

Is human-like speech in robots deception?

Julia Rosén
University of Skövde
Skövde, Sweden
julia.rosen@his.se

Erik Lagerstedt
University of Skövde
Skövde, Sweden
erik.lagerstedt@his.se

Maurice Lamb
University of Skövde
Skövde, Sweden
maurice.lamb@his.se

KEYWORDS

human-robot interaction, deception, ethics, robo-identity

1 INTRODUCTION

The aim of this extended abstract is to discuss how speech and voice in robots could impact user expectations, and how we, within the human-robot interaction (HRI) research community, ought to handle human-like speech both in research and in the development of robots. Human-like speech refers to both emotions that are expressed through speech and the synthetic voice profile by the robot. The latter is especially important as artificial human-like speech is becoming indistinguishable from actual human speech. Together, these characteristics may cause certain expectations of what the robot is and what it is capable of which may impact both the immediate interactions between a user and robot, as well as a user's future interactions with robots. While there are many ethical considerations around robot designs, we focus specifically on the ethical implications of speech design choices as these choices affect user expectations. We believe this particular dimension is of importance because it not only effects the user immediately, but also the field of HRI, both as a field of research and design. The stance on deception may vary across the different domains that robots are used within; for example, there is a wider acknowledgment of deception in scientific research compared to commercial use of robots. Some of this variation may turn on technical definitions of deception for specific areas or cases. In this paper, we will take on a more general understanding of deception as an attempt to distort or withhold facts with the aim to mislead. We mainly focus on the research perspective of deception, however, there is an overlap between the domains and thus we hope to contribute to the wider discussion of deception with robot use, including the commercial use of such artifacts. Deception is quite a complex issue within the HRI research field as it may not always be used overtly or unintentionally, thus potentially falling through the cracks of an ethical board approval screening. Since roboticists are aware of the robot's capabilities, deception may not even be something that is considered or used outright while developing different aspects of the robot. This can be seen as a version of the problem highlighted by Cooper [3], who argued that a reason for some software to be poorly designed is that the programmers unintentionally created software that fit them, and not the intended users. The roboticists might have similar tacit knowledge, that is then built into the robot. However, the users have certain expectations, and no actual experience, of robots which makes certain robot designs possibly deceptive.

Our own point of departure regarding ethics and HRI begins from the perspective of established protocols concerning ethical principles involving human participants in research [1, 4, 21], as well as the ongoing discussion of deception in HRI research [e.g., 7,

9, 13–15, 17]. While these ethical perspectives are grounded in legal and practical consideration, a more general theoretical perspective on ethics and technology can also be useful. In ethical discourse, there is generally an actor who can be considered ethical, an action that has ethical implications, and a stakeholder that is affected by the action [8]. Actors and stakeholders may occupy one or many on several identities, e.g., individual, group, or community. When considering what is right or wrong, an ethical framework must be chosen as well. Many professional ethics guidelines are rooted in utilitarian and deontological perspectives on ethics. According to utilitarian perspectives, an action is evaluated as right or wrong in terms of the consequences of the action, for example, do the consequences make one better or worse off. Thus, a utilitarian perspective of deceptive human-like speech in robots would focus on the consequences of using human-like speech in a robot [6, 13, 17]. These consequences may be considered at the individual user level, for society, for robotics communities or for any other various actor or stakeholder potentially affected by the robot. Deontological perspectives, often held in contrast to utilitarianism, evaluate an action according to whether or not the action violates general duty that is prescribed to them as a member of society or a particular identity that they may have. The consequence of the action does not matter, even if the violation of a duty would lead to a good outcome. Many professional ethical frameworks such as the ACM Code of Ethics [5] use modified deontological formulations of duties such as "do not lie" in providing guidance to professionals in a field. Thus, a deontological perspective of the use of human-like speech in robots might require consideration of what duties (professional or otherwise) HRI researchers and designers should uphold, regardless of the consequences [6, 13, 17]. In the field of Ethics, it may be necessary to hold to one of these theoretical ethical frameworks, practically, we can use guidance from each to identify a framework that is both agreeable and practical.

