Testing the Error Recovery Capabilities of Robotic Speech

Krantz, Amandus; Stedtler, Samantha; Balkenius, Christian; Fantasia, Valentina

2023

Document Version:
Early version, also known as pre-print

Link to publication

Citation for published version (APA):
Krantz, A., Stedtler, S., Balkenius, C., & Fantasia, V. (in press). *Testing the Error Recovery Capabilities of Robotic Speech*. Abstract from The Imperfectly Relatable Robot, HRI'23, Stockholm, Sweden.

Total number of authors:
4

Testing the Error Recovery Capabilities of Robotic Speech

Amandus Krantz*
amandus.krantz@lth.lu.se
Lund University
Lund, Sweden

Samantha Stedtler*
samantha.stedtler@lucs.lu.se
Lund University
Lund, Sweden

Christian Balkenius
christian.balkenius@lucs.lu.se
Lund University
Lund, Sweden

Valentina Fantasia
valentina.fantasia@lucs.lu.se
Lund University
Lund, Sweden

ABSTRACT

Trust in Human-Robot Interaction is a widely studied subject, and yet, few studies have examined the ability to speak and how it impacts trust towards a robot. Errors can have a negative impact on perceived trustworthiness of a robot. However, there seem to be mitigating effects, such as using a humanoid robot, which has been shown to be perceived as more trustworthy when having a high error-rate than a more mechanical robot with the same error-rate. We want to use a humanoid robot to test whether speech can increase anthropomorphism and mitigate the effects of errors on trust. For this purpose, we are planning an experiment where participants solve a sequence completion task, with the robot giving suggestions (either verbal or non-verbal) for the solution. In addition, we want to measure whether the degree of error (slight error vs. severe error) has an impact on the participants' behaviour and the robot's perceived trustworthiness, since making a severe error would affect trust more than a slight error. Participants will be assigned to three groups, where we will vary the degree of accuracy of the robot's answers (correct vs. almost right vs. obviously wrong). They will complete ten series of a sequence completion task and rate trustworthiness and general perception (Godspeed Questionnaire) of the robot. We also present our thoughts on the implications of potential results.

CCS CONCEPTS

• **Computing methodologies → Cognitive science**; • **Human-centered computing → HCI theory, concepts and models**.

KEYWORDS

robotics, trust, human-robot interaction, anthropomorphism

ACM Reference Format:
Amandus Krantz, Samantha Stedtler, Christian Balkenius, and Valentina Fantasia. 2023. Testing the Error Recovery Capabilities of Robotic Speech. In *Proceedings of The Imperfectly Relatable Robot*. ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

*Both authors contributed equally to this research.

The Imperfectly Relatable Robot, March 13, 2023, Stockholm, Sweden

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

We have previously shown that possessing the ability to speak can potentially be efficient at mitigating loss of trust in case of a trivial failure of operation [16]. We theorized then that this effect might be caused by an increase in perceived intelligence, which in turn affects trust. However, another possible explanation is that rather than increasing the perceived robotic intelligence, speech instead changes the perceived intelligence to be more human-like. Since humans are generally accepted as not being perfect, failures and errors in an agent with human-like intelligence would not be punished as harshly as if the same agent was perceived as having a robot-like intelligence, which are often thought of as being perfect. A similar effect was found by Ragni et al. [21] who saw that collaborative performance in cognitive tasks decreased when the robot made mistakes, but that this also brought a significant increase in positive emotions towards the robot. Mirnig et al. [18] also found the same effect and suggest that it is linked to the Pratfall Effect, a psychological phenomenon wherein people with a high perceived competency are viewed more favourably after making a mistake since it makes them appear less superhuman. The exact mechanics and extent of this effect are still debated, and its effect on the perceived trustworthiness of the robot is unknown.

Here, we present plans for an experiment intended to examine how the severity and rate of the error are judged depending on if the robot uses verbal or non-verbal communication in a collaborative sequence completion task. We hypothesize that we again should see a lower loss of trust when the robot speaks, but that the effect does not remain when errors become more frequent or severe.

