.

5.2 Perceived Interpersonal Closeness

The results of the IOS scale showed the mean score of $M = 3.42$, and the standard deviation of $sd = 1.51$ across all participants. Table 5 shows the number of participants for each diagram.

5.3 Perceived Human Nature Attributes

The findings of the HN questionnaire are displayed in Fig. 6. As shown, the Pepper robot received high ratings in all dimensions of HN except for its emotional responsiveness ($M = 528.64, sd = 275.22$). The mean scores and standard deviations for warmth ($M = 716.84, sd = 234.79$), perceived openness ($M = 685.8, sd = 261.79$) were all above the neutral choice, and agency or liveliness had the highest mean value ($M = 742.66, sd = 229.38$).

Perceived Listening Behaviour: Fig. 7 shows that the scores for the robot's listening behaviour during presentations ($M = 829.98, sd = 154.841$), attentiveness ($M = 837.6, sd = 156.58$), and engagement ($M = 787.32, sd = 179.435$) were notably high.

5.4 General Perception of the Robot as a Public Speaking Coach

We evaluated how well the robot performed as a coach from the participants' perspectives. Figure 8 shows the mean and standard deviation of all measurements were high, including competence ($M = 818.02, sd = 201.92$), perceived helpfulness ($M = 912.16, sd = 126.21$), willingness for future use ($M = 834.2, sd = 178.2$), perceived feedback

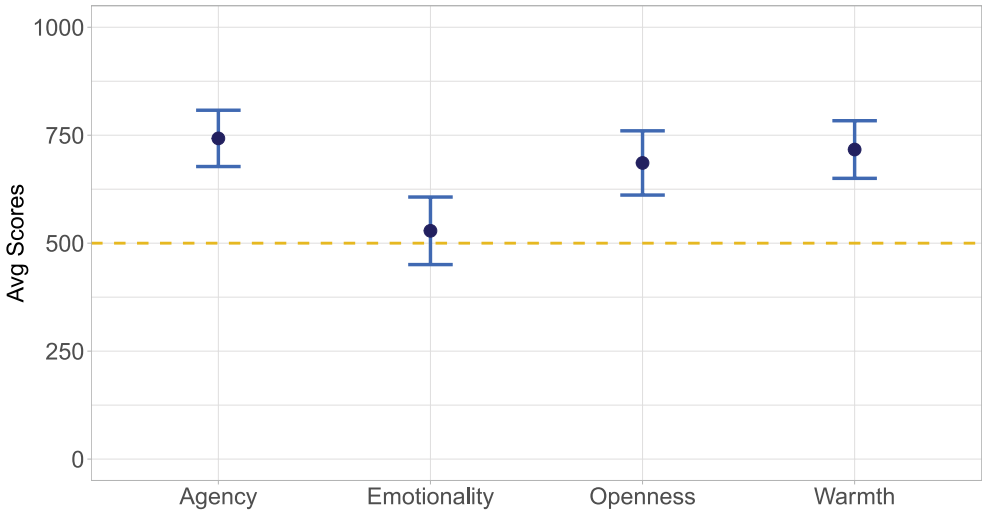

Fig. 6 Human nature characteristics attributed to the pepper robot. Dark blue points show the mean score for each attribute and error bars indicate 95% confidence intervals. The yellow line shows the midpoint on the slider

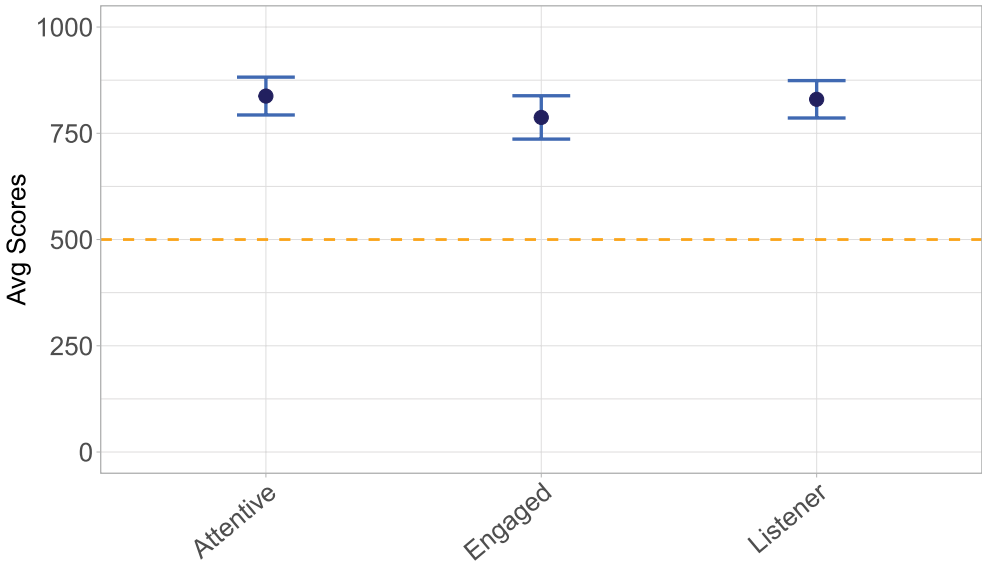

Fig. 7 Perceptions of active listening on the robot. Dark blue points show the mean score for each item and error bars indicate 95% confidence intervals. The orange line shows the midpoint on the slider

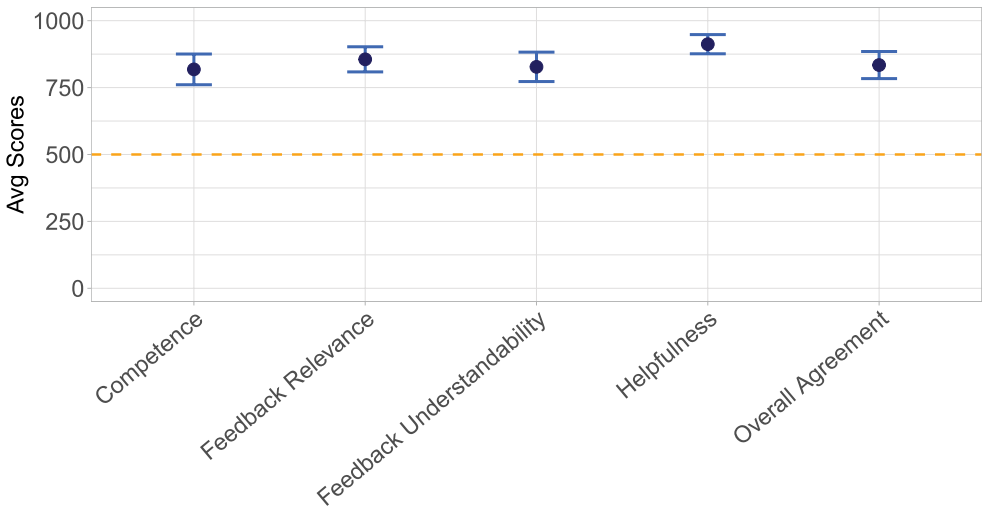

Fig. 8 General perceptions of the robot as a public speaking rehearsal coach and the feedback provided by the robot. Dark blue points show the mean score of each item and error bars indicate 95% confidence intervals. The yellow line indicates the neutral choice

comprehensibility ($M = 827.5, sd = 193.31$), and perceived feedback relevancy ($M = 855.56, sd = 165.12$).