2 DECEPTION IN HUMAN-ROBOT INTERACTION

Deception can occur for many reasons and in many contexts, we will focus on two contexts relevant to academic HRI researchers. The first context is deception that may occur, purposely or unintentionally, as a part of research. Examples of this kind of research include using Wizard of Oz methodologies or using cover stories regarding the purpose of research. The second context is deception that may occur when a robot is designed, purposely or unintentionally, in order to leverage deception to accomplish some design objective. Deception in these cases may include relying on fictional robots as examples when communicating to participants, and images and videos of fictional robots are even used for data collection purposes [11]. Deception generally is the act of misleading someone regarding the actual state of affairs, for example, claiming that a

teleoperated robot is acting autonomously. This can lead to inappropriately high expectations of robot capabilities, which in turn might skew expectations of professionals for whom robots might be targeted, for example within the healthcare domain [2]. In the case of relying on fictional robots in researching interaction with real robots is the tacit suggestion from a credible source (the researcher) that the fictional robots are more plausible or imminent than they actually are, which highlights the necessity for proper debriefing between researcher and participant. In particular since the fictional robots are often explicitly associated with either a utopian or dystopian message [11].

While the ethics of deception in the context product design and marketing can be quite fuzzy, there are industry standards and guidelines to rely on [e.g., 5, 18]. In more academic contexts, there is substantial and clear guidance on ethics in human subjects research which can include any HRI experiment where humans are involved. Regarding research and deception the American Psychological Association defines deception in research as *"any distortion of or withholding of fact with the purpose of misleading others. For example, a researcher who has not disclosed the true purpose of an experiment to a participant has engaged in deception."*¹. Proper ethical conduct should be considered in any research field, and special care needs to be considered when using human participants [1, 4, 21].

In the context of research, deception occurs when participants are deceived while participating in a study. This deception can take many forms, and some deceptions may be unintentional, it is important that the researchers aim to identify all potential sources of deception prior to beginning the study. Sometimes, in order to obtain certain data, deception is necessary as some results would not be revealed otherwise [1]. Deception should only be used when the results are important enough to risk deceiving participants. The benefits of the study need to outweigh the risk of the deception and all other options needs to have been ruled out. A challenge for researchers is that they often must balance the duty to advance knowledge with utilitarian concerns involving quantifying the value of findings and harms of deceptive study designs. When deception is used in research, one critical step aimed at minimizing negative consequences of deception includes debriefing participants regarding all deceptions used in a study. This is for the benefit of the participant and the research community at large as it reduces the risk of breakdowns in trust resulting from undisclosed deceptive practices. Debriefing may also be used to correct any misconceptions the participants may have [1]. While exactly what counts as deception or too much deception in research is a debated topic, it remains a common practice in many research fields, including the HRI field [e.g., 7, 9, 13–15, 17]. Because the nature of deception in HRI is so complex, some deception may fall through the cracks while considering the risks of the study. Human-like speech may not be considered deception by all, but it could be deceptive due to the participant's naivete towards robot capabilities.

3 POTENTIAL DECEPTION IN HUMAN-LIKE ROBOT SPEECH

Moore [10] proposed the habitability gap. This gap presents the mismatch between the capabilities and expectations of voice-based agents. The author explains that, as flexibility of the system increases, so does usability. This happens up to the point where users do not know the capabilities of the system which leads to an inability to use the system. The shape of this curve is similar to the Uncanny Valley effect (i.e., robots that may trigger eerie feelings based on its human-like design) which suggests that a similar attention to the design and presentation of the robot could help with the overall usability [10]. Thus, if a user interacts with a robot and it has a human-like voice, there will be high expectations of the robot's capabilities. It is this mismatch that could be considered deception as the robot is not actually capable of being (almost) human. That is, when a robot voice is designed to be human-like, the user may be deceived into thinking the robot is human-like.