The experiment seeks to answer the following research questions:

- RQ1: Does the level of accuracy of the robot's advice have an impact on trust ratings, willingness to cooperate, and task performance in a sequence completion task?
- RQ2: Do trust ratings drop less for speaking robots when giving advice that is obviously wrong compared to mute robots, i.e. can anthropomorphism mitigate the effect of errors on trust?

2 BACKGROUND

2.1 Anthropomorphism and Speech

Most studies investigating the influence of anthropomorphism have focused on physical features, often by comparing entirely different kinds of robots [1, 9]. Yet, the sounds and speech that robots utter might be relevant for our perception and interaction with them as well. The usage of language can also signal comprehension on the robot's side and make humans more prone to give commands to the robot [28].

Human-like speech has also been found to affect the level of trust people have in robots. Cameron et al. [5] found that robots that give verbal explanations when an error occurs are not penalized as harshly in evaluations of their trustworthiness as robots that are mute. This effect may not only be due to the contents of the speech (i.e. explanations or apologies), but may instead be due to the ability to speak itself, which many robots do not possess. We showed in previous work that this might be the case [16], when we saw that a robot that simply had the ability to speak was penalized less when evaluating their trustworthiness than a robot that was mute. Similar results were found by [24] who saw that social behaviours have a positive effect on trust and acceptance of robots. Specifically, their participants unanimously voted for speech as the preferred social behaviour.

We have theorized that this effect of speech increasing trust is due to speech causing an increase in perceived intelligence or capability, which has been correlated with trust [11]. Possessing speech could be seen as a signal of several other abilities, such as understanding of language. However, it is also possible that rather than increasing perceived intelligence, speech implies enough human-like abilities that the nature of the perceived intelligence changes from robot-like intelligence to anthropomorphic intelligence. If this is the case, human users might assume that the robot possesses similar capabilities and shortcomings as themselves. The user would then evaluate the robot as they would evaluate a human, rather than as a robot.

2.2 Trust and Errors

A person's tendency to trust an agent is significantly shaped by how reliable it is and on how well it performs [22, 27]. For instance, Daronnat et al. [6] found that trust ratings could be best predicted by including both reliance on an agent, performance and task complexity, as well as the participant's age and gender. On the other hand, when robots commit errors, trust decreases drastically [8, 23]. How much trust is affected by faulty behaviours depends on different factors, such as the type of robot. Ahmad et al. [1] identified a three-way interaction between robot type, error rate and trust ratings. They showed that the humanoid robot was trusted less than the mechanic robot in the low error-rate condition, while it was trusted more in the high error-rate condition. Other aspects that could impact trust in this context are the timing and extent of the error and how often an error occurs [8, 23]. Salem et al. [25] found that an erroneous robot had a negative impact on perceived reliability and trustworthiness compared to a perfect robot, however it did not affect willingness to cooperate. Moreover, it seems like erroneous robots evoke more positive affective responses, but result in lower task performance from the participant's side [21].

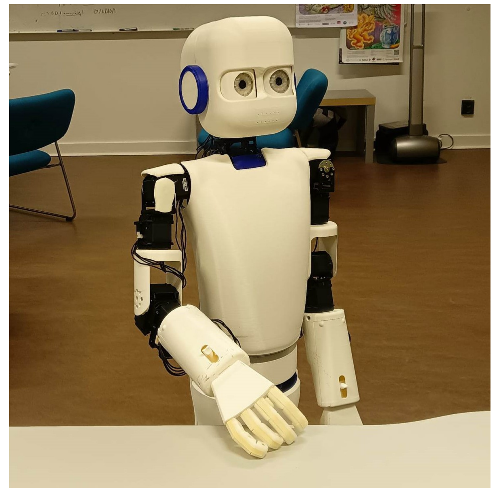

Figure 1: Epi, the humanoid robot that will be used in the proposed experiment.