Further analysis: The results of the linear regression analysis revealed significant correlations between perceptions of the robot's active listening behaviour and the assessment of the robot's Human Nature attributes. According to the results, perceived non-verbal active listening behaviour in the robot was significantly correlated with perceived emotional responsiveness ($se = 0.227, t = 3.232, p < .01$), perceived warmth ($se = 0.19, t = 4.294, p < .001$), perceived openness ($se = .209, t = 4.717, p < .001$), and perceived agency ($se = .22, t = 2.439, p < .05$) in the robot. Furthermore, participants' ratings of interpersonal closeness with the robot, as measured through the IOS questionnaire, are positively correlated with perceived non-verbal active listening behaviour in the robot ($se = .001, t = 2.061, p < .05$).

Perceived competence was correlated with perceived feedback relevancy (*se*=.106, *t*=6.99, $p < .001$) and perceived attentiveness of the robot (*se*=.112, *t*=2.51, $p < .05$). Similarly, the helpfulness of the system was correlated with feedback relevancy (*se*=.113, *t*=3.374, $p < .01$) and with the NARS scores (*se*=2.319, *t*=-2.24, $p < .05$). Furthermore, willingness to use the robot as a public speaking rehearsal coach at the university had positive correlations with perceived feedback relevancy (*se*=.099, *t*=7.459, $p < .001$) and negative correlations with NARS (se=2.036, t=-2.436, p < .05).

5.5 Expert Evaluation

In Fig. 9, the evaluations by the expert are presented, with a focus on feedback relevancy as well as the instructional quality of the feedback. Feedback relevancy ($M = 843.22, sd = 107.07$) and instructional quality of the feedback ($M = 842.11, sd = 92.76$) were assessed to be on the positive extremes, indicating agreement between the human coach and the robot's evaluations of the presenters.

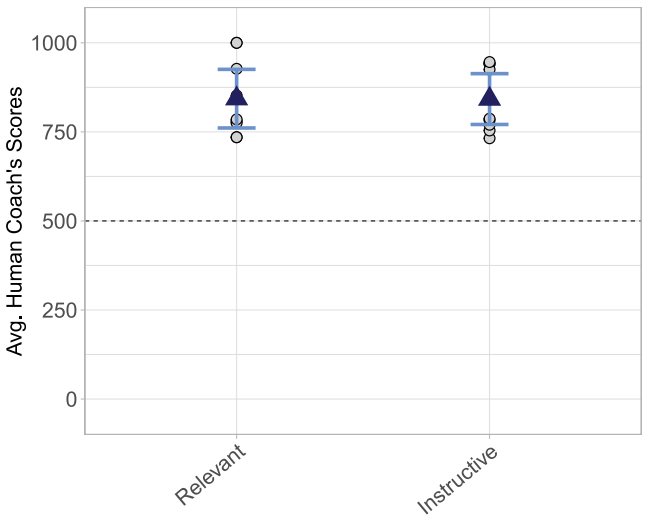

Fig. 9 Perceptions of the human coach regarding the feedback relevancy and feedback instruction quality provided by the robot as a coach. The dark blue triangle shows the average human coach's ratings over nine videos, and the error bars indicate 95% confidence intervals

In the qualitative analysis, we transcribed the human coach's responses recorded after each video for improving the robot's feedback into a textual format using the AssemblyAI speech-to-text transcription API. Subsequently, we conducted a qualitative analysis of the transcribed text to gain valuable insights into the robot's performance and identify areas for improvement in feedback, as perceived by the expert human coach. For the qualitative analysis, first the researcher who interviewed the human coach identified the main points from the expert's responses regarding aspects when he agreed or disagreed with the robotic coach. These points were then shared with the co-authors. The research team then jointly reviewed and summarized the points to retain the most important ones. The main identified points are shown in Table 6.

The results of the final two questions Table 4 aimed at obtaining an overall assessment of the robot's performance are as follows. For the first question, the score provided by the human coach was 918, indicating his overall positive evaluation of the robot's performance as a coach. The second question was answered by the human coach with a score of 838, demonstrating his general support for having an agent at the university to assist with presentation rehearsals.

6 Discussion

We conducted an evaluation study utilizing the Pepper robot as a public speaking rehearsal coach. The robot delivered verbal feedback to presenters, addressing both positive and negative aspects while providing suggestions for improvement. The selection of the Pepper robot for this study was motivated by its anthropomorphic design and its diverse expressive features, including the ability to display arm movements and hand gestures. We aimed to understand participants' acceptance and intention to use the robot for presentation rehearsals, assess their perceived interpersonal closeness, evaluate the robot's human nature attributes, and gauge the robot's competence as a coach. Additionally, we sought expert evaluation to assess the robot's feedback on some selected presentations.

6.1 Acceptance and Intention to Use

The ratings that participants gave to the Almere model construct support H1–1 on accepting the technology and intending to use it. Specifically, the constructs that are

Table 6 The results of the qualitative analysis of the open-ended question after each video in the feedback session with the human expert, where points of agreement and disagreement were identified

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

correlated with Intention to Use, namely Perceived Usefulness, Perceived Ease of Use, Perceived Enjoyment, Trust, and Attitude towards technology, received high scores on average. According to the results of the Almere model, participants reported low evoked anxious reactions when rehearsing with the robot. This has both positive and negative implications for the experiment. On the positive side, participants felt low anxiety, partly because they were aware that the robot could not form a negative judgment of them. This anxiety reduction could be associated with the fear of negative evaluation that people may feel in front of humans. Therefore, the robot could provide participants with a comfortable rehearsal session without the fear of judgment. On the negative side, a counterfactual to consider is that the scenario did not effectively replicate a real presentation session. Thus, participants' experiences in the current experiment may not be comparable to what they would experience in a real presentation.

The findings indicate that Almere constructs received high average scores, particularly in Attitude Towards the Agent, Perceived Usefulness, Trust, Perceived Sociability, Perceived Ease of Use, Perceived Enjoyment, and Perceived Adaptiveness. By examining the interrelations between Almere constructs [92], we can infer participants' inclinations and their willingness to use the system, as reflected in the high scores given to the robot.

Apart from the results considering the average scores, participants with higher self-reported public speaking anxiety exhibited less trust in the robot's feedback and perceived lower usefulness in the system. Moreover, higher public speaking anxiety led to a decline in the intention for future use. It is important to note that these negative correlations do not necessarily imply that people with higher public speaking anxiety may have a negative opinion about the system, as the error bars and confidence intervals in ratings showed very small intervals. However, this can indicate a lower preference associated with people with higher public speaking anxiety compared to those with lower levels of anxiety, which may be related to their reluctance or discomfort with the mere act of delivering presentations. For further elaboration on this hypothesis, future research can conduct interviews with individuals with high self-reported public speaking anxiety to understand their feelings when giving presentations in front of the robot.

