2023 32nd IEEE International Conference on Robot and
Human Interactive Communication (RO-MAN)
August 28-31, 2023. Paradise Hotel, Busan, Korea

2023 32nd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN) | 979-8-3503-3670-2/23/$31.00 ©2023 IEEE | DOI: 10.1109/RO-MAN57019.2023.10309583

What Do People Think of Social Robots and Voice Agents as Public Speaking Coaches?

Delara Forghani[1*], Moojan Ghafurian[1], Samira Rasouli[1], Chrystopher L. Nehaniv[1,2] and Kerstin Dautenhahn[1]

***Abstract*— Social robots have the potential to serve as coaches for public speaking training. To design successful social robots, it is important to understand the expectations and perceptions of prospective users of such robots. In this paper, we present thematic analyses of comments made by 168 participants in an online study where participants watched videos of agents in the role of a public speaking coach. The study had a between-participant design with three conditions: two conditions with a humanoid social robot in either (1) active listening mode, i.e., using non-verbal backchanneling, or (2) passive listening mode, and (3) a voice assistant agent. The themes identified and discussed can contribute to the development of social robots and other agents as public speaking coaches.**

I. Introduction

Public speaking is a valuable skill that is crucial for achieving success in both academic and professional settings [1]–[3]. Rehearsing is essential to delivering an effective presentation [4]. Although public speaking coaches are helpful for improving one's public speaking skills, practicing with a human coach can be difficult due to cost, limited access, time constraints, and public speaking anxiety [4], [5]. As a result, designing interactive technologies like robots or voice assistants as coaches can serve as a supplementary aid to the currently available resources and provide assistance to individuals who may not obtain access to or do not feel at ease with a human coach.

Studies have used interactive systems to assess their efficacy in providing verbal and nonverbal feedback on the communication skills of presenters [4], [6]–[9]. However, there is yet limited knowledge regarding how people would perceive and evaluate different interactive agents with varied levels of social embodiment, anthropomorphism, and non-verbal social behaviour as potential coaches for public speaking. Two examples of these interactive agents are social robots and smart speaker-based voice assistants.

Research has compared the impact of a social robot and a voice assistant on users' entertainment levels by examining them as game peers [10]. Moore and Urakami (2022) used three voice user interfaces (VUIs), a voice-only system, a smart speaker, and a social robot, to compare their distracting effects on users' cognitive functionality, and defined the benefits and drawbacks of each of the agents, for example, the voice-only VUI had the minimum distraction [11]. The physically embodied VUI (the smart speaker) scored the highest on perceived calm, which may be due to its minimalist design with only essential features, and the ability to associate the voice with a specific source, making the user feel most comfortable [11]. The social embodied VUI (the social robot) was the most likeable VUI [11].

In addition to analyzing the type of agent used, our investigation explores the impact of active versus passive listening in social robots. Specifically, we studied the impact of nonverbal backchanneling on participants' perceptions of the robot as a coach. Backchannel cues refer to verbal and nonverbal signals that listeners use to convey their understanding, attention, agreement, and interest in continuing the conversation (e.g. nodding or "hmm" utterances) [12]. Bodie et al. found that in a human-human self-disclosure session, behaviours associated with actively listening, verbal and nonverbal, were perceived as indicating a higher level of emotional awareness and led to a greater degree of emotional improvement [13].

To understand what would be preferred in terms of the agent type and social (backchanneling) behaviours for an agent as a public speaking coach, this online study explored the potential of a social robot (with and without backchanneling behaviour shown through head nodding) and a smart speaker (a voice assistant agent), as public speaking coaches.

For designing any type of technology, it is important to understand users' opinions, needs, and preferences. To the best of our knowledge, there is not enough information available on thematic and qualitative analyses of participants' perceptions regarding the use of different types of interactive agents in the context of public speaking training. Therefore, we report on thematic analyses of participants' answers regarding the potential application of three interactive agents as public speaking coaches. Our **Research Question** is: *What are participants' impressions, opinions and preferences regarding three types of agents in the role of a public presentation coach?*

II. Related work

Agent Comparison: Voice as a social cue can change how people perceive machines [14]. However, non-verbal behaviour can be equally important in human-human [13] and human-robot interaction [12], [15], e.g. for social coordination and signaling [14]. Social robots are designed to interact with humans, and thus, non-verbal cues can be used to improve their ability to convey their internal state, complementing speech [16]. Multiple studies have compared intelligent agents with different embodiments that

This research 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.

