

Robot public speakers' effect on audience affective reaction and attention allocation

Jacqueline Urakami
Tokyo Institute of Technology
Tokyo, Japan
urakami28@gmail.com

Marie-Luce Bourguet
Queen Mary U. of London
London, UK
marie-luce.bourguet@qmul.ac.uk

Gentiane Venture
Tokyo U. of Agriculture and Technology
Tokyo, Japan
venture@cc.tuat.jp

Social robots delivering public speeches have a wide range of practical applications as stand-ins for educators, experts, or entertainers. The goal of our work is to investigate how a social robot should be programmed to deliver an effective public speech. Applying a mixed methods research design to collect quantitative and qualitative data, we have conducted a study, which compares a human speaker with a semi-anthropomorphic social robot speaker (the SoftBank Pepper robot). The robot was programmed to mimic the behaviour patterns of the human delivering the same speech. The study results show that the robot is perceived as intelligent and rational, which contributes to effective delivery of the message content. However, the robot struggles with actively engaging the audience and with establishing an emotional connection. In addition, the behavioural patterns that appear natural in the human speaker tend to be distracting in the robot. Suggestions for the improved design of robot public speakers are discussed, which include implementing rhetoric skills, exploiting and synchronising the robot's specific communication channels, and creating a robot persona.

Public speaking, audience, social robot, body gestures, facial expressions, affect, visual attention.

1. INTRODUCTION

Public speaking is a form of communication in which a single speaker addresses an audience to inform, persuade, or entertain [Trenholm 2017]. The speaker is usually physically distant from the audience and therefore must speak clearly and use gestures and visual aids to be seen and heard. Generally, we would expect a human speaker to deliver the message. But what about a social robot delivering a public speech or a lecture? Could a social robot deliver a message as well as a human speaker? A robot public speaker could be useful in a variety of circumstances. Robots could represent speakers or teachers around the world, increasing the reach of the message. The "human speaker" would not have to travel and could deliver a speech at the same time in different locations. In educational environments, a robot speaker could deliver a lecture content in a manner similar to a human lecturer, freeing up teacher resources for individual student support.

Robots certainly still lack the social skills necessary to deliver an effective speech. Restrictions in movement and emotional expressiveness limit the range of expressions a robot can exhibit. On the other hand, because of their enduring novelty and appeal robots can attract large crowds wanting to listen to them. Building on the fascination many have with robots, we believe that robot public speakers can be ambassadors for a good purpose to deliver important messages to a wider audience.

The goal of our research is to investigate how a robot public speaker should be programmed to deliver an effective message by comparing the robot speaker to a skilled human speaker. Even though knowing that the robot speaker at this stage will likely under-perform compared to the human speaker we believe that such a comparison is useful to identify weak as well as strong points of the robot speaker. By collecting quantitative as well as qualitative data we conducted a study to inform suggestions for the improvement of the robot speaker. The results of this study will help develop robotic speakers that can deliver engaging, entertaining, and compelling messages to a wider audience, bringing us one step closer to a realistic future with stand-in robotic speakers. The study was conducted with social distancing measures in place, so participants rated videos of the robot and human speech in an online experiment.

2. THEORETICAL BACKGROUND

2.1 Delivering an effective message

How to successfully deliver a message has been studied for many centuries. Already the ancient Greek philosopher Aristotle described in his studies of rhetoric, how to successfully influence an audience using three methods: Ethos, pathos, and logos. Ethos is about establishing personal credentials, persuading the audience to trust the speaker, acknowledging the speaker's competence;

pathos is about the ability to evoke an emotional response and get the audience to feel and be emotionally involved; and logos is about making reasonable, logical arguments to get the audience thinking.

Modern communication theory has taken a slightly different approach to communication skills and to what makes a good speech. In the education domain, teachers' communication skills, which are identified as one of the most effective contributors to student's achievement, have been examined under five dimensions: empathy, transparency, equality, effectiveness, and competence [Gulec and Leylek 2018]. They have also been described as involving verbal, nonverbal and para-verbal components [Muste 2016]. In [Bambaeeroo and Shokrpour 2017], the impact of the teachers' non-verbal communication is emphasised.

Overall, modern approaches agree with ancient philosophers that credibility (e.g., trustworthiness), identification (e.g., perceived connection between the speaker and the listener), and appeal (e.g., emotional appeal) are important factors in effective delivery of a message [Hovland et al. 1953] [Higgins and Walker 2012]. We have therefore based our evaluation of the effective delivery of a speech on these three aspects, naming them for simplicity after the Greek rhetoric doctrine of ethos, pathos, and logos.