Participants with higher negative attitudes toward robots, as measured by the Negative Attitudes toward Robots Scale (NARS), exhibited more anxiety toward the robot, as measured by the anxiety construct (ANX) in the Almere model. This aligns with prior research that showed a positive correlation between negative attitudes towards robots, perceived anxiety, and subsequent behaviours such as communication avoidance with robots [15]. NARS was also negatively associated with the attitude towards technology(ATT), and intention to use (ITU) in the Almere model.

6.2 Perceived Closeness

As a result of analyzing the IOS scale, H1–2 is partially supported by the available data, as 70% of the participants selected from "Some Overlap" to "Most overlap" in terms of their perceived interpersonal closeness with the robot. Moreover, participants who perceived the non-verbal active listening behaviour in the robot reported more interpersonal connections and rapport with it. This observation is consistent with previous research on active listening behaviour in human-human counselling that can contribute to an increased sense of closeness in interpersonal relationships [101, 103]. One possible explanation for that is the perception of active listening activates a brain region called the ventral striatum that plays an important role in processing rewards [104]. This region can encode a wide array of rewards, including those of abstract rewards known as "warm glow", which is the pleasant feeling experienced when receiving acts of kindness [104]. The recognition of active listening can also represent a warm glow that initiates a positive feeling in participants during the interaction with the robot.

6.3 Perceived Human Nature

The evaluation of human nature attributes in the robot showed scores leaning more toward human-like characteristics than machine-like ones, and they surpassed the neutral midpoint. Our results showed that participants ascribed three of the four measured human nature attributes to the robot, including warmth, openness, and agency. Human nature (HN) traits distinguish factors between humans and machines [94]. It is conceivable that attributing these traits to a robot not only makes it more human-like in people's perceptions but also less machine-like at the same time [105]. Higher perceived human nature attributes can contribute to creating a more familiar communication experience for presenters, akin to interacting with a human. This, in turn, has the potential to enhance the amount of information shared, the manner in which information is transferred, and the overall duration of the interaction [65]. Consequently, the results partially support H1–3.

Among perceived human nature attributes in the robot, emotional responsiveness stood out as relatively lower compared to other attributes. Emotional responsiveness is consistent with the concept of empathy, which refers to the ability to recognize 1's emotional state and respond with appropriate affect [106]. In this study, the robot's role as a coach was to provide objective and impartial evaluations of

participants' presentations without incorporating affective responses. Therefore, it is reasonable that participants did not perceive emotion in the robot. Additionally, it is important to note that designing the robot's personality or emotional responses was not within the scope of this study.

According to [107], perceptions of anthropomorphism and the sense of human nature in an agent can be linked to the agent's embodiment as well as its verbal and non-verbal social behaviour. We found positive correlations between the perceived human nature attributes of the robot and the perceived robot's active listening behaviour during the presentations. This shows that participants who perceived the robot as more actively listening during the presentations rated it more on its human nature traits.

6.4 Evaluation of the Robot as a Public Speaking Coach

The results of the perceived active listening behaviour in the robot revealed that participants did perceive the non-verbal backchanneling, leaning forward, and face tracking as means of active listening, consequently indicating attentiveness in the Pepper robot. These findings provide support for H1–4, as participants gave high scores to the Pepper robot as a coach, both in terms of its competence, helpfulness, and future use and the relevance of the robot's feedback compared to their self-evaluations. Our findings reveal that participants' ratings of the robot's competence, helpfulness, and participants' willingness to use the system are correlated with the relevancy and accuracy of the feedback provided. It is also correlated with perceived active listening behaviour in the system. It can be inferred that participants' evaluations of the system's effectiveness as a rehearsal coach are linked to both the robot's appropriate verbal feedback and its non-verbal active listening behaviour.

6.5 System Evaluation by a Human Coach

We evaluated our system from both participants' perspectives and an expert's perspective. The ratings provided by the human coach for the robot's feedback demonstrated a high level of agreement between his evaluations of presentations and the robot's assessments, on average. However, there are areas where further improvement is possible. In the end, the human coach was also asked to evaluate the usefulness of such a system and its potential future application at the university, and he provided high scores for both.

The questions requiring the human coach's rating focused on the accuracy of the feedback regarding the presenter's performance and its instructive aspect for the presenter. A qualitative analysis of the human coach's answer to the open-ended question concerning improving the robot's feedback revealed some limitations and highlighted areas for future enhancements. According to the human coach, the robot's analysis of articulation rate and voice volume was nearly accurate. However, in terms of assessing variations in pitch, he suggested that, in general, the robot could have been less lenient on participants and could have expected more variations in intonation as satisfactory. Highlighting the cultural aspects of expectations regarding variations in pitch, the human coach pointed out that in North America, a charismatic and engaging speech usually requires a broad range of pitch variations with specific stress on keywords in presentations. Similar considerations apply to the evaluation of eye contact, as there were cases where the human coach anticipated a higher degree of eye contact maintenance during the presentations. Further improvement suggestions provided by the human coach are explained in Sect. 7.

According to the human coach, analyzing speech rate could be added in addition to the articulation rate to incorporate the number and duration of silent pauses in speech. As speech rate serves as a valuable measure for assessing speech fluency. Therefore, the robot can analyze and identify silent pauses in the presenter's speech, and provide suggestions for when the presenter should take a deliberate pause to reduce the overuse of filler words.

7 Limitations and Future Work

The system described in this paper has certain limitations that warrant consideration for future research. One limitation is that the robot evaluates presentations using an overall performance assessment based on an average-based strategy, without considering subtle flaws that may occur prominently during specific intervals of the presentation. In the future, we plan to enhance the system by providing the robot with visualization memory, for building personalized interactions with participants and being able to track participants' improvements during multiple sessions. This improvement aims to offer more in-depth and personalized feedback to each participant. Moreover, calibrating participants' voices before the rehearsals to determine their average pitch and speaking rate could make the system more personalized, which can be addressed in the future.

As it was beyond the scope of this study, we did not involve any comparison with a human coach. Note that the main objective of this study was to assess the user acceptance and experience of using a robot as a public speaking coach to explore the potential of such technology-based interventions to help mitigate the current issues of limited accessibility to human coaches.

The current system does not account for accented speech, a significant consideration in a culturally diverse society

such as Canada. Some languages exhibit stress-timed or syllable-timed characteristics [108], which may affect the compatibility with varying pitch during speech. The participants' first language and the cultural context they grew up in significantly influenced their presentation delivery styles, including intonation patterns and eye contact tendencies. Notably, in North America, there is a general preference for more pitch variation and maintaining eye contact during presentations. Students residing in North America often need to adapt to these preferences for various reasons, including job prospects.