[1]Department of Electrical and Computer Engineering, [2]Department of Systems Design Engineering, University of Waterloo, Waterloo, Ontario, Canada

*delara.forghani@uwaterloo.ca

either communicated solely through voice or expressed non-verbal social behaviour during interactions, to gain insights into how user experience and task performance would be affected [4], [16]–[18]. For example Kontogiorgos et al [16] compared a smart speaker, a robotic head resembling a human face that only interacted through speech and the same robotic head that also communicated through gaze and head movements(social) to analyze how people respond to various agents in a guided cooking task. The results suggested that anthropomorphism and the use of social non-verbal behaviour in the agent positively impacted its perceived sociability but negatively impacted task completion time [16]. No significant difference in agents' perceived intelligence was reported [16]. Participants' responses to two open-ended questions yielded two key themes. First, participants recognized the distinction in the embodiments of the agents, however, they did not perceive any disparities between the anthropomorphic robot and the anthropomorphic social robot. Second, most of the participants preferred the anthropomorphic social robot for the cooking task [16].

In the context of a robotic coach, Trinh et al. (2017) [4] conducted a study using an anthropomorphic robotic head as a coach in a within-participants design. The robot provided verbal feedback on participants' speech, eye contact, and content coverage. Participants were reported to give better presentations when coached by the robot, compared to those who received visual-only or voice-only feedback [4]. The voice-only feedback in this study was conducted without identifying a device as the source of the voice [4]. Trinh et al. [4] did semi-structured interviews with participants. The responses were transcribed and analyzed, leading to the identification of three main themes. These themes included the facilitation of rehearsal sessions, improvement in the quality of presentations, and a varied impact on participants' confidence [4]. Wang et al. [7] focused on utilizing Amazon Alexa as a tutor for reducing public speaking anxiety. A qualitative analysis of responses to three open-ended questions resulted in three main themes: usefulness, weaknesses, and suggestions for improving Alexa's effectiveness as a tutor. Kimani et al. [19] conducted interviews and performed thematic analysis to investigate participants' perceptions following their interaction with a virtual agent functioning as a public speaking tutor. Several themes emerged from the analysis, including the agent's ability to connect with presenters and sense their fears, and participants' intentions to use the agent for future presentations. However, in the realm of public speaking training, limited research has explored qualitative analysis using different interactive agents.

Backchanneling: Typically, backchannels refer to the verbal and/or non-verbal cues that listeners use to indicate their comprehension, attentiveness, agreement, or willingness to sustain a coordinated conversation with a speaker [20], [21]. The use of backchanneling can enhance user engagement with virtual conversational agents and social robots. Gratch et al. (2007) suggested that a virtual listener agent that used positive listening feedback, such as head nods, could create greater engagement and establish a stronger rapport than a human listener [15]. Head nods, as a form of nonverbal backchannel, communicate attentiveness and agreement [22], [23]. Larger and more frequent nods typically indicate agreement, while smaller nods suggest active participation in the conversation [24]. Murali et al. (2021) utilized a virtual agent to provide support for presenters through nonverbal backchannel cues, including smiling, head nods, raised eyebrows, and a thumbs-up gesture. Through self-reports and physiological measures, such as heart rate (HR) and heart rate variability (HRV), they found significant reductions in public speaking anxiety among presenters who received this type of support, compared to a control condition that involved a text display and no agent. The study involved semi-structured interviews, which were transcribed and coded using thematic analysis. The resulting themes included the advantages of the agent in enhancing confidence, providing moral support, displaying interest, and potential problems associated with its use [25].