Thus, it seems relevant to include both the cognitive/affective and the behavioural dimension of trust as they don't seem to always go together [15, 17].

2.3 Trust and Anthropomorphism

Anthropomorphism describes the extent to which an agent holds human-like features and which allows for attributing motivations, aims and beliefs onto it [10]. Together with task performance, anthropomorphism is one of the parameters that positively affect trust (Hancock et al. 2011). For example, participants prefer robots with anthropomorphic features, such as a humanoid face, over robots without such features [4, 13].

Anthropomorphism can also have the opposite effect, as de Visser et al. [7] showed. Participants initially relied more on computer agents than avatar or human agents in solving a digit pattern recognition task. However, after an error, trust declined more rapidly for the non-anthropomorphic agent. This is in line with Ahmad et al. [1] who discovered the same trend, with the anthropomorphic robot being trusted more in the high error-rate condition than the non-anthropomorphic robot. The findings that participants trusted the non-anthropomorphic agent more in the beginning can be explained with the automation bias. According to that concept, trust will be higher for machine-like agents in the beginning, and will change with the observed feedback [20].

3 METHODOLOGY

Our proposed method has a human participant complete a series of number sequence completion tasks using physical tiles. They will be given a sequence of numbers on tiles (e.g. 1, 2, 3, 4) and have to find the next number in the sequence (in this case 5). To aid them in the task, they will be able to ask for advice from the robot Epi (See Fig. 1); a humanoid robotics platform that is being developed at Lund University [14]. Epi has 2 degrees of motion in its head (yaw and tilt) and full motion in its two arms, including two anthropomorphic grippers. Its eyes have 1 degree of freedom (yaw), adjustable pupil size, and adjustable intensity of the illumination in the pupils.

When asked, Epi will indicate which number it thinks is next in the sequence. This can be done either non-verbally, by pointing at a tile, or verbally, by saying something to the effect of "The answer is 5". The verbal answer is intended to signal human-like intelligence following findings in our previous work, while the non-verbal answer signals robot-like intelligence, or the absence of human-like intelligence.

The answer Epi gives can have one of three different levels of accuracy: correct, almost right (e.g. off by 1), or obviously wrong (e.g. off by 20). This would allow us to gauge how wrong the robot is allowed to be before trust starts to decline. For example, if a robot that is perceived to have robot-like intelligence gives an answer that is very obviously wrong, would a participant still go along with the answer because they believe robots to be better at maths than themselves? Or, would the robot be perceived as broken and thus punished more harshly in the trust evaluation?

3.1 Measures

To evaluate the trust relationship, we will use the Trust Perception Scale for Human-Robot Interaction (TPS-HRI) developed by Schaefer [26], which gives the participants a series of questions wherein they'll be asked to estimate how often the robot will behave in a certain way. The scale will be filled out once before seeing Epi display any behaviours, to establish the participants' baseline level of trust and enable comparison of the changes in trust, as recommended by O'Neill [19]. Since the trust relation is believed to be constantly changing [3, 11], the TPS-HRI will also be filled out after every sequence completion task.

While the characteristics of the robot has been found to be the main factor that affects trust in HRI [12], it is still possible that the participants have some preconceived notions about robots that might affect the development of their trust. To control for this, the participants will be asked to fill out the Negative Attitudes Towards Robots Scale (NARS) before seeing Epi [29]. NARS provides estimations on negative attitudes in general, as well as more specific sub-scales on negative attitudes towards situations with robots, social influence of robots, and emotions in interactions with robots.

To confirm whether verbal communication does promote perceptions of human-like intelligence, we will also use the Godspeed scales for anthropomorphism and perceived intelligence [2].

Since Ragni et al. [21] found that task performance and positive perceptions about the robot were not necessarily linked, we will also collect data on how many of the sequences the participants solved, whether they used Epi's help, and their response times.