Regarding pitch variation, future developments could adopt a more granular approach by analyzing the intonation of specific utterances. For instance, rising intonations may convey uncertainty or seek feedback, while falling intonations can indicate confidence or assurance in a statement [109–111]. Analyzing these nuances in intonation during specific phrases could enhance the system's ability to provide more detailed and context-aware feedback on participants' presentation styles.

Additionally, a potential avenue for future research is to demonstrate the robot's understanding of the content by enabling it to pose questions related to different slides after the presentation concludes. This capability could enhance the robot's explicit demonstration of attentiveness and comprehension of the content. Furthermore, it could encourage participants to rehearse question-and-answer practices.

Another valuable addition to the rehearsal coach could involve enhancing its ability to establish synchrony with the participants, thus improving the sense of interpersonal closeness. Previous studies have highlighted the significance of synchrony in communication and entertainment, where speakers adjust their speech in harmony with their partners, including factors such as pitch, intensity, and jitter [112]. This aspect warrants further analysis and exploration in future research. Our study was not a controlled experiment, as our aim was to explore a potential application for a social robot; therefore, it did not involve any control/baseline condition. However, future work can include a baseline condition with a human coach and/or conditions involving other agents for the purpose of comparison.

A limitation of our study is that the expert evaluation involved only one human coach due to our limited time frame. In future work multiple human coaches could be involved in evaluating the robot. Future work could also consider longer presentations, e.g. 10–15 minutes which is typical of a conference presentation. However, a 3-minute presentation, as used in our study, is not unrealistic. There are frequent course presentations and 3-minute thesis presentation competitions in university higher education. In these contexts, participants are required to strictly adhere to the time limit and deliver an effective presentation within 3 minutes. Future work may also compare different types of social robots.

8 Conclusion

We developed an autonomous public speaking robot system to analyze vocal aspects of participants' performance during their presentations and their audience orientation.

Fifty participants rated the Pepper robot coach with high scores in terms of acceptance and intention to use. Participants perceived a high relevance between their performance and the feedback provided by the robot. They also experienced the robot showing non-verbal back-channelling during their presentation. There was a positive correlation between perceived competency by the robot and the aforementioned variables, namely, feedback relevancy and expressing active listening. Therefore, two of the functionalities that a social robot may need to possess to be accepted by users as a public speaking coach are proper feedback and active listening.

The majority of participants also reported some interpersonal closeness with the social robot as a coach. Participants also attributed human-like qualities to the robot. Further analysis showed that perceived human-like qualities were related to the perceived non-verbal active listening exhibited by the robot. Participants who had negative attitudes towards robots in general and those with self-reported high levels of public speaking anxiety indicated less intention for future use as well as less acceptance towards the robot.

The expert human coach's assessment generally confirmed that the robot coaching system demonstrated a high level of feedback relevance and informativeness. However, some suggested improvements to the robot's feedback will be addressed in future work.

Appendix A Presentation Material

History of Canada Day

Canada's national holiday is celebrated on **July 1**. It is called “Canada Day” and is celebrated to appreciate Canadian history, culture, and achievements.

July 1, 1867: Canada was born with the British North America Act , also known as the Constitution Act.

1879: A federal law makes July 1 a statutory holiday as the "Anniversary of Confederation," which is later called "Dominion Day." – an opportunity for everyone in Canada to celebrate!

From 1958 to 1968: The government organizes celebrations for Canada's national holiday every year. A typical format includes events on Parliament Hill in Ottawa. Parliament Hill is the home of the Parliament of Canada. The celebrations involve, for example, ceremonies in the morning and at sunset, followed by a concert and fireworks. These are many opportunities to celebrate Canada!

From 1968 to 1979: A large multicultural celebration and a concert are broadcast from Parliament Hill on television across the country. The main celebrations (called "Festival Canada") are held throughout July.

1981: In many major Canadian cities, July 1st is celebrated with fireworks, a tradition that is still practiced today and one that is enjoyed by a great number of Canadians

1982: "Dominion Day" officially becomes **Canada Day**, which is what today we are celebrating every year on July 1st.

2014: Canadian Heritage organizes the 147th Canada Day celebrations. On Canada Day, several festivities are organized.

2017: A wide range of activities across Canada mark the 150th Anniversary of Confederation- the birthday of Canada!

Source link

Fig. A1 Presentation notes

Fig. A2 Presentation slide 1

Fig. A3 Presentation slide 2

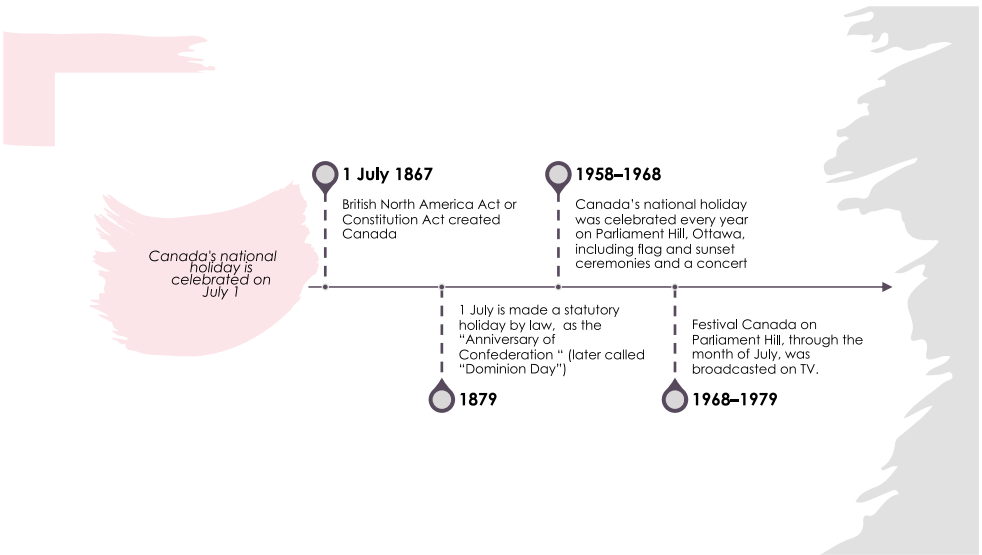

Fig. A4 Presentation slide 3

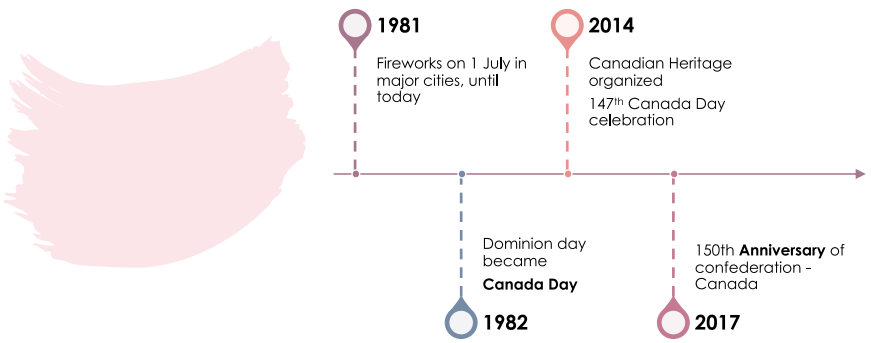

Fig. A5 Presentation slide 4

Acknowledgements The authors would like to express their appreciation to Dr. Kent Williams for evaluating the system and sharing his valuable ideas for improving it, and to Pourya Aliasghari for assisting in data analysis.