III. Method

An online, between-participants study with three conditions was carried out using Amazon Mechanical Turk (MTurk) to understand participants' perspectives on two types of interactive agents (i.e., voice assistant agent and a social robot) as public speaking coaches, including social robots with and without backchanneling behaviour (as a proxy for active listening). Therefore, the study had three conditions: (a) **ActiveListenR**: A social robot (Pepper robot [26]) provided verbal feedback. It also showed backchanneling through nodding during the presentation; (b) **PassiveListenR**: A social robot (Pepper robot) provided verbal feedback; and (c) **VoiceAgent**: A voice agent (Google Nest Mini [27]) provided verbal feedback.

A. Procedures and Measures

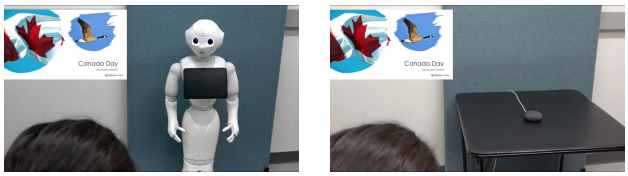
First, participants provided their consent to participate by reviewing the study information.

Step 1) Demographics and pre-experimental questionnaires: Participants were asked general questions, about their demographics, experiences with public speaking classes, and personalities.[1]

Step 2) Watching a Video of an agent interacting with a student: Participants watched a video where a student delivered a presentation on the topic of "History of Canada Day"[2] to an agent (the type of agent changed based on the experiment condition). The presentation took approximately one minute and thirty seconds. Following that, the agent provided feedback to the student for approximately two minutes. The verbal feedback included expressing appreciation for the presentation and was primarily about the vocal modulation skills and prosodies in speech. The way we structured the feedback was to acknowledge positive aspects (e.g. 'You demonstrated a good speech rate in your presentation!") and

[1]These questions are not discussed in detail due to page limits and will be reported elsewhere.

[2]Note: the presentation content was sourced from the official website of the Government of Canada (canada.ca).



(a) (b)

Fig. 1. (a) Picture of the experimental setup with the social robot (Pepper). (b) Picture of the experimental setup with the voice agent (Google Nest Mini).

then offer suggestions for improvement using the "feedback sandwich" technique. (e.g. "Try not to use pauses too often or for too long!") [4].

Participants could either replay the video or proceed to the questionnaires. To prevent participants from engaging in other tasks while the video was playing, the video would pause if the current window or tab was not the active window.

For consistency purposes, we used the Pepper robot's voice in all conditions. Moreover, in all the experimental conditions, the audio and content of the student's presentation and the coach agent's feedback were identical. They were recorded once and played in all of the videos. The presentation slides appeared on the top-left corner of the screen and would disappear after the student's presentation (see Figure 1. The camera used to record the videos was positioned in the same location, and behind the student's head so the student cannot be identified). Depending on the experimental condition, the agent's non-verbal behaviour varied as below:

Condition 1 - ActiveListenR: The Pepper robot displayed non-verbal backchanneling through head nodding during the presentation. To achieve this a researcher, using Wizard of Oz, had control over the timing of the robot's head nods. The researcher monitored the student's speaking voice and utilized cues such as the end of utterances and brief pauses to initiate the robot's head nods [15]. As head nods can convey varied meanings based on their amplitude and frequency [23], and to avoid participants' attention becoming biased towards one specific head motion in the robot, we employed two modes of nodding randomly alternating between them. One consisted of a gradual upward and downward motion of the head, and the other involved two consecutive upward and downward motions that were faster in speed and smaller in amplitude. During the feedback, along with the pre-recorded audio, the robot used hand gestures and arm movements (to indirectly demonstrate the differences between voice agents and robots). The gestures and movements of the Pepper robot were created with *Choregraphe* [28], and their timing was tailored to match the accompanying verbal content.

Condition 2 - PassiveListenR: The robot remained stationary with a fixed head position and did not use backchanneling during the presentation. The robot's location and orientation were consistent with Condition 1. The feedback content was identical to Condition 1.