4 EXPECTED RESULTS AND CONCLUSION

Developing strategies to signal trustworthiness, aid trust calibration and assist humans in problem-solving tasks will be useful for successful human-robot interaction. Displaying fallibility can be useful to calibrate user expectations and avoid automation bias (and thus not overestimate the robot's capabilities). However, it is important to understand the impact of potential errors and error size on the interaction and how negative impacts can be reduced.

We have presented an experimental setup to study the impact of robotic speech on trust in a number sequence completion task. While most studies focus on physical features in the context of anthropomorphism, we are planning to test the same robot (physically) and to merely vary the presence of speech capabilities. Thereby, we expand on previous findings that speech can lead to higher perceived competence and soften the negative influence of errors on perceived trustworthiness. The main aim of the experiment is to compare a humanoid robot with speech to the same robot without speech and measure how it affects self-reported trust, perceived anthropomorphism, performance and willingness to take the robot's advice. More specifically, we want to see whether speech can mitigate the negative effect of errors on trust and whether this might be due to higher perceived anthropomorphism. Overall, we expect performance to decrease and hesitation rate to increase with more obviously wrong answers (low accuracy), but this effect should be more strong for the robot without speech. We hypothesize that speech might indicate humanlike capabilities and, with that, make the robot seem more trustworthy. However, this effect might only hold for errors that are slightly wrong. While erroneous robots seem to be preferred over perfect robots because it makes them seem more human-like [21], there might be a limit to how wrong a robot can be. For instance, a robot that gives slightly wrong advice might be perceived as more anthropomorphic and trustworthy than a robot that gives advice that is completely off.

In the future, results from this study could inform robotic design by taking not only physical features but also language capabilities into account. If we find that speech can reduce the negative impact of errors on trust, speech could be used strategically in robot behaviour to smoothen the interaction and facilitate human-robot collaboration. Lastly, finding out how different degrees of accuracy affect trust could help to anticipate how serious of an error the robot can 'allow' itself to make and thus, how much energy should be dedicated to solving tasks accurately.

ACKNOWLEDGMENTS

This work was partially supported by the Wallenberg AI, Autonomous Systems and Software Program - Humanities and Society (WASP-HS) funded by the Marianne and Marcus Wallenberg Foundation and the Marcus and Amalia Wallenberg Foundation.

REFERENCES

[1] Muneeb Imtiaz Ahmad, Jasmin Bernotat, Katrin Lohan, and Friederike Eyssel. 2019. Trust and Cognitive Load During Human-Robot Interaction. arXiv:1909.05160 [cs]

[2] Christoph Bartneck, Dana Kulić, Elizabeth Croft, and Susana Zoghbi. 2009. Measurement Instruments for the Anthropomorphism, Animacy, Likeability, Perceived Intelligence, and Perceived Safety of Robots. *International Journal of Social Robotics* 1, 1 (Jan. 2009), 71–81. https://doi.org/10.1007/s12369-008-0001-3

[3] Kirsimarja Blomqvist. 1997. The Many Faces of Trust. *Scandinavian journal of management* 13, 3 (Sept. 1997), 271–286. https://doi.org/10.1016/S0956-5221(97)84644-1

[4] Elizabeth Broadbent, Vinayak Kumar, Xingyan Li, John Sollers 3rd, Rebecca Q. Stafford, Bruce A. MacDonald, and Daniel M. Wegner. 2013. Robots with Display Screens: A Robot with a More Humanlike Face Display Is Perceived To Have More Mind and a Better Personality. *PLOS ONE* 8, 8 (Aug. 2013), e72589. https://doi.org/10.1371/journal.pone.0072589

[5] David Cameron, Stevienna de Saille, Emily C. Collins, Jonathan M. Aitken, Hugo Cheung, Adriel Chua, Ee Jing Loh, and James Law. 2021. The Effect of Social-Cognitive Recovery Strategies on Likability, Capability and Trust in Social Robots. *Computers in human behavior* 114 (Jan. 2021), 11. https://doi.org/10.1016/j.chb.2020.106561

[6] Sylvain Daronnat, Leif Azzopardi, Martin Halvey, and Mateusz Dubiel. 2021. Inferring Trust From Users' Behaviours; Agents' Predictability Positively Affects Trust, Task Performance and Cognitive Load in Human-Agent Real-Time Collaboration. *Frontiers in Robotics and AI* 8 (2021), 14.