Author Contributions DF, SR, MG, MJ, CLN, KD designed the study. DF implemented the apparatus, performed the experiment, conducted the data analysis, and drafted the manuscript. SR, MG, MJ, CLN and KD reviewed and edited the manuscript, and also supervised the project.

Funding This preparation of this manuscript was undertaken, in part, thanks to funding from the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Canada 150 Research Chairs Program.

Data Availability Due to ethics considerations we don't think we can make the raw data (students' speeches) available.

Declarations

Compliance with Ethical Standards The study received ethics approval from the University of Waterloo Human Research Ethics Board.

Conflict of Interest The authors declare that they have no conflict of interest.

References

1. Breazeal C, Dautenhahn K, Kanda T (2016) Social robotics. Springer International Publishing, Cham, pp 1935–1972. https://doi.org/10.1007/978-3-319-32552-1_72
2. Rasouli S, Gupta G, Nilsen E et al (2022) Potential applications of social robots in robot-assisted interventions for social anxiety. Int J Soc robot 14(5):1–32. https://doi.org/10.1007/s12369-021-00851-0
3. Goodman A, Communications C (2006) Why bad presentations happen to good causes: and how to ensure they Won't happen to yours. Cause Commun. https://books.google.ca/books?id=tcsPAQAAMAAJ
4. Grieve R, Woodley J, Hunt SE, McKay A (2021) Student fears of oral presentations and public speaking in higher education: a qualitative survey. J Furth High Educ 45(9):1281–1293. https://doi.org/10.1080/0309877X.2021.1948509
5. Russell G, Topham P (2012) The impact of social anxiety on student learning and well-being in higher education. J Ment Health 21(4):375–385. PMID: 22823093, https://doi.org/10.3109/09638237.2012.694505
6. Trinh H, Asadi R, Edge D et al (2017) Robocop: a robotic coach for oral presentations. Proc ACM Interact Mob Wearable Ubiquitous Technol 1(2). https://doi.org/10.1145/3090092
7. Bishop J, Bauer K, Becker E (1970) A survey of counseling needs of male and female college students. J Educ Chang Coll Student Devel 39:205–210
8. Trinh H, Yatani K, Edge D (2014) Pitchperfect: integrated rehearsal environment for structured presentation preparation. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. Association for Computing Machinery, vol 14, New York, NY, USA, CHI, 1571–1580). https://doi.org/10.1145/2556288.2557286
9. Tejwani V, Ha D, Isada C (2016) Observations: public speaking anxiety in graduate medical education–A matter of interpersonal and communication skills? J Grad Med Educ 8(1):111
10. Wang J, Yang H, Shao R et al (2020) 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. Association for Computing Machinery, vol 20, New York, NY, USA, CHI, 1–13. https://doi.org/10.1145/3313831.3376561

11. Rasouli S, Ghafurian M, Nilsen ES et al (2024) University students' opinions on using intelligent agents to cope with stress and anxiety in social situations. Comput In Hum Behav 153:108072. https://doi.org/10.1016/j.chb.2023.108072; https://www.sciencedirect.com/science/article/pii/S0747563223004235
12. Ayres J (1988) Coping with speech anxiety: the power of positive thinking. Commun Educ 37(4):289–296. https://doi.org/10.1080/03634528809378730
13. Bodie GD (2010) A racing heart, rattling knees, and ruminative thoughts: defining, explaining, and treating public speaking anxiety. Commun. Educ 59(1):70–105. https://doi.org/10.1080/03634520903443849
14. Morrison T (2017) The Book on Public speaking. Morgan James Publishing
15. Nomura T, Kanda T, Suzuki T et al (2008) Prediction of human behavior in human–robot interaction using psychological scales for anxiety and negative attitudes toward robots. IEEE Trans robot 24(2):442–451. https://doi.org/10.1109/TRO.2007.914004
16. Wang X, Zeng H, Wang Y et al (2020) Voicecoach: interactive evidence-based training for voice modulation skills in public speaking. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. Association for Computing Machinery, New York, NY, USA, CHI '20, pp 1–12. https://doi.org/10.1145/3313831.3376726
17. Anderson PL, Zimand E, Hodges LF et al (2005) Cognitive behavioral therapy for public-speaking anxiety using virtual reality for exposure. Depress Anxiety 22(3):156–158
18. Hoque ME, Courgeon M, Martin JC et al (2013) Mach: my automated conversation coach. In Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing. Association for Computing Machinery, New York, NY, USA, UbiComp '13, pp 697–706. https://doi.org/10.1145/2493432.2493502
19. Kimani E, Bickmore T, Trinh H et al (2019) You'll be great: virtual agent-based cognitive restructuring to reduce public speaking anxiety. In 2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII), pp 641–647. https://doi.org/10.1109/ACII.2019.8925438
20. Schneider J, Börner D, van Rosmalen P et al (2015) Presentation trainer, your public speaking multimodal coach. In Proceedings of the 2015 ACM on International Conference on Multimodal Interaction. Association for Computing Machinery, New York, NY, USA, ICMI '15, pp 539–546. https://doi.org/10.1145/2818346.2830603
21. Tanveer MI, Lin E, Hoque ME (2015) 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. Association for Computing Machinery, New York, NY, USA, IUI '15, pp 286–95. https://doi.org/10.1145/2678025.2701386
22. van den Berghe R, Verhagen J, Oudgenoeg-Paz O et al (2019) Social robots for language learning: a review. Rev Educ Res 89(2):259–295. https://doi.org/10.3102/0034654318821286
23. Nomura T, Kanda T, Suzuki T et al (2020) Do people with social anxiety feel anxious about interacting with a robot? AI Soc 35(2):381–390. https://doi.org/10.1007/s00146-019-00889-9
24. Asadi R, Trinh H, Fell HJ et al (2017) Intelliprompter: speech-based dynamic note display interface for oral presentations. In Proceedings of the 19th ACM International Conference on Multimodal Interaction. Association for Computing Machinery, New York, NY, USA, ICMI '17, p 172–180. https://doi.org/10.1145/3136755.3136818
25. Forghani D, Ghafurian M, Rasouli S et al (2024) Evaluating people's perceptions of an agent as a public speaking coach. Paladyn 15(1):20240004. https://doi.org/10.1515/pjbr-2024-0004
26. Antony VN, Cho SM, Huang CM (2023) Co-designing with older adults, for older adults: robots to promote physical activity. In Proceedings of the 2023 ACM/IEEE International Conference on Human-Robot Interaction. Association for Computing Machinery, New York, NY, USA, HRI '23, 506–515. https://doi.org/10.1145/3568162.3576995
27. Foster ME, Candelaria P, Dwyer LJ et al. (2023) Co-design of a social robot for distraction in the paediatric emergency department. In Companion of the 2023 ACM/IEEE International Conference on Human-Robot Interaction. Association for Computing Machinery, New York, NY, USA, HRI '23, pp 461–465. https://doi.org/10.1145/3568294.3580127
28. Steen M (2013) Co-design as a process of joint inquiry and imagination. Des Issues 29(2):16–28. https://direct.mit.edu/desi/article-pdf/29/2/16/1715163/desi_a_00207.pdf
29. Zamenopoulos T, Alexiou K (2018) Co-design as collaborative research. University of Bristol/AHRC Connected Communities Programme
30. Judge TA, Cable DM (2004) The effect of physical height on workplace success and income: preliminary test of a theoretical model. J Retailing Appl Phychol 89(3):428–441. https://doi.org/10.1037/0021-9010.89.3.428
31. Lee Koay K, Syrdal DS, Walters ML et al (2007) Living with robots: investigating the habituation effect in participants' preferences during a longitudinal human-robot interaction study. In RO-MAN 2007 - The 16th IEEE International Symposium on Robot and Human Interactive Communication, pp 564–569. https://doi.org/10.1109/ROMAN.2007.4415149
32. Walters ML, Koay KL, Syrdal DS et al (2009) Preferences and perceptions of robot appearance and embodiment in human-robot interaction trials. https://api.semanticscholar.org/CorpusID:8049157
33. Weiss A, Wurhofer D, Bernhaupt R et al (2010) A methodological adaptation for heuristic evaluation of hri. In 19th International Symposium in Robot and Human Interactive Communication, pp 1–6). https://doi.org/10.1109/ROMAN.2010.5598735
34. Haider F, Cerrato L, Campbell N et al (2016) Presentation quality assessment using acoustic information and hand movements. In 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp 2812–2816. https://doi.org/10.1109/ICASSP.2016.7472190
35. Hincks R (2005) Measuring liveliness in presentation speech. In Ninth European Conference on Speech Communication and Technology
36. Traunmüller H, Eriksson A (1995) The perceptual evaluation of f0 excursions in speech as evidenced in liveliness estimations. J Acoustical Soc Am 97:1905–1915. https://doi.org/10.1121/1.412942
37. DeCoske MA, White SJ (2010) Public speaking revisited: delivery, structure, and style. Am J Health Syst Pharm 67(15):1225–1227
38. Berger S, Niebuhr O, Peters B (2017) Winning over an audience–a perception-based analysis of prosodic features of charismatic speech. In Proceedings 43rd Annual Conference of the German Acoustical Society, Kiel, Germany, pp 1454–1457
39. Holladay S, Coombs W (1993) Communicating visions: an exploration of the role of delivery in the creation of leader charisma. Manag Commun q - MANAG COMMUN Q 6:405–427. https://doi.org/10.1177/0893318993006004003
40. Rosenberg A, Hirschberg J (2005) Acoustic/Prosodic and lexical correlates of charismatic speech. Interspeech, https://api.semanticscholar.org/CorpusID:735627
41. Niebuhr O, Brem A, Novák-Tót E et al (2016) Charisma in business speeches: a contrastive acoustic-prosodic analysis of steve jobs and Mark Zuckerberg. In Speech Prosody Special Interest Group, 8th Speech Prosody Conference; Conference date: 31-05-2016 Through 03-06-2016