2.2 Social robots in public settings

Social robots can be an effective medium for instruction and communication with users in a public setting. Robots show promise as information assistants and store clerks, among others. Numerous studies have used robots as tour guides in museums and exhibitions [Matsumoto et al. 2020; Wang and Christensen 2018; You and Lin 2019]. Many of these studies focus on how robots can initiate interactions with visitors [Iio et al. 2020] [Rashed et al. 2015] or convey information [Velentza et al. 2019] [Velentza et al. 2020]. A meta-review on robots in public spaces [Mubin et al. 2018] showed that educational scenarios for public speaking robots are the most popular applications and most of these focus on providing information. In such positions, robots will be expected to be experts in their field. However, expertise alone is not enough; effectively communicating expertise is also necessary to build trust and ensure compliance with the information they provide.

The presence of a physical robot in the same location as the learner enables co-location and can

support collaboration between the robot and the learner. The most common form of robot tutoring is one-to-one tutoring, where a robot teaches a learner individually, e.g., [Ramachandran et al. 2019] [Baxter et al. 2017] and only few studies have examined robots teaching a group of learners., e.g., [Huang and Hoorn 2018] [Edwards et al. 2016]. A review of social robots as tutors [Belpaeme et al. 2018] showed that robots perform well when the educational task is limited, approximating the performance of human tutors in cognitive tasks. However, it remains unclear what characteristics of a robot tutor may contribute to learning success. In a study by Striepe et. al. [Striepe et al. 2021] an emotional and a neutral robot produced comparable effects in participants' sense of immersion in the content and affective responses during a storytelling task. Konjin and Hoorn [Konjin and Hoorn 2020] reported that underachieving students benefited more from a neutral robot than from one that exhibited a social behaviour. Contrary to previous research our study does not focus on learning outcome but rather on message delivery as one aspect that might be important for a successful robot speaker and robot teacher likewise.

2.3 Research question

Previous studies have demonstrated that robots can effectively be used for storytelling [Striepe et al. 2021], teaching [Saerbeck et al. 2010] [Brown and Howard 2013], social interaction, rehabilitation and companionship [Uluer et al. 2020]; but very few studies have compared humans and robots in these roles. The few studies that have compared human and robot speakers focused on the audience's comprehension and retention of the information delivered by the speaker [Li et al. 2016] [Palanica et al. 2019]. For our study, we created and filmed a speech delivered by a robot that mimicked the gestures and motions of a human speaker but fell short of being completely human-like as it didn't display facial expressions and spoke in a monotonous voice. The choices we made when creating the robot's behaviours were dictated by two major reasons: (1) dealing with the mechanical constraints of the robot, and (2) avoiding the uncanny valley effect [Seyama and Nagayama 2007]. The main research questions were: Can a social robot, which mimics the gestures and motions of a skilled human speaker, deliver a speech that triggers desirable audience's affective response? The main aim of this research was to uncover the strengths and weaknesses of a social robot when given the role of a public speaker or teacher, and to devise robot-specific speech delivery strategies.

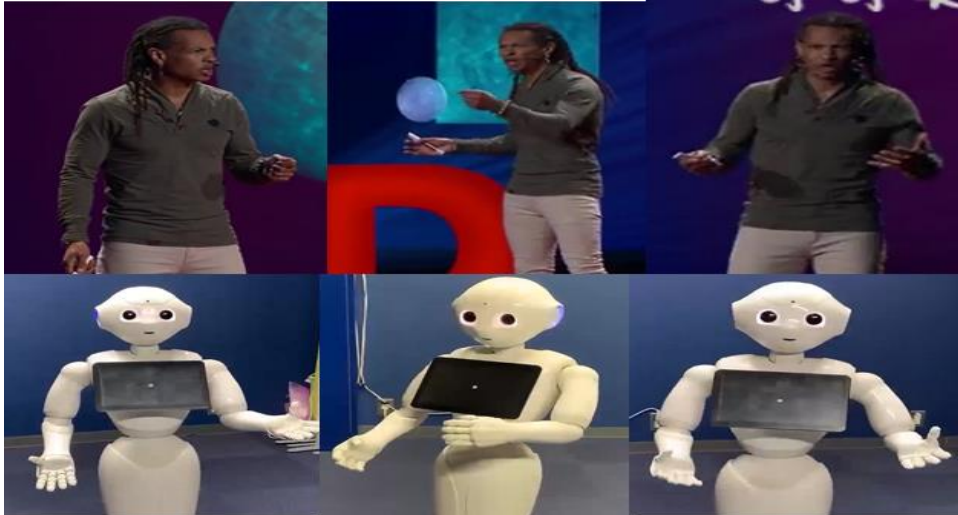


Figure 1: Human and robot speakers

3. METHOD

3.1 Participants

Twenty-eight university students (14 males, 14 females) participated voluntarily in this study. Participants were on average 25 years old ($M = 24.61$; $SD = 7.51$). Approval was granted by the Ethics Committee of the University. Informed consent was obtained from all individual participants.