When designing a voice for a robot, much like when designing the appearance of the robot, decisions have to be made regarding how much to lean in to anthropomorphism (or other similar metaphors such as zoomorphism). The high fidelity of the projected face of the robot Furhat, in combination with a human-sounding voice, is an example of reliance on anthropomorphism despite the physical shape of the robot is only mimicking a human head. Another strategy is to include aspects in the design that highlight the artificial or non-human nature of the machine, despite the physical shape of the robot mimicking a full human body (e.g., Nao and Pepper robots). Arguably, robots like Furhat are more deceptive since the machine is designed to appear as possessing a human mind, whereas robots like Nao are clearly displaying their artificialness. However, the appropriate voice of the Nao is insinuating that it is the necessary or natural voice of the robot, which is not the actual case. The speakers of the robot are capable of expressing a vast range of sounds and the choice of voice is to a large extent arbitrary. The appropriateness of the voice of a biological agent is, in contrast, determined by physical properties that can be visible to some extent (such as size), the environment in which the agent tend to live (e.g., vast open planes or dense forests), and the purpose of the voice (e.g., conspecific communication, inter-species signaling). A lifetime of experiences of the world has lead to heuristics on what to expect, agents who's voices align with those expectations sound appropriately. For that reason, it might be deceptive to identify and use the appropriate voice, in the sense of the most expected voice, of a robot as it will enforce the idea that the robot is not only consistently designed, but actually has abilities that are necessary consequences emerging from underlying properties.

Another way in which relying on an appropriately robotic voice might be problematically deceptive is that the robot might be *perceived to not* be deceptive by disclosing its robotic nature. This can be problematic since such self-disclosure might lead to increased trust and positive attitudes towards the robot, while encouraging the user to disclose information about themselves [20]. Further, there are some studies in HRI that aim to manipulate participants' reactions by emulating emotions in the robot. Thus, human-like speech relates both to how to make the voice sound human-like and how to make the voice express human-like emotions. For example

¹APA Dictionary of Psychology. Deception. Retrieved February 16, 2022, from <https://dictionary.apa.org/deception>

Niculescu et al. [12], found that pitching a robot's voice, including language cues such as empathy and humor, had an effect on the participants' overall feelings towards the robots. Thus, there is a direct link between emulating emotions in robots and psychological effect in participants. Since robots are not actually capable of feeling emotions, using such strategies are particularly deceptive since it is purposely making someone believe something that is not true, that is, emulated emotions in robots is deception since designers want to *make* users *believe* that the robots can feel emotions [16].

4 CONCLUDING REMARKS

What benefits does human-like voice in robots offer and does it deceive users (ultimately causing psychological harm)? Moreover, do the benefits outweigh the potential harm? These questions could be generalized to any deception that may occur in HRI research, and we therefore intend to contribute to the wider discussion on deception and ethics in HRI. We have previously found that ethical principles is generally under-reported in HRI research [15], and we advocate for increased transparency within the field. Although the issues regarding what is ethically correct when studying or developing human-like speech in robots are too complex to solve in this work, we do believe that increasing the reporting of ethical issues is one way to handle such issues. By being transparent regarding deception, whether it is purposeful or unintentional, we as a community can continue the discussion of how to handle HRI research ethically. However, although it is an important step, not all ethical issues can be solved by transparency within the research community. Incorporating debriefing sessions for participants in conjunction with studies is a standard in psychology that can help with conveying appropriate expectations and understanding to non-experts. In addition, it is important to be vigilant and develop methods for ethical robotics as the field (both academic and commercial) develops and new issues are discovered.

Apart from discussing ways in which the voices of robots can be deceptive, we have mainly focused on the problematic aspects of deception, and through the guidelines focused on deontological perspectives of ethics. There are, however, occasions where deception is more appropriate, at least from a utilitarian perspective. For instance, Matthias [9] offered four instances where deceptive robots in healthcare is morally permissible: a) it is in the best interest of the patient, b) the patient is able to have increased autonomy by being able to make more choices and control the robot, c) the patient is aware of, and able to stop, the deception, and, d) no direct or indirect harm can happen to the patient. For example, the patient may rely on a specific service from the robot (e.g., reminders to take medication) which implies that the patient needs be informed of the actual capabilities and services of the robot to be informed of what they can expect from it. We believe that this framework can be adapted to when human-like speech in robots is used, regardless of the application area.

Regarding the benefits of human-like speech; according to Moore [10], if the end-goal of HRI is to maximize the effectiveness of the interaction, a robot should have a robot sounding voice since a human-like voice will create unrealistic expectation because the other capabilities of the robot does not meet the human-like standard (i.e., the habitability gap). With this line of argument, there

may not be any general benefits to human-like voice, and thus (purposely or unintentionally) deception is not necessary. There are, however, circumstances where robots are more or less trusted based on whether their voice sounds human or artificial *in combination with* the behavior of the robot [19], so the consequences of deception might not easily be generalizable. This can serve as yet another example of the necessity of allowing the issues to be complex when working towards a more ethical HRI. It might not be possible to control users' expectations, but to understand how they are formed and how they interact with other aspects, is an important step to minimize deception.