42. Niebuhr O, Voße J, Brem A (2016) What makes a charismatic speaker? A computer-based acoustic-prosodic analysis of steve jobs tone of voice. Comput in Hum Behav 64:366–382. https://doi.org/10.1016/j.chb.2016.06.059; https://www.sciencedirect.com/science/article/pii/S0747563216304873
43. Gallo C (2009) Uncovering Steve jobs' presentation secrets. https://www.bloomberg.com/news/articles/2009-10-06/uncovering-steve-jobs-presentation-secrets,. Accessed 8 Mar 2024
44. Hotz RL (2014) How to train your voice to be more charismatic - wsj. https://www.wsj.com/articles/how-to-train-your-voice-to-be-more-charismatic-1417472214. Accessed 8 Mar 2024
45. Sutter JD (2011) When it comes to presentation, mark zuckerberg is no steve jobs. https://www.cnn.com/2011/TECH/innovation/07/07/zuckerberg.facebook.presentation/index.html
46. Saukh O, Maag B (2019) Quantle: fair and honest presentation coach in your pocket. In 2019 18th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN), pp 253–264
47. Trinh H, Ring L, Bickmore T (2015) Dynamicduo: co-presenting with virtual agents. In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. Association for Computing Machinery, New York, NY, USA, CHI '15, pp 1739–748. https://doi.org/10.1145/2702123.2702584
48. Gordon G, Spaulding S, Westlund JK et al (2016) Affective personalization of a social robot tutor for children's second language skills. In Proceedings of the AAAI conference on artificial intelligence
49. Pachidis T, Vrochidou E, Kaburlasos V et al (2019) Social robotics in education: state-of-the-art and directions. In Advances in Service and Industrial Robotics: Proceedings of the 27th International Conference on Robotics in Alpe-Adria Danube Region (RAAD 2018), Springer, pp 689–700
50. Smakman MHJ, Konijn EA, Vogt P et al (2021) Attitudes towards social robots in education: enthusiast, practical, troubled, sceptic, and mindfully positive. Robotics 10(1). https://doi.org/10.3390/robotics10010024, https://www.mdpi.com/2218-6581/10/1/24
51. Vogt P, van den Berghe R, de Haas M et al (2019) Second language tutoring using social robots: a large-scale study. In 2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI), pp 497–505. https://doi.org/10.1109/HRI.2019.8673077
52. Belpaeme T, Kennedy J, Ramachandran A et al (2018) Social robots for education: a review. Sci robot 3(21):eaat5954
53. Pai RY, Shetty A, Dinesh TK et al (2024) Effectiveness of social robots as a tutoring and learning companion: a bibliometric analysis. Cogent Bus & Manag 11(1):2299075
54. Fasola J, Mataric MJ (2012) Using socially assistive human–robot interaction to motivate physical exercise for older adults. Proc IEEE 100(8):2512–2526. https://doi.org/10.1109/JPROC.2012.2200539
55. Saerbeck M, Schut T, Bartneck C et al (2010) Expressive robots in education varying the degree of social supportive behavior of a robotic tutor. pp 1613–1622. https://doi.org/10.1145/1753326.1753567
56. Lemaignan S, Newbutt N, Rice L et al (2024) "It's important to think of pepper as a teaching aid or resource external to the classroom": a social robot in a school for autistic children. Int J Soc robot 16(6):1083–1104. https://doi.org/10.1007/s12369-022-00928-4
57. Kim ES, Leyzberg D, Tsui KM et al (2009) How people talk when teaching a robot. In Proceedings of the 4th ACM/IEEE International Conference on Human Robot Interaction. Association for Computing Machinery, New York, NY, USA, HRI '09, pp 23–30. https://doi.org/10.1145/1514095.1514102
58. Lee KM, Jung Y, Kim J et al (2006) Are physically embodied social agents better than disembodied social agents?: the effects of physical embodiment, tactile interaction, and people's loneliness in human–robot interaction. Int J Hum Comput Stud 64(10):962–973. https://doi.org/10.1016/j.ijhcs.2006.05.002; https://www.sciencedirect.com/science/article/pii/S1071581906000784
59. Pan Y, Steed A (2016) A comparison of avatar-, video-, and robot-mediated interaction on users' trust in expertise. Front Robot AI 3. https://www.frontiersin.org/articles/10.3389/frobt.2016.00012
60. Pereira A, Martinho C, Leite I et al (2008) Icat, the chess player: the influence of embodiment in the enjoyment of a game. pp 1253–1256. https://doi.org/10.1145/1402821.1402844
61. Powers A, Kiesler S, Fussell S et al (2007) Comparing a computer agent with a humanoid robot. In Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction. Association for Computing Machinery, New York, NY, USA, HRI '07, pp 145–152. https://doi.org/10.1145/1228716.1228736
62. Janeczko Z, Foster ME (2022) A study on human interactions with robots based on their appearance and behaviour. In Proceedings of the 4th Conference on Conversational User Interfaces. Association for Computing Machinery, New York, NY, USA, CUI '22. https://doi.org/10.1145/3543829.3544523
63. Thunberg S, Arnelid M, Ziemke T (2022) Older adults' perception of the furhat robot. In Proceedings of the 10th International Conference on Human-Agent Interaction. Association for Computing Machinery, New York, NY, USA, HAI '22, pp 4–12. https://doi.org/10.1145/3527188.3561924
64. Aron A, Melinat E, Aron EN et al (1997) The experimental generation of interpersonal closeness: a procedure and some preliminary findings. Pers Soc Psychol Bull 23(4):363–377. https://doi.org/10.1177/0146167297234003
65. Laban G, George JN, Morrison V et al (2021) Tell me more! assessing interactions with social robots from speech. Paladyn, J Behavioral Robot 12(1):136–159. https://doi.org/10.1515/pjbr-2021-0011
66. Kalegina A, Schroeder G, Allchin A et al (2018) Characterizing the design space of rendered robot faces. In Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction. Association for Computing Machinery, New York, NY, USA, HRI '18, pp 96–104. https://doi.org/10.1145/3171221.3171286
67. Kontogiorgos D, Pereira A, Andersson O et al (2019) The effects of anthropomorphism and non-verbal social behaviour in virtual assistants. In Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents. Association for Computing Machinery, New York, NY, USA, IVA '19, pp 133–140. https://doi.org/10.1145/3308532.3329466
68. Boersma P, Weenink D (2021) Praat: doing phonetics by computer [computer program]. http://www.praat.org/. Version 6.1.38, Retrieved January 2, 2021
69. Jadoul Y, Thompson B, de Boer B (2018) Introducing Parselmouth: a python interface to Praat. J phon 71:1–15. https://doi.org/10.1016/j.wocn.2018.07.001
70. Oertel C, Lopes J, Yu Y et al (2016) 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. Association for Computing Machinery, New York, NY, USA, ICMI '16, pp 21–28. https://doi.org/10.1145/2993148.2993188
71. Jacewicz E, Fox RA, O'Neill C et al (2009) Articulation rate across dialect, age, and gender. Lang Var Change 21(2):233–256
72. Quené H (2008) Multilevel modeling of between-speaker and within-speaker variation in spontaneous speech tempo. J Acoustical Soc Am 123(2):1104–1113
73. de Jong Nh, Wempe T (2009) Praat script to detect syllable nuclei and measure speech rate automatically. Behav Res Methods 41(2):385–390. https://doi.org/10.3758/BRM.41.2.385