Condition 3 - VoiceAgent: A Google Nest Mini device was positioned on a table in front of the student, in a similar location that the robot was located in the other two conditions. During the presentation, the Google Nest device was in listening mode (a behaviour similar to when Google Nest is getting input in a real setting, with the four lights illuminated, this was added especially for participants who are familiar with this device) but remained silent. The feedback content was identical to the previous conditions.

Step 3) Post-experimental questionnaires: Aside from questionnaires that involved participants' ratings of different aspects of agents which are not reported here[3], participants were asked to provide feedback on the use of robots and voice assistant agents as public speaking coaches through two open-ended questions. In the first question, in the ActiveListenR and PassiveListenR conditions, participants were asked to provide their opinions on the idea of using robots as public speaking coaches. In contrast, those in the VoiceAgent condition were asked to provide their opinions on the idea of using voice assistant agents as public speaking coaches. Participants were asked to rate the statement "Voice assistant agents can be good coaches" (or "Robots can be good coaches" depending on the condition). The answer was on a continuous scale ranging from "Strongly disagree" to "Strongly agree". Participants were asked to provide explanations for their ratings.

Participants were also asked about preferences for a coaching agent given four options, i.e., (1) the same robot/voice assistant agent they had seen in the video, (2) a different robot/voice assistant agent from the one they had seen, (3) a different type of interactive system (for instance, voice assistant agent for those in robot conditions and vice versa), or (4) a human coach. In an open-ended follow-up question, we asked participants why they preferred each of the agents or a human. Here, we present a thematic analysis of participants' comments regarding their preferences.

B. Participants

We recruited 184 participants via Amazon Mechanical Turk, considering a minimum approval rate of 80%, at least 50 HITs for qualifications, and participants were in US or Canada. 130 participants were from Canada and 54 from the United States. A total of 16 records were discarded from the study due to failing more than two attention checks. The gender distribution was 66 females, 100 males, and 2 who preferred not to disclose their gender. The age range of participants was 19 to 67 years old, with a mean age of 35.60. We had a final total of 58 participants in the ActiveListenR condition, 56 in the PassiveListenR condition, and 54 in the VoiceAgent condition. The study received ethics clearance from the University of Waterloo Human Research Ethics Board.

[3]We asked participants questionnaires regarding their attitude towards agents and compared their ratings of different agents using designed and standard questionnaires (e.g., RoSAS [29] and parts of Godspeed [30]), which are beyond the scope of this paper.

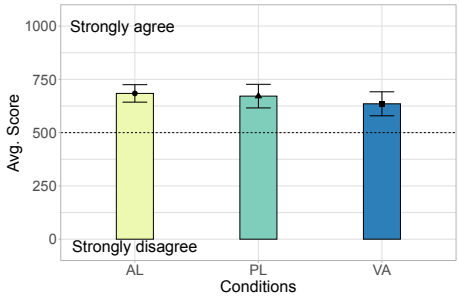


Fig. 2. Ratings on whether participants agreed with using the agent they saw as a public speaking coach (rated from "Strongly disagree" to "Strongly agree"). Mean values and error bars (95% confidence intervals) are shown. Dashed line indicates the neutral choice. Recorded responses changed from 0 to 1000. AL: ActiveListenR, PL: PassiveListenR, VA: VoiceAgent.

C. Thematic Analysis

Participants' comments were analyzed using the six-step Thematic Analysis (TA) method developed by Braun and Clarke [31]. Codes (labels assigned to pieces of text) were first generated from extracted excerpts, and then for one of the analyses (i.e., the first analysis where we had more comments), a theme was accepted if we found at least 20 codes related to it across all conditions. Our second analysis looked at the salient reasons for selecting a choice in a multiple-choice question, and we used 4 as the minimum threshold to accept a theme since some options (e.g., a different agent) were selected less than others. Themes were then reviewed and revised, and sub-themes were created mainly based on positive vs. negative attitudes in each theme. One author conducted the TA of the entire dataset and another author independently analyzed 20% of the data. We calculated the inter-rater reliability using kappa's agreement score [32]. The results of calculating weighted kappa was 0.873 ($kappa = 0.823, SE = 0.054, 95\%, CI = [0.717, 0.929]$), which shows "almost perfect agreement".