[7] Ewart J. de Visser, Samuel S. Monfort, Ryan McKendrick, Melissa A. B. Smith, Patrick E. McKnight, Frank Krueger, and Raja Parasuraman. 2016. Almost Human: Anthropomorphism Increases Trust Resilience in Cognitive Agents. *Journal of Experimental Psychology: Applied* 22 (2016), 331–349. https://doi.org/10.1037/xap0000092

[8] Munjal Desai, Mikhail Medvedev, Marynel Vázquez, Sean McSheehy, Sofia Gadea-Omelchenko, Christian Bruggeman, Aaron Steinfeld, and Holly Yanco. 2012. Effects of Changing Reliability on Trust of Robot Systems. In *Proceedings of the Seventh Annual ACM/IEEE International Conference on Human-Robot Interaction (HRI '12)*. Association for Computing Machinery, New York, NY, USA, 73–80. https://doi.org/10.1145/2157689.2157702

[9] Carl F. DiSalvo, Francine Gemperle, Jodi Forlizzi, and Sara Kiesler. 2002. All Robots Are Not Created Equal: The Design and Perception of Humanoid Robot Heads. In *Proceedings of the 4th Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques (DIS '02)*. Association for Computing Machinery, New York, NY, USA, 321–326. https://doi.org/10.1145/778712.778756

[10] Nicholas Epley, Adam Waytz, and John T. Cacioppo. 2007. On Seeing Human: A Three-Factor Theory of Anthropomorphism. *Psychological Review* 114 (2007), 864–886. https://doi.org/10.1037/0033-295X.114.4.864

[11] Ella Glikson and Anita Williams Woolley. 2020. Human Trust in Artificial Intelligence: Review of Empirical Research. *Academy of management annals* 14, 2 (March 2020), 627–660. https://doi.org/10.5465/annals.2018.0057

[12] Peter A. Hancock, Deborah R. Billings, Kristin E. Schaefer, Jessie Y. C. Chen, Ewart J. de Visser, and Raja Parasuraman. 2011. A Meta-Analysis of Factors Affecting Trust in Human-Robot Interaction. *Human factors* 53, 5 (Oct. 2011), 517–527. https://doi.org/10.1177/0018720811417254

[13] Nicholas Hertz and Eva Wiese. 2017. Social Facilitation with Non-Human Agents: Possible or Not? *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 61, 1 (Sept. 2017), 222–225. https://doi.org/10.1177/1541931213601539

[14] Birger Johansson, Trond A Tjøstheim, and Christian Balkenius. 2020. Epi: An Open Humanoid Platform for Developmental Robotics. *International Journal of Advanced Robotic Systems* 17, 2 (March 2020), 11. https://doi.org/10.1177/1729881420911498

[15] Devon Johnson and Kent Grayson. 2005. Cognitive and Affective Trust in Service Relationships. *Journal of Business Research* 58, 4 (April 2005), 500–507. https://doi.org/10.1016/S0148-2963(03)00140-1

[16] Amandus Krantz, Christian Balkenius, and Birger Johansson. 2022. Using Speech to Reduce Loss of Trust in Humanoid Social Robots. In *SCRITA Workshop Proceedings (arXiv:2208.11090)*. IEEE, Naples, Italy, 4. arXiv:2208.13688 [cs]

[17] J. David Lewis and Andrew Weigert. 1985. Trust as a Social Reality. *Social Forces* 63, 4 (June 1985), 967–985. https://doi.org/10.1093/sf/63.4.967

[18] Nicole Mirnig, Gerald Stollnberger, Markus Miksch, Susanne Stadler, Manuel Giuliani, and Manfred Tscheligi. 2017. To Err Is Robot: How Humans Assess and Act toward an Erroneous Social Robot. *Frontiers in Robotics and AI* 4 (2017), 15.