74. Feinberg DR (2022) Parselmouth praat scripts in python. https://osf.io/6dwr3/
75. Boersma P, Van Heuven V (2001) Speak and unspeak with praat. Glot Int 5(9/10):341–347
76. Batrinca L, Stratou G, Shapiro A et al (2013) Cicero - towards a multimodal virtual audience platform for public speaking training. In: Aylett R, Krenn B, Pelachaud C et al (eds) Intelligent virtual agents. Springer, Berlin, Heidelberg, pp 116–128
77. Marchand E, Uchiyama H, Spindler F (2016) Pose estimation for augmented reality: a hands-on survey. IEEE Trans Visual Comput Graphics 22(12):2633–2651. https://doi.org/10.1109/TVCG.2015.2513408
78. Nomura T, Kanda T, Suzuki T et al (2004) Psychology in human-robot communication: an attempt through investigation of negative attitudes and anxiety toward robots. In RO-MAN 2004. 13th IEEE International Workshop on Robot and Human Interactive Communication (IEEE Catalog No.04TH8759), pp 35–40. https://doi.org/10.1109/ROMAN.2004.1374726
79. Prochazka J, Ovcari M, Durinik M (2020) Sandwich feedback: the empirical evidence of its effectiveness. Learn Motiv 71:101649. https://doi.org/10.1016/j.lmot.2020.101649, https://www.sciencedirect.com/science/article/pii/S0023969020301429
80. Bottini S, Gillis J (2021) A comparison of the feedback sandwich, constructive-positive feedback, and within session feedback for training preference assessment implementation. J Organ Behav manage 41(1):83–93. https://doi.org/10.1080/01608061.2020.1862019
81. Debnath K (2021) Online clinical teaching: a simple model to facilitate students' communication and clinical reasoning skills on distance learning e-platform (2016). MedEdpublish 9:272
82. Dohrenwend A (2002) Serving up the feedback sandwich. Fam Pract Manag 9(10):43–46
83. Lin GSS, Tan WW, Hashim H et al (2023) The use of feedback in teaching undergraduate dental students: feedback sandwich or ask-tell-ask model? BMC Oral Health 23(1):417
84. Freepik (2010). https://www.freepik.com/. Accessed 2024-03-08
85. Fischer K, Lohan K, Saunders J et al (2013) The impact of the contingency of robot feedback on hri. In 2013 International Conference on Collaboration Technologies and Systems (CTS), pp 210–217. https://doi.org/10.1109/CTS.2013.6567231
86. Palinko O, Sciutti A, Schillingmann L et al (2015) Gaze contingency in turn-taking for human robot interaction: advantages and drawbacks. In 2015 24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN), pp 369–374. https://doi.org/10.1109/ROMAN.2015.7333640
87. Sandini G, Sciutti A, Rea F (2019) Movement-based communication for humanoid-human interaction. Springer, Netherlands, Dordrecht, pp 2169–2197. https://doi.org/10.1007/978-94-007-6046-2_138
88. Tudor AD, Mustatea I, Poeschl S et al (2014) Responsive audiences — nonverbal cues as reactions to a speaker's behavior. In 2014 IEEE Virtual Reality (VR), pp 119–120. https://doi.org/10.1109/VR.2014.6802080
89. McCroskey JC (1997) An introduction to rhetorical communication. Routledge. https://books.google.ca/books?id=WQuCVToxdpwC
90. McCroskey JC, Beatty MJ, Kearney P et al (1985) The content validity of the prca-24 as a measure of communication apprehension across communication contexts. Commun Q 33(3):165–173
91. Gosling SD, Rentfrow PJ, Swann WB (2003) A very brief measure of the big-five personality domains. J Res personality 37(6):504–528. https://doi.org/10.1016/S0092-6566(03)00046-1; https://www.sciencedirect.com/science/article/pii/S0092656603000461
92. Heerink M, Kröse B, Evers V et al (2010) Assessing acceptance of assistive social agent technology by older adults: the almere model. Int J Soc robot 2(4):361–375. https://doi.org/10.1007/s12369-010-0068-5
93. Aron A, Aron EN, Smollan D (1992) Inclusion of other in the self scale and the structure of interpersonal closeness. J Pers Soc psychol 63(4):596
94. Haslam N, Loughnan S, Kashima Y et al (2008) Attributing and denying humanness to others. Eur Rev Soc psychol 19(1):55–85. https://doi.org/10.1080/10463280801981645
95. Funke F, Reips UD (2012) Why semantic differentials in web-based research should be made from visual analogue scales and not from 5-point scales. Field method 24(3):310–327. https://doi.org/10.1177/1525822X12444061
96. Matejka J, Glueck M, Grossman T et al (2016) 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. Association for Computing Machinery, vol 16, New York, NY, USA, CHI, 5421–5432. https://doi.org/10.1145/2858036.2858063
97. Bousmalis K, Mehu M, Pantic M (2013) Towards the automatic detection of spontaneous agreement and disagreement based on nonverbal behaviour: a survey of related cues, databases, and tools. Image Vision Comput 31(2):203–221. https://doi.org/10.1016/j.imavis.2012.07.003, https://www.sciencedirect.com/science/article/pii/S0262885612001059 affect Analysis In Continuous Input
98. Dennis A, Kinney S (1998) Testing media richness theory in the new media: the effects of cues, feedback, and task equivocality. Inf Syst Res 9:256–274. https://doi.org/10.1287/isre.9.3.256
99. Gratch J, Wang N, Okhmatovskaia A et al (2007) Can virtual humans be more engaging than real ones? In: Jacko J (ed) Human-computer interaction. HCI intelligent multimodal interaction environments. Springer, Berlin, Heidelberg, pp 286–297
100. Sidner CL, Lee C, Kidd CD et al (2005) Explorations in engagement for humans and robots. Artif Intell 166(1):140–164. https://www.sciencedirect.com/science/article/pii/S0004370205000512
101. Gladstein GA (1977) Empathy and counseling outcome: an empirical and conceptual review. The Couns Psychol 6(4):70–79. https://doi.org/10.1177/001100007700600427
102. Bozdogan H (1987) Model selection and akaike's information criterion (aic): the general theory and its analytical extensions. Psychometrika 52(3):345–370. https://doi.org/10.1007/BF02294361
103. Duan C, Hill CE (1996) The current state of empathy research. J Couns psychol 43(3):261
104. Yoshihara HKA (2015) Perceiving active listening activates the reward system and improves the impression of relevant experiences. Soc Neurosci 10(1):16–26. pMID: 25188354, https://doi.org/10.1080/17470919.2014.954732
105. Złotowski J, Strasser E, Bartneck C (2014) Dimensions of anthropomorphism: from humanness to humanlikeness. In Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction, pp 66–73
106. Park S, Whang M (2022) Empathy in human–robot interaction: designing for social robots. Int J Environ Res Pub Health and Public Health 19(3):1889
107. Eyssel F, Hegel F, Horstmann G et al (2010) Anthropomorphic inferences from emotional nonverbal cues: a case study. In 19th International Symposium in Robot and Human Interactive Communication, pp 646–651. https://doi.org/10.1109/ROMAN.2010.5598687
108. Roach P (1982) On the distinction between 'stress-timed'and 'syllable-timed'languages. Ling Controversies 73:79
109. Holmes J (1986) Functions of you know in women's and men's speech. Lang Soc 15(1):1–21. https://doi.org/10.1017/S0047404500011623