IV. Results

Figure 2 shows participants' ratings of how they felt about using agents (the one they saw) as a public speaking coach. The average ratings are relatively high (around 700 out of 1000) for all experimental conditions, suggesting that participants were generally supportive of the idea of a social robot or a voice assistant agent as a public speaking coach.

A total of six main themes were identified. For each, we identified 3 to 5 sub-themes to better represent the comments. The themes and sub-themes resulting from the thematic analysis of participants' comments are presented in Table I, and discussed in the following subsections. As the quotes represent themes (which are consistent in all experimental conditions), they are added as examples, they can be from any of the three experimental conditions (VA: VoiceAgent, AL: ActiveListenR, PL: PassiveListenR).

A. Theme 1 - Technology Acceptance

125 participants stated their views about technology acceptance. This theme describes perceptions of the technology and the extent to which they accept it and are willing to use it in the future. It is categorized into four sub-themes.

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

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

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

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

B. Theme 2 - Advantages/Shortcomings of Technology

30 comments were regarding this theme. Participants discussed the capabilities or limitations of this technology and how it may or may not be able to address certain issues.

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

2.2 Does not help with fear/anxiety of public speaking: 4 participants stated that the agent does not help with the fear of public speaking in front of a live audience. "*They wouldn't help too much with the fear of speaking before a live audience (PL).*" "*I think it can be a helpful starting point to get you used to the cadence of public speaking, but I don't think it will actually help you get over the anxiety of it (AL).*"

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

TABLE I

THEMATIC ANALYSIS RESULTS ON OPINIONS ABOUT THE AGENTS. THE PERCENTAGE OF RESPONSES RELATED TO EACH SUB-THEME IS PROVIDED FOR EACH CONDITION. TOTAL PERCENTAGE OF PARTICIPANTS COMMENTING ON EACH THEME IS SHOWN BELOW THE DASHED LINES. WHILE ONE PARTICIPANT COMMENT MIGHT BE COUNTED IN MORE THAN ONE SUB-THEME, DEPENDING ON THE CONTENT. AS A RESULT OF THIS OVERLAP, SUB-THEMES PERCENTAGES MAY NOT ADD UP TO THEIR RESPECTIVE TOTAL PERCENTAGES.

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

people who are struggling with the general basics of public speaking (AL)."

C. Theme 3 - Attitude towards Agent

43 participants expressed their attitudes about the agent they had seen in the video.

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

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

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

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

D. Theme 4 - Interaction Perceptions

21 participants shared their views about the interaction they saw in the video, which relates to how participants envisioned and perceived the interaction in general.

4.1 Interaction is odd: 4 comments in the ActiveListenR and the PassiveListenR conditions were related to interactions

with the social robot and speaking with it felt strange. "*It feels somehow strange but I believe it would be helpful to consider it (AL).*" "*However, it still feels strange having the interaction with a robot instead of a human. I think I would become more comfortable with them over time (PL).*"

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

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

E. Theme 5 - Agent vs. Human

29 comments were related to comparisons between the agent they observed as a coach with a human coach.

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

5.2 A human coach should be next to the agent: 3 participants indicated having a human coach beside the agent could be beneficial. "*I think though that it would be harder to only work with a robot, I think both a robot and human coach would work best together (PL).*" "*It might be better, in some instances, for students to have one other human person in the room to help with additional feedback (AL).*"

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

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

F. Theme 6 - Feedback

46 comments reflected participants' opinions about the feedback content and the agent's voice (identical in all conditions).

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

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

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

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

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

Finally, participants were asked whether they preferred the same agent they interacted with, another agent, or a human coach. Table II displays the results of the thematic analysis regarding the reasons cited by participants for their choice, as well as the number of participants who preferred each of the choices for a public speaking coach.