[19] Onora O'Neill. 2020. Questioning Trust. In *The Routledge Handbook of Trust and Philosophy*. Routledge.

[20] Raja Parasuraman, Ewart de Visser, Eva Wiese, and Poornima Madhavan. 2014. Human Trust in Other Humans, Automation, Robots, and Cognitive Agents: Neural Correlates and Design Implications. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 58, 1 (Sept. 2014), 340–344. https://doi.org/10.1177/1541931214581070

[21] Marco Ragni, Andrey Rudenko, Barbara Kuhnert, and Kai O. Arras. 2016. Errare Humanum Est: Erroneous Robots in Human-Robot Interaction. In *2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 501–506. https://doi.org/10.1109/ROMAN.2016.7745164

[22] Paul Robinette, Ayanna M. Howard, and Alan R. Wagner. 2017. Effect of Robot Performance on Human–Robot Trust in Time-Critical Situations. *IEEE Transactions on Human-Machine Systems* 47, 4 (Aug. 2017), 425–436. https://doi.org/10.1109/THMS.2017.2648849

[23] Alessandra Rossi, Kerstin Dautenhahn, Kheng Lee Koay, and Michael L. Walters. 2017. How the Timing and Magnitude of Robot Errors Influence Peoples' Trust of Robots in an Emergency Scenario. In *Social Robotics (Lecture Notes in Computer Science)*, Abderrahmane Kheddar, Eiichi Yoshida, Shuzhi Sam Ge, Kenji Suzuki, John-John Cabibihan, Friederike Eyssel, and Hongsheng He (Eds.). Springer International Publishing, Cham, 42–52. https://doi.org/10.1007/978-3-319-70022-9_5

[24] Alessandra Rossi, Fernando Garcia, Arturo Cruz Maya, Kerstin Dautenhahn, Kheng Lee Koay, Michael L. Walters, and Amit K. Pandey. 2019. Investigating the Effects of Social Interactive Behaviours of a Robot on People's Trust during a Navigation Task. In *Towards Autonomous Robotic Systems (Lecture Notes in Computer Science)*, Kaspar Althoefer, Jelizaveta Konstantinova, and Ketao Zhang (Eds.). Springer International Publishing, Cham, 349–361. https://doi.org/10.1007/978-3-030-23807-0_29

[25] Maha Salem, Gabriella Lakatos, Farshid Amirabdollahian, and Kerstin Dautenhahn. 2015. Would You Trust a (Faulty) Robot? Effects of Error, Task Type and Personality on Human-Robot Cooperation and Trust. In *10th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 8.

[26] Kristin E. Schaefer. 2016. Measuring Trust in Human Robot Interactions: Development of the "Trust Perception Scale-HRI". In *Robust Intelligence and Trust in Autonomous Systems*, Ranjeev Mittu, Donald Sofge, Alan Wagner, and W.F. Lawless (Eds.). Springer US, Boston, MA, 191–218. https://doi.org/10.1007/978-1-4899-7668-0_10

[27] T.B. Sheridan. 1989. Trustworthiness of Command and Control Systems. In *Analysis, Design and Evaluation of Man–Machine Systems 1988*. Elsevier, 427–431. https://doi.org/10.1016/B978-0-08-036226-7.50076-4

[28] Valerie K. Sims, Matthew G. Chin, Heather C. Lum, Linda Upham-Ellis, Tatiana Ballion, and Nicholas C. Lagattuta. 2009. Robots' Auditory Cues Are Subject to Anthropomorphism. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 53, 18 (Oct. 2009), 1418–1421. https://doi.org/10.1177/154193120905301853

[29] Dag Sverre Syrdal, Kerstin Dautenhahn, Kheng Lee Koay, and Michael L. Walters. 2009. The Negative Attitudes towards Robots Scale and Reactions to Robot Behaviour in a Live Human-Robot Interaction Study. *Adaptive and Emergent Behaviour and Complex Systems* (April 2009), 7.