3.2 Material

We programmed the social robot Pepper [Robotics 2021] to deliver a short speech in its robotic voice while mimicking the behaviour of a human speaker delivering the same speech (gestures, pace, and body movements). The Choregraphe software used to program Pepper has a built-in virtual robot that can be used for the design and testing of the robot's movements. The movements are composed from motion boxes, and when several boxes are chained from the input to the output box, the sequence of the robot's motion and behaviours is complete and can be run using a real robot.

A five and a half minutes long TED talk by Moriba Jah about "the world's first crowd-sourced space traffic monitoring system" was chosen [Jah 2019]. Choosing a TED talk implied that the speaker was likely to demonstrate high public speaking skills, which was setting the bar high for the robot, but also an opportunity of instilling in the robot Pepper realistic public speaking behaviours by mimicry.

The two speakers, human and robot, had in common the content of the speech (same words), the pace of delivery, and the body gestures and movements, including arms and hands gestures, torso and neck orientations (within the mechanical

constraints of the robot). The differences lied in their appearance (see Fig. 1), their voice and the expressivity of their face.

3.3 Questionnaire (see Appendix)

The videos were embedded in a questionnaire and order of videos was counterbalanced. After watching each video participants responded to questions about their affective reaction and attention allocation.

3.3.1 Self-reported affective reaction

Participants' affective reactions were measured on the dimensions: unpleasant/pleasant (i.e., valence); calm/excited (i.e., arousal); and tired/awake with 7-point Likert scales. Questions were based on the valence-arousal model of emotions [Watson and Tellegen 1985]. The dimension tired/awake was added based on a previous study by McAdams et al. [McAdams et al. 2017].

3.3.2 Perception of the speech and attention distribution

Participants rated ten statements on 7-point Likert scales regarding the perceived logos (3 items), ethos (4 items) and pathos (3 items) of the speeches. Two questions regarded attention distribution and distraction. In these questions, participants ranked to what extent different features of the presenter and the background (e.g., face, hand and arm gestures, body torso and chest, motion of the entire body, background) attracted their attention or were seen as distracting.

3.4 Self-reported impressions

Additionally, participants were asked to provide qualitative feedback in the form of a short essay describing their impressions of the speakers.

4. RESULTS

4.1 Evaluation of affective reactions

The robot speaker was rated as giving a neutral impression (valence), feeling of calm (arousal dimension), and tired (awake dimension). The human speaker created a positive impression, triggering arousal, and feeling of being awake (see Fig. 2). In comparison, the human speaker evoked a more positive affect (valence) ($t(27) = -5.41, p < .001$), and was rated as being more exciting (arousal) ($t(27) = -6.44, p < .001$) and keeping listeners more awake ($t(27) = -7.70, p < .001$) compared to the robot speaker.

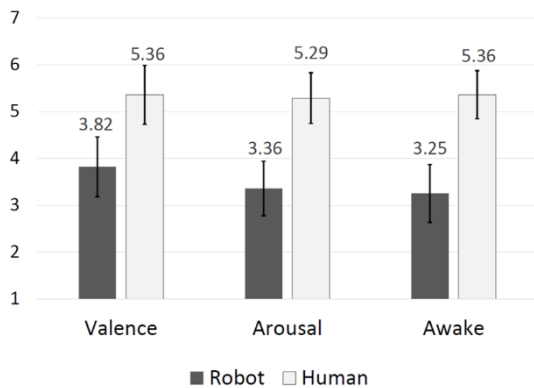


Figure 2: Affective reactions

The robot speaker was perceived as being smart, critical, and rational (logos), was evaluated as neutral in terms of identifiability, competence and trustworthiness (ethos), and received lowest ratings for being engaging and emotionally stimulating (pathos) (see Fig. 3). The human speaker received high scores across all three dimensions as being able to engage the audience emotionally and stimulate imagination (pathos), being trustworthy and competent (ethos), and was rated high in being rational, smart and critical (logos). Compared to the robot, the human speaker received higher ratings across all dimensions, i.e., for pathos ($t(27) = -7.79, p < .001$), ethos ($t(27) = -7.61, p < .001$) and logos ($t(27) = -3.76, p < .001$).