V. Discussion

The results showed high agreement scores for using social robots and voice assistant agents as public speaking coaches. Participants' responses to open-ended questions provided valuable insights into the strengths and weaknesses of these systems in the context of a public speaking coach. Among the advantages mentioned by participants, regardless of the agents' embodiment, the most prominent were the *helpfulness* of the feedback and the *impartiality and honesty* of the feedback. *Ease of interaction* with these agents was also mentioned as useful for individuals with public speaking anxiety, as it might allow them to practice before speaking in front of an audience. Lastly, participants highlighted the *usefulness* of having these agents for people who want to improve their fundamental public speaking skills, especially university and college students. Participants also pointed out

TABLE II

PARTICIPANTS' AGENT PREFERENCES IN EACH CONDITION. NUMBERS ARE DIVIDED BY THEMES. THE "TOTAL NUMBER (PERCENTAGE)" OF PARTICIPANTS PREFERRING THE SAME AGENT, A DIFFERENT AGENT, ANOTHER INTERACTIVE SYSTEM, OR A HUMAN ARE SHOWN FOR EACH CONDITION.

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

several shortcomings of those agents. The primary limitation identified by participants was the agent's presumed *inability to perceive and analyze non-verbal communication*, prosody, body language, and emotional states of the presenter during the presentation. Similarly, some participants felt that the feedback provided by the agents was too generic and lacked detail. Another disadvantage/limitation was the *robotic-sounding voice* of the agent, which some participants found uncomfortable and made the feedback difficult to understand. Note, in our study, we used the robot's voice in all three conditions. We tried to adjust the voice pitch and frequency to be as understandable as possible, but a few participants in the voice assistant agent condition expressed a preference for the voice of Google Nest instead. The goal of this paper was to understand the comments of participants about different types of coaching agents, and future work would be needed to further compare preferences for these agents. Note, the findings of this online study may vary from live interactions, especially in detecting head nodding. In future research, we aim to replicate the study in-person. Participants' responses regarding their preference for a coach suggested that most participants *preferred a human coach*. This was mainly due to the comprehensive and compassionate feedback that a human can provide, as commented by the participants, as well as a human's experience and knowledge, making a human more trustworthy.

While we cannot make strong claims based on the Thematic Analyses, we noticed some differences between the number of comments in different conditions. For example, comments related to the sub-theme on *agent being unresponsive/weird* were only made by participants in the PassiveListenR condition. One possible explanation is that the non-verbal backchanneling behaviour of the ActiveListenR positively influenced this aspect, and participants did not expect responses from a voice assistant similar to what they expect from robots. Therefore, only the robot that was still during the presentation was perceived unresponsive/weird. Also, the sub-themes related to *interaction being odd*, *agent being better than humans in some aspects*, and *a human coach should be used along with the agent* were only found in comments of participants in the two robot conditions. Another interesting observation is that although the audio and feedback were identical in all conditions (recorded once and used in all conditions), the sub-theme related to *feedback/voice being unclear* was only found in comments of participants in the robot conditions, perhaps because the less visually distracting effect of voice assistant agents made participants pay more attention to the feedback content [11].

Although there was a strong preference toward human coaches as compared with the agents, some of the themes suggested that the agents were still perceived as good alternatives when human coaches are not present, or when used along with human coaches. These results are also in line with our research agenda to develop a robot as a public speaking coach, not to replace human coaches, but to provide an alternative and complementary medium.

VI. CONCLUSION

We presented thematic analyses on data collected in an online study with 168 participants, where a type of agent (i.e., a social robot with non-verbal backchanneling, a social robot without backchanneling, and a voice agent) acted as public speaking coaches. The goal of this study was to better understand participants' opinions and perceived benefits and shortcomings of social robots and voice assistant agents as public speaking coaches. We presented a range of themes that contained both positive and negative comments about different aspects of the agents. As the use of social robots as public speaking coaches is a relatively new context, we hope that the themes presented in this paper can inform the design and use of social robots and other agents as public speaking coaches.

ACKNOWLEDGMENTS

This research 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. We thank Pourya Aliasghari and Sahand Shaghaghi for assisting with the design of the web interface used in this study.

