Speaking of Trust - Speech as a Measure of Trust

Ella Velner
p.c.velner@utwente.nl
University of Twente
Enschede, The Netherlands

Khiet P. Truong
k.p.truong@utwente.nl
University of Twente
Enschede, The Netherlands

Vanessa Evers
vanessa.evers@ntu.edu.sg
Nanyang Technological University
Singapore, Singapore

ABSTRACT

Since trust measures in human-robot interaction are often subjective or not possible to implement real-time, we propose to use speech cues (on what, when and how the user talks) as an objective real-time measure of trust. This could be implemented in the robot to calibrate towards appropriate trust. However, we would like to open the discussion on how to deal with the ethical implications surrounding this trust measure.

KEYWORDS

trust, speech, human-robot interaction, measure, real-time

ACM Reference Format:

Ella Velner, Khiet P. Truong, and Vanessa Evers. 2021. Speaking of Trust - Speech as a Measure of Trust. In *HRI 2021, March 09–11, 2021, Virtual*. ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Researchers in Human-Robot Interaction (HRI) are often interested in the trust people have in the robot. Understandable, since trust can be a prerequisite for relationship formation and self-disclosure [22, 26]. Another reason for measuring trust, is to assess whether there is appropriate trust, meaning that the trust matches the capabilities of the robot [13]. Overtrust can lead to complacency and dangerous situations, while undertrust can lead to underuse and suboptimal use of the robot's capabilities [7, 13]. At the moment trust is mostly measured by either (self-reported) subjective measures (e.g., questionnaires) that might be influenced by people-pleasing behavior of the subjects, or behavioral measures that constrain the experiment design (e.g., trust games). Moreover, these measures are most often not measured real-time, so it is not possible to achieve optimal trust balance *while* interacting with the robot. Hence, we propose to explore how objective and real-time sensing of trust can be achieved.

Building on the knowledge that speech carries signals of emotions and attitudes [6, 11, 21, 23], we propose to use the user's speech to measure the trust they have in the robot, by looking at *what* they say, *when* they say it and *how* they say it. We are especially interested in how this could help the field of *child*-robot interaction, since subjective measures are even more problematic for this user group [3]. In this short paper we will discuss how we envision a real-time measure of trust in human-robot interaction, but also address the ethical implications surrounding real-time trust calibration.

TRAITS 2021, March 12, 2021, Virtual

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

2 TRUST MEASURES

To measure trust, we can distinguish between three types of measures: subjective, active objective and passive objective measures. We refer to subjective measures as self-reports from the user. While these are widely used by researchers ([10, 20, 27]), they are administered after the actual interaction, hence they cannot be used as a real-time measure for the robot to sense the trust level during the interaction. Furthermore, they rely on the honesty and understanding of the user. While this is already somewhat problematic for adults, when children are involved, questionnaires could be