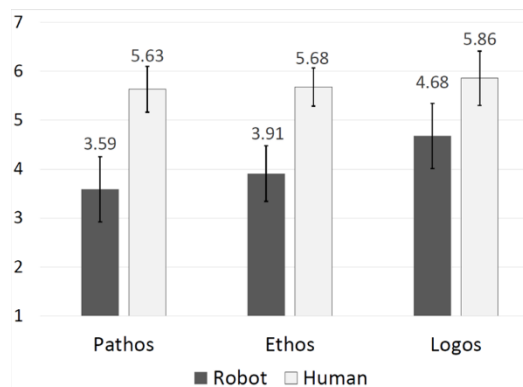


Figure 3: Evaluation of Pathos, Ethos and Logos

4.2 Attention allocation

Participants ranked five different areas of the speaker and the video according to how much attention they paid to each area, as a previous study based on eye tracking [Bourguet et al. 2020] has shown that viewers are able to accurately report on their visual attention distribution when viewing a video.

For the robot speaker, participants allocated most of their attention to the hands ($Md = 2.00$), body ($Md = 2.00$), face ($Md = 3.00$) and torso ($Md = 3.00$), and least attention was given to the background ($Md = 5.00$). For the human speaker most attention was given to the face ($Md = 1.00$), followed by the hands ($Md = 2.00$), body motion ($Md = 3.00$), torso ($Md = 4.00$), and least attention was given to the background ($Md = 5.00$).

In a similar way we asked participants which areas of the speaker and the video they found most distracting. Interestingly the areas that were regarded as attracting attention in the video of the robot speaker, were also evaluated as being distracting; hand ($Md = 2.00$) and body motion ($Md = 2.00$); torso ($Md = 3.00$); face ($Md = 4.00$); and background ($Md = 5.00$). For example, the robot's hands and body motion were regarded as attracting attention but were also seen as causing distraction. On the contrary, for the human speaker areas causing distraction differed from areas attracting attention, with the background causing the most distraction ($Md = 1.50$), followed by body motion ($Md = 2.50$), hands ($Md = 3.00$) and face ($Md = 3.00$), and least the torso ($Md = 4.00$). As indicated by the data, the face of the human speaker attracted the most attention but was ranked only 3rd in being distracting.

4.3 Self-reported impressions

Twenty out of twenty-eight participants provided detailed feedback in the form of a short essay about their impressions of the speakers. The essays were qualitatively analysed aiming to explore significant themes mentioned by the participants. They reflected especially upon four areas: 1) The voice of the robot; 2) Gestures; 3) Motions; and 4) Facial expressions.

4.3.1 Voice

Ten participants mentioned that the voice of the robot lacked intonation and was monotonous. The lack of intonation in the voice was associated with difficulties to understand the important points in the speech ($n = 2$), and with making the speech less engaging ($n = 3$). This might also be related to difficulties in staying focused and requiring more energy to understand what the robot had said ($n = 2$). Furthermore, eight participants mentioned a lack of emotionality in the robot's speech. However, one

participant mentioned that the lack of emotion might be an advantage, because the speech sounded more scientific and trustworthy. Positive aspects of the robot speaker were that it conveyed information well ($n = 3$), had a pleasant appearance ($n = 1$), and created a good first impression ($n = 2$).

4.3.2 Gestures

The gestures of the robot speaker were modelled according to the human speaker. However, participants especially mentioned that the gestures were distracting ($n = 5$) and unnatural ($n = 6$). From the listeners' perspectives, the robot's gestures appeared to be incongruent ($n = 2$) and did not seem to support the speech ($n = 4$). In this regard, participants mentioned that gestures "did not provide additional information", "seemed meaningless", "did not change with the speech content", and "were hard to understand".

4.3.3 Motions

Similarly, participants reported that the whole body motion of the robot was distracting ($n = 4$). The sound coming from the motors of the robot when moving around seemed to have especially contributed to the feeling of distraction. Furthermore, participants felt that the motions were unnatural ($n = 9$). This impression seemed to come from unsynchronised motions ($n = 1$), and the robots body facing towards the front when moving side way ($n = 1$).

4.3.4 Facial expressions

Six participants mentioned the lack of facial expressions of the robot speaker. It was said that the lack of facial expressions made the speech less engaging ($n = 3$), gave the impression of an emotionless speaker ($n = 1$) and the speech was boring ($n = 1$).

Participants also provided some suggestions on how the robot speaker could be improved. While seven participants suggested that the robot should be more human-like, especially its voice ($n = 3$), or movements ($n = 1$), two participants preferred a robot that was less human-like. Four participants suggested that the robot should interact and engage more with the audience by maintaining eye contact, asking questions, or adding humour. Furthermore, it was suggested that the robot speaker would improve by having facial expressions and showing emotions.