REFERENCES

- [1] American Psychological Association. 2017. *Ethical principles of psychologists and code of conduct*. <http://www.apa.org/ethics/code/>
- [2] E. Billing, J. Rosén, and J. Lindblom. 2019. Expectations of robot technology in welfare. In *The second workshop on social robots in therapy and care in conjunction with the 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI 2019)*, Daegu, Korea, March 11–14 2019.
- [3] A. Cooper. 2004. *The Inmates are Running the Asylum: Why high tech products drive us crazy and how to restore the sanity*. Sams.
- [4] European Commission. 2018. *Ethics in Social Science and Humanities*. https://ec.europa.eu/research/participants/data/ref/h2020/other/hi/h2020_ethics-soc-science-humanities_en.pdf
- [5] Association for Computing Machinery. 2018. *ACM Code of Ethics and Professional Conduct*. Association for Computing Machinery. <https://www.acm.org/code-of-ethics>
- [6] A. Henschel, R. Hortensius, and E.S. Cross. 2020. Social Cognition in the Age of Human-Robot Interaction. *Trends in Neurosciences* (2020).
- [7] B.F. Malle, K. Fischer, J.E. Young, A.J. Moon, and E. Collins. 2020. Trust and the Discrepancy Between Expectations and Actual Capabilities. In *Human-Robot Interaction: Control, Analysis, and Design*, Dan Zhang and Bin Wei (Eds.). Cambridge Scholars Publishing, Chapter 1, 1–23.
- [8] R. O. Mason. 1995. Applying ethics to information technology issues. *Commun. ACM* 38, 12 (1995), 55–57.
- [9] A. Matthias. 2015. Robot lies in health care: When is deception morally permissible? *Kennedy Institute of Ethics Journal* 25, 2 (2015), 169–162.
- [10] R.K. Moore. 2017. Is spoken language all-or-nothing? Implications for future speech-based human-machine interaction. In *Dialogues with Social Robots*. Springer, 281–291.
- [11] O. Mubin, K. Wadibhasme, P. Jordan, and M. Obaid. 2019. Reflecting on the presence of science fiction robots in computing literature. *ACM Transactions on Human-Robot Interaction (THRI)* 8, 1 (2019), 1–25.
- [12] A. Niculescu, B. van Dijk, A. Nijholt, H. Li, and S.L. See. 2013. Making social robots more attractive. *International journal of social robotics* 5, 2 (2013), 171–191.
- [13] L.D. Riek. 2012. Wizard of oz studies in hri. *Journal of Human-Robot Interaction* 1, 1 (2012).
- [14] L.D. Riek and D. Howard. 2014. A code of ethics for the human-robot interaction profession. (2014).
- [15] J. Rosén, J. Lindblom, and E. Billing. 2021. Reporting of Ethical Conduct in Human-Robot Interaction Research. In *International Conference on Applied Human Factors and Ergonomics*. Springer, 87–94.
- [16] J. Rosén, J. Lindblom, E. Billing, and M. Lamb. 2021. Ethical Challenges in the Human-Robot Interaction Field. *arXiv preprint arXiv:2104.09306* (2021).
- [17] A. Sharkey and N. Sharkey. 2020. We need to talk about deception in social robotics! *Ethics and Information Technology* (2020), 1–8.
- [18] The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. 2017. *Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems* (2 ed.). IEEE. http://standards.ieee.org/develop/indconn/ec/autonomous_systems.html
- [19] I. Torre, J. Goslin, L. White, and D. Zanatto. 2018. Trust in artificial voices: A "congruency effect" of first impressions and behavioural experience. In *Proceedings of the Technology, Mind, and Society*. 1–6.
- [20] L. van Uffelen, J.B. van Waterschoot, and M. Theune. 2021. Show Me Yours if I Show You Mine: Self-disclosure in Conversational Agents. In *Workshop Robo-Identity: Artificial identity and multi-embodiment*. Boulder (virtual), USA.
- [21] World Medical Association. 2018. *Declaration of Helsinki*. <https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/>