110. McMurtry CM, McGrath PJ, Asp E et al (2007) Parental reassurance and pediatric procedural pain: a linguistic description. J Sport Hist of Pain 8(2):95–101. https://doi.org/10.1016/j.jpain.2006.05.015; https://www.sciencedirect.com/science/article/pii/S1526590006008595
111. Warren P (2016) Uptalk: the phenomenon of rising intonation. Cambridge University Press
112. Borrie SA, Lubold N, Pon-Barry H (2015) Disordered speech disrupts conversational entrainment: a study of acoustic-prosodic entrainment and communicative success in populations with communication challenges. Front psychol 6:1187

Delara Forghani is currently doing PhD in Human-Computer Interaction at the Cheriton School of Computer Science at the University of Waterloo. She got her master's from the Department of Electrical and Computer Engineering at the University of Waterloo in 2023 focusing on Human-Robot Interaction and Social Robotics. Her research interests expands interaction design, exploring new interaction methods and input techniques.

Samira Rasouli is a Postdoctoral Fellow in the Electrical and Computer Engineering Department at the University of Waterloo, Canada. She received her PhD in Electrical and Computer Engineering, specializing in Pattern Analysis and Machine Intelligence (PAMI), from the University of Waterloo in 2023. Her research interests include human–computer/robot interaction, social robotics, artificial intelligence, and assistive technologies.

Moojan Ghafurian is an Assistant Professor at the Department of Systems Design Engineering at the University of Waterloo and Director of the Emotionally Intelligent and Trustworthy Agents (EITA) Research Laboratory. She got her PhD from the Pennsylvania State University in 2017, and her research interests include human-computer/robot interaction, social robotics, affective computing, and cognitive science.

Melanie Jouaiti is an Assistant Professor in the School of Computer Science at the University of Birmingham. She got her PhD from the Université de Lorraine in 2020, and her research interests include socially assistive robotics, and healthcare technology.

Chrystopher L. Nehaniv Ph.D. (University of California, Berkeley) is a Mathematician, a Computer Scientist, a Complex Systems expert, and a full Professor with the Department of Systems Design Engineering and the Department of Electrical and Computer Engineering, University of Waterloo, where he has co-founded the Social and Intelligent Robotics Research Laboratory and directs Waterloo Algebraic Intelligence and Computation Laboratory. His research interests include automata and information theory, AI robotics, and constructive biology areas, such as cognition, experience, and interaction.

Kerstin Dautenhahn Dr. rer. nat., is a full professor and Canada 150 Research Chair in Intelligent Robotics (Laureate) at the University of Waterloo in Canada. She is an IEEE Fellow for her contributions to Social Robotics and Human-Robot Interaction, and a Fellow of the Royal Society of Canada. Her work is widely cited, and she frequently gives keynote talks at international conferences. Dr. Dautenhahn's research focuses on basic research as well as assistive technology applications.