5. DISCUSSION

The goal of our study was to investigate how a robot public speaker should be programmed to deliver an effective message by comparing it to a skilled human speaker. The robot speaker showed potential in being perceived as smart and rational (important factors to deliver the content of a speech)

but had difficulties to engage the audience because of its lack of emotional interaction. This was most evident in the measured affective responses on the dimensions of valence, arousal, and wakefulness, all of which trended more toward neutral. In other words, the robot did not elicit negative affective responses, but it failed to evoke positive affective responses necessary to spark interest and engage the audience in meaningful ways. Although the robot's body movements and gestures mimicked the human speaker, they were perceived as distracting in the robot, and were criticised as unnatural and incoherent.

The audience's comments on the incoherence of the robot's behaviour are interesting and have potential implications on the design of a robot speaker. Significantly, the audience's attention allocation is very different for the robot and for the human speaker. In the human speaker condition, the audience allocates most of its attention to the face, stressing the importance of facial expressions and possibly eye and lip movements when giving a speech (one participant felt disturbed by the fact that the robot's mouth wasn't moving). In the robot speaker condition, the audience allocates most of its attention to the most mobile parts of the robot, i.e., its hands and body, which, in the absence of facial expressions, eye and lip movements, become the main nonverbal channels that participants rely on for a better reception of the speech. However, the robot's gestures and motion alone failed to create adequate emotional interaction and affective response in the audience, resulting in the low ratings of the robot's pathos. Furthermore, these gestures, when performed by the robot are judged distracting and unhelpful, which is not the case for the human speaker.

A speaker's behaviour is fundamentally multimodal and is globally perceived as the result of combining various communication channels (e.g., facial expressions, voice, intonation, gestures, motion, gaze). Designing a robot speaker by mimicking some of the behaviours of a human speaker is not enough to create an effective robot speech performance. It is in fact counterproductive to strive for human-likeness on one of the communication channels, when the other channels are perceived as non-human-like. Instead, completely new sets of multimodal behaviours should be devised, taking advantage of robot-specific communication channels (e.g., availability of screens, lights, and sound effects) alongside human inspired gestures, postures, and facial expressions when available. These robot-specific multimodal behaviours are more likely to meet audience's expectations and create an effective speech delivery.

This study had to be conducted online, but ideally, we wanted to compare the video of a human speaker with a robot giving a speech in front of a

larger audience. We expect that a robot "in person" can create a stronger effect on the audience than just watching a video of it. A further limitation of the study is that we relied on participants' self-evaluation for affective reactions. Adding physiological measures (e.g., eye movements, posture) as done in a study by Bourguet et al. [Bourguet et al. 2020] would increase the reliability of the measurements.

6. IMPLICATIONS

Effective human public speaking skills are too complex still to be imitated by social robots. They include not only appropriate use of gestures and precise control and inflection of the voice, but also choice of vocabulary and register, use of humour and enthusiasm, ability to develop a good rapport with the audience and effective use of questions [Abella and Cutamura 2019]. Given the complexity of the task, a better approach to the design of a robot speaker should take advantage of the robot's specificity. Its behaviour should be designed to reinforce its existing strengths rather than undermine them. We recommend the four following areas for improving a robot speaker: (1) implementing rhetoric skills; (2) exploiting robot specific communication channels; (3) synchronising communication channels; and (4) creating a robot persona.

For a human speaker speech preparation and training often involve a considerable amount of effort. Considering the limited emotional expressiveness of a robot and the problems associated with the uncanniness of overly anthropomorphic and emotional robots, a robot's public speech must first of all effectively use language. This can be achieved by telling a compelling story, using humour, and emphasising the core message. In addition, greater audience engagement can be achieved through rhetorical questions and by delivering a message from the robot's perspective.

Robots should take advantage of their unique communication capabilities that come from their appearance, built-in sensors, and/or unique output channels. For example, screens, projections, specific body movements, and sounds can accompany the speech to increase comprehension and improve entertainment value. Nonverbal communication is an important part of every speech and robots could use unique nonverbal codes such as olfactory displays or light and colours.

Creating redundancy by using multiple synchronised communication channels is also important. A human's speech is a synchronised display of voice, gestures, facial expressions, and body movements, and humans are sensitive to conflicting messages being sent through different communication channels [Trenholm 2017]. Therefore, it is important

to synchronise the messages sent by a robot's multimodal channels to avoid uncanny expression (e.g., synchronisation of speech content, gestures, body movements, and voice).

Creating a robot displaying specific personality characteristics could increase trust, competence and identification with the speaker increasing its perceived ethos. Depending on the purpose of the speech or audience, the robot can be an expert, an outsider bringing in a new perspective, or the person from next door. It would be very interesting to see what kind of persona is the most successful in delivering a specific message to a specific audience. The goal would be to design a set of behaviours which reinforce the robot's public identity (ethos) while matching the topic of the speech and the audience it is delivered to.