REFERENCES

[1] A. K.-F. Lui, S.-C. Ng, and W.-W. Wong, "A novel mobile application for training oral presentation delivery skills," in *Technology in Education. Technology-Mediated Proactive Learning*, J. Lam, K. K. Ng, S. K. Cheung, T. L. Wong, K. C. Li, and F. L. Wang, Eds. Springer Berlin Heidelberg, 2015, pp. 79–89.

[2] D. Apriyanti, Syofiani, S. Ramadhan, and Mukhaiyar, "Improving students public speaking skill through instagram," in *Proc. Intl. Conf. on Language, Literature, and Education (ICLLE 2018)*. Atlantis Press, 2018/12, pp. 268–271.

[3] J. Hernández-March, M. Martín del Peso, and S. Leguey, "Graduates' skills and higher education: The employers' perspective," *Tertiary Education and Management*, vol. 15, no. 1, pp. 1–16, Mar 2009.

[4] H. Trinh, R. Asadi, D. Edge, and T. Bickmore, "Robocop: A robotic coach for oral presentations," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 1, no. 2, jun 2017.

[5] A. W. Blöte, M. J. W. Kint, A. C. Miers, and P. M. Westenberg, "The relation between public speaking anxiety and social anxiety: a review," *J. Anxiety Disord.*, vol. 23, no. 3, pp. 305–313, Apr. 2009.

[6] M. I. Tanveer, E. Lin, and M. E. Hoque, "Rhema: A real-time in-situ intelligent interface to help people with public speaking," in *Proc. Intl. Conf. on Intelligent User Interfaces*, ser. IUI '15. ACM, 2015, p. 286–295.

[7] J. Wang, H. Yang, R. Shao, S. Abdullah, and S. S. Sundar, "Alexa as coach: Leveraging smart speakers to build social agents that reduce public speaking anxiety," in *Proc. Conf. on Human Factors in Computing Systems*, ser. CHI '20. ACM, 2020, p. 1–13.

[8] H. Trinh, K. Yatani, and D. Edge, "Pitchperfect: Integrated rehearsal environment for structured presentation preparation," in *Proc. Conf. on Human Factors in Computing Systems*, ser. CHI '14. New York, NY, USA: ACM, 2014, p. 1571–1580.

[9] X. Wang, H. Zeng, Y. Wang, A. Wu, Z. Sun, X. Ma, and H. Qu, "Voicecoach: Interactive evidence-based training for voice modulation skills in public speaking," *CoRR*, vol. abs/2001.07876, 2020.

[10] K. Pollmann, C. Ruff, K. Vetter, and G. Zimmermann, "Robot vs. voice assistant: Is playing with pepper more fun than playing with alexa?" in *Intl. Conf. on Human-Robot Interaction*, ser. HRI '20. ACM, 2020, p. 395–397.

[11] B. A. Moore and J. Urakami, "The impact of the physical and social embodiment of voice user interfaces on user distraction," *International Journal of Human-Computer Studies*, vol. 161, p. 102784, 2022.

[12] E. Ayedoun, Y. Hayashi, and K. Seta, "Communication strategies and affective backchannels for conversational agents to enhance learners' willingness to communicate in a second language," in *Artificial Intelligence in Education*, E. André, R. Baker, X. Hu, M. M. T. Rodrigo, and B. du Boulay, Eds. Cham: Springer International Publishing, 2017, pp. 459–462.

[13] G. Bodie, A. Vickery, K. Cannava, and S. Jones, "The role of "active listening" in informal helping conversations: Impact on perceptions of listener helpfulness, sensitivity, and supportiveness and discloser emotional improvement," *Western Journal of Communication*, vol. 79, no. 2, pp. 151–173, Mar. 2015.

[14] C. Nass and J. Steuer, "Voices, boxes, and sources of messages," *Human Communication Research*, vol. 19, no. 4, pp. 504–527, 1993. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1468-2958.1993.tb00311.x

[15] J. Gratch, N. Wang, A. Okhmatovskaia, F. Lamothe, M. Morales, R. J. van der Werf, and L.-P. Morency, "Can virtual humans be more engaging than real ones?" in *Human-Computer Interaction. HCI Intelligent Multimodal Interaction Environments*, J. A. Jacko, Ed. Springer Berlin Heidelberg, 2007, pp. 286–297.