7. ROBOT EMBODIMENT

Finally, the question of different robots' embodiment on their ability to deliver an effective speech remains. Pepper is a human-sized semi-anthropomorphic social robot, friendly looking and approachable, who has already been used in many different social settings. He can perform a wide range of arm, hand, and body movements, although he cannot display facial expressions. But what effect would have a completely non-anthropomorphic robot speaker on its audience?

We have decided to replicate the Pepper study using a different robot, which is non-anthropomorphic in its shape but can display facial expressions. We chose the Anki Vector robot (see Fig. 4), with the expectation that well timed facial expressions by Vector will add the social cues that have been missing in Pepper. Facial expressions are a powerful tool to convey emotions and our study has shown that participants paid close attention to the face of the human speaker.

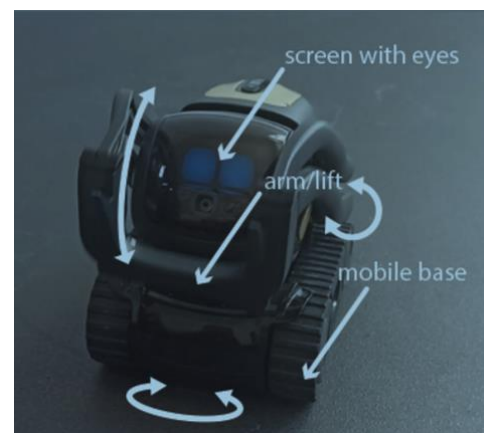


Figure 4: The non-anthropomorphic Anki Vector robot

By displaying facial expressions, we expect that Vector will be better able to establish an emotional

connection to the audience compared to Pepper. However, programming a non-anthropomorphic robot to mimic a human speaker poses new challenges. A systematic scheme to retarget human motions and facial expressions into the various modalities of the robot is necessary. In [venture et al], such a scheme based on video analysis that extracts relevant information from a human speaker's behaviour and send them to the robot is described (see Fig. 5). The robot is then controlled using this information depending on its interaction modalities and movement abilities. We have used this scheme to program Vector to deliver the same speech as Pepper and will soon conduct a new user study to compare the effect of the two robot speakers on their audience affective reactions and attention allocation.

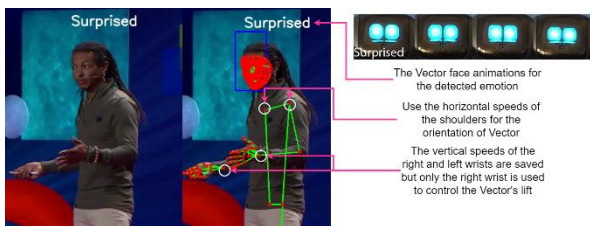


Figure 5: Retargeting of human behaviour to the non-anthropomorphic Anki Vector robot [Venture et al.]

8. CONCLUSION

We believe that employing robot speakers as stand-ins for public speakers is feasible, if the robot's speech and behaviour is designed to be engaging and if the robot is able to create an emotional connection to the audience.

The Pepper study investigated the impact of a semi-anthropomorphic social robot with no facial expressions on audience affective reactions and attention allocation. It provides valuable insights into the possible strengths and limits of a robot speaker. Nonverbal communication is an important part of every speech and robots should use unique nonverbal codes to accompany their speech. With Pepper, we found that the number and amplitude of gestures (especially the beat gestures) should be reduced to achieve a better match with the robot's perceived persona, and to avoid the unwanted distraction they created.

Future work is to compare Pepper and Vector in terms of their ability to evoke the positive affective responses necessary to spark interest, motivate the audience to listen, and engage the audience in meaningful ways. In a future study, Vector's single arm's gestures will still be inspired from human performance using video data information extraction and retargeting. They are however going to be more limited both in frequency and amplitude to achieve a better match with the speaker itself, i.e., Vector. A new user study will be designed to find out how a

"human-like" performance (Pepper's) compares with a more robot-like performance, where the Vector robot's gestures coincide with the human gestures but are very different in terms of type and amplitude. We anticipate that Vector's gestures, combined with facial expressions, will be found more natural and less distracting. It is however harder to anticipate if and how this behaviour will contribute to evoke positive affective responses from the audience, which is something Pepper failed to achieve despite its semi-anthropomorphic and friendly appearance.

ACKNOWLEDGMENTS

We thank all the volunteers who participated in our study.

9. REFERENCES

- Abella, R.C. and Cutamora, J.C. (2019) Approach to Public Speaking Skills Development in an Educational Organization: A Grounded Theory. *European Journal of Education Studies* 6(3) 232-264.
- Bambaerero, F. and Shokrpour, N. (2017) The impact of the teachers' non-verbal communication on success in teaching. *Journal of Advances in Medical Education & Professionalism* 5(2) 51–59.
- Baxter, P., Ashurst, E., Read, R., Kennedy, J. and Belpaeme, T. (2017) Robot education peers in a situated primary school study: Personalisation promotes child learning. *PloS one* 12(5):e0178126.
- Belpaeme, T., Kennedy, J., Ramachandran, A., Scassellati, B. and Tanaka, F. (2018) Social robots for education: A review. *Science robotics* 3(21):eaat5954.
- Bourguet, M.L., Xu, M., Zhang, S., Urakami, J. and Venture, G. (2020) The Impact of a Social Robot Public Speaker on Audience Attention. In *Proceedings of the 8th International Conference on Human-Agent Interaction*, online, 10-13 November 2020, pp 60–68.
- Brown, L. and Howard, A.M. (2013) Engaging children in math education using a socially interactive humanoid robot. In *13th IEEE-RAS International Conference on Humanoid Robots (Humanoids)*, Atlanta, GA, 15-17 October 2013, pp 183–188.
- Edwards, A., Edwards, C., Spence, P.R., Harris, C. and Gambino, A. (2016) Robots in the classroom: Differences in students' perceptions of credibility and learning between "teacher as robot" and "robot as teacher". *Computers in Human Behavior* 65, 627–634.

- Gulec, S. and Leylek, B.S. (2018) Communication Skills of Classroom Teachers According to Various Variables. *Universal Journal of Educational Research* 6(5) 857–862. <https://doi.org/10.13189/ujer.2018.060507>
- Higgins C., and Walker, R. (2012) Ethos, logos, pathos: Strategies of persuasion in social/environmental reports. In *Accounting Forum*, Vol. 36. Elsevier, 194–208.
- Hovland, C.I., Janis, I.L. and Kelley, H.H. (1953) *Communication and persuasion; psychological studies of opinion change*. Yale University Press.
- Huang, I.S. and Hoorn, J.F. (2018) Having an Einstein in class. Teaching maths with robots is different for boys and girls. In *13th World Congress on Intelligent Control and Automation (WCICA)*, Changsha, China, 4-8 July 2018, pp 424–427.
- Iio, T., Satake, S., Kanda, T., Hayashi, H., Ferreri, F. and Hagita, N. (2020) Human-like guide robot that proactively explains exhibits. *International Journal of Social Robotics* 12(2) 549–566.
- Jah, M. (2019) The world's first crowdsourced space traffic monitoring system. https://www.ted.com/talks/moriba_jah_the_world_s_first_crowdsourced_space_traffic_monitoring_system?language=en (retrieved 20 May 2022)
- Konijn, E.A. and Hoorn, J.F. (2020) Robot tutor and pupils' educational ability: Teaching the times tables. *Computers & Education* 157:103970.
- Li, J., Kizilcec, R., Bailenson, J. and Ju, W. (2016) Social robots and virtual agents as lecturers for video instruction. *Computers in Human Behavior* 55, 1222–1230.
- Matsumoto, K., Yamada, H., Imai, M., Kawamura, A., Kawachi, Y., Nakamura, T. and Kurazume, R. (2020). Quasi-Zenith Satellite System-based Tour Guide Robot at a Theme Park. In *IEEE/SICE International Symposium on System Integration (SII)*. IEEE, 1212–1217.
- McAdams, S., Douglas, C. and Vempala, N.N. (2017) Perception and modeling of affective qualities of musical instrument sounds across pitch registers. *Frontiers in psychology* 8, Art. 153, 1-19.
- Mubin, O., Ahmad, M.I., Kaur, S., Shi, W. and Khan, A. (2018) Social robots in public spaces: A meta-review. In *International Conference on Social Robotics. ICSR 2018. Lecture Notes in Computer Science*, vol 11357. Springer, Cham, 213–220.
- Muste, D. (2016) The Role of Communication Skills in Teaching Process. In the *European Proceedings of Social Behavioural Sciences EpSBS. Future Academy*. <https://doi.org/10.15405/epsbs.2016.12.52>
- Palanica, A., Thommandram, A. and Fossat, Y. (2019) Adult Verbal Comprehension Performance is Better from Human Speakers than Social Robots, but only for Easy Questions. *International Journal of Social Robotics* 11(2) 359–369.
- Ramachandran, A., Sebo, S.S. and Scassellati, B. (2019) Personalized robot tutoring using the assistive tutor POMDP (AT-POMDP). In *Proceedings of the AAAI Conference on Artificial Intelligence*, 33(1) 8050–8057.
- Rashed, M.G., Suzuki, R., Lam, A., Kobayashi, Y. and Kuno, Y. (2015) Toward museum guide robots proactively initiating interaction with humans. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*, Portland, Oregon, 2-5 March 2015. Extended Abstracts 1–2.
- Softbank Robotics. The robot Pepper. <https://www.softbankrobotics.com/emea/en/pepper> (retrieved 20 May 2022)
- Saerbeck, M., Schut, T., Bartneck, C. and Janse, M.D. (2010) Expressive robots in education: varying the degree of social supportive behavior of a robotic tutor. In *Proceedings of the 28th Int. conference on human factors in computing systems, CHI 2010, Atlanta, Georgia, USA, 10-15 April 2010*, pp 1613–1622.
- Seyama, J. and Nagayama, R.S. (2007) The Uncanny Valley: Effect of Realism on the Impression of Artificial Human Faces. *Presence: Teleoperators and Virtual Environments* 16(4) 337–351.
- Striepe, H., Donnermann, M., Lein, M. and Lugin, B. (2021) Modeling and evaluating emotion, contextual head movement and voices for a social robot storyteller. *International Journal of Social Robotics* 13, 441-457.
- Trenholm, S. (2017) *Thinking through communication: An introduction to the study of human communication*. Routledge.
- Uluer, P., Kose, H., Oz, B.K., Aydinalev, T.C. and Barkana, D.E. (2020) Towards an affective robot companion for audiology rehabilitation: How does pepper feel today? In *29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, Naples, Italy, 31 Aug.-4 Sept. 2020, pp 567-572.
- Velentza, A.M., Heinke, D. and Wyatt, J. (2019) Human interaction and improving knowledge through collaborative tour guide robots. In *28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, New Delhi, India, 14-18 October 2019, pp 1–7.
- Velentza, A.M., Heinke, D. and Wyatt, J. (2020) Museum robot guides or conventional audio