[16] D. Kontogiorgos, A. Pereira, O. Andersson, M. Koivisto, E. Gonzalez Rabal, V. Vartiainen, and J. Gustafson, "The effects of anthropomorphism and non-verbal social behaviour in virtual assistants," in *Proc. Intl. Conf. on Intelligent Virtual Agents*, ser. IVA '19. ACM, 2019, p. 133–140.

[17] M. Luria, G. Hoffman, and O. Zuckerman, "Comparing social robot, screen and voice interfaces for smart-home control," in *Proc. Conf. on Human Factors in Computing Systems*, ser. CHI '17. ACM, 2017, p. 580–628.

[18] G. Laban, J.-N. George, V. Morrison, and E. S. Cross, "Tell me more! assessing interactions with social robots from speech," *Paladyn, Journal of Behavioral Robotics*, vol. 12, no. 1, pp. 136–159, 2021.

[19] E. Kimani, T. Bickmore, H. Trinh, and P. Pedrelli, "You'll be great: Virtual agent-based cognitive restructuring to reduce public speaking anxiety," in *2019 8th Intl. Conf. on Affective Computing and Intelligent Interaction (ACII)*, 2019, pp. 641–647.

[20] V. H. Yngve, "On getting a word in edgewise," in *CLS-70*. University of Chicago, 1970, pp. 567–577.

[21] H. H. Clark, *Using Language*, ser. 'Using' Linguistic Books. Cambridge University Press, 1996.

[22] K. Bousmalis, M. Mehu, and M. Pantic, "Towards the automatic detection of spontaneous agreement and disagreement based on nonverbal behaviour: A survey of related cues, databases, and tools," *Image and Vision Computing*, vol. 31, no. 2, pp. 203–221, 2013,

[23] C. Oertel, J. Lopes, Y. Yu, K. A. F. Mora, J. Gustafson, A. W. Black, and J.-M. Odobez, "Towards building an attentive artificial listener: On the perception of attentiveness in audio-visual feedback tokens," in *Proc. Intl. Conf. on Multimodal Interaction*, ser. ICMI '16. ACM, 2016, p. 21–28.

[24] H. M. Rosenfeld and M. Hancks, "The nonverbal context of verbal listener responses," *The relationship of verbal and nonverbal communication*, vol. 25, 1980.

[25] P. Murali, H. Trinh, L. Ring, and T. Bickmore, "A friendly face in the crowd: Reducing public speaking anxiety with an emotional support agent in the audience," in *Proc. Intl. Conf. on Intelligent Virtual Agents*, ser. IVA '21. ACM, 2021, p. 156–163.

[26] "Softbank robotics - group: Global site." [Online]. Available: https://www.softbankrobotics.com/

[27] "Google nest mini." [Online]. Available: https://store.google.com/product/google_nest_mini?hl=en-US&pli=1

[28] "Softbank robotics documentation what's new in naoqi 2.5?" [Online]. Available: http://doc.aldebaran.com/2-5/software/choregraphe/index.html

[29] C. M. Carpinella, A. B. Wyman, M. A. Perez, and S. J. Stroessner, "The Robotic Social Attributes Scale (RoSAS): Development and Validation," in *2017 12th ACM/IEEE Intl. Conf. on Human-Robot Interaction (HRI*, 2017, pp. 254–262.

[30] C. Bartneck, D. Kulić, E. Croft, and S. Zoghbi, "Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots," *International Journal of Social Robotics*, vol. 1, no. 1, pp. 71–81, Jan 2009.

[31] V. Braun and V. Clarke, "Using thematic analysis in psychology," *Qualitative Research in Psychology*, vol. 3, pp. 77–101, 01 2006.

[32] J. R. Landis and G. G. Koch, "The measurement of observer agreement for categorical data," *Biometrics*, vol. 33, no. 1, pp. 159–174, 1977.