guides? An experimental study. *Advanced Robotics* 34(24) 1571–1580.

Venture, G., Muraccioli, B., Bourguet, M.L. and Urakami, J. (2022) Can robots be good public speakers? 16th ACM International Conference on Tangible Embedded and Embodied Interaction, 13-16 February 2022, Daejeon, Korea & online, pp 1-6.

Wang, S. and Christensen, H.I. (2018) Tritonbot: First lessons learned from deployment of a long-term autonomy tour guide robot. In 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN), 27-31 August 2018 Nanjing, China, pp 158–165.

Watson, D. and Tellegen, A. (1985) Toward a consensual structure of mood. *Psychological bulletin* 98(2) 219-235.

You, H.C. and Lin, K.W. (2019) Gendered Tour-guide Robots and Their Influence on User Attitude and Behavior. In Proceedings of the 5th International ACM In-Cooperation HCI and UX Conference, 1-9 April 2019, Jakarta, Indonesia, pp 32-35.

A QUESTIONS USED IN THE SURVEY TO EVALUATE THE HUMAN AND ROBOT SPEAKERS

A.1 Self-reported affective response

• How negative / positive did the speech make you feel?

negative 1 2 3 4 5 6 7 positive

• How calm / excited did the speech make you feel?

calm 1 2 3 4 5 6 7 excited

• How tired / awake did the speech make you feel?

tired 1 2 3 4 5 6 7 awake

A.2 Evaluation of speech performance

Please indicate on a scale of 1 (strongly disagree) till 7 (strongly agree) how much you agree or disagree with the following statements.

- The speech stimulated my imagination (Pathos)
- It was fun to watch the speech (Pathos)
- The speech was engaging (Pathos)
- I can identify with the speaker (Ethos)
- The speaker is competent (Ethos)
- The speaker is knowledgeable (Ethos)
- The speaker is trustworthy (Ethos)
- The speaker is smart (Logos)

• The speaker is rational (Logos)

• The speaker is critical (Logos)

A.3 Attention allocation

• Where do you look at most often while watching the video? Please rank the following items starting with the item you pay the most attention to and finishing with the item you pay the least attention to!

– Face

– Hand and arm gestures

– Body torso / chest

– Entire body moving

– Background

• Which areas of the video were distracting for you while watching? Please rank the following items starting with the item that was most distracting and finishing with the item that was least distracting.

– Face

– Hand and arm gestures

– Body torso / chest

– Entire body moving

– Background

A.4 Direct comparison of both speakers

In the last five questions, please compare the two presenters/presentations with each other. Move the slider either towards the Robot Pepper or Moriba Jah according to your preference. Moving the slider to the left indicates you prefer Robot Pepper. Moving the slider to the right indicates you prefer Moriba Jah.

• Which presentation was more impressive?

• Which presenter did you like most?

• Who's performance was better?

• If you could watch another presentation of the presenters, who's presentation do you want to see?

• Who would you choose as an instructor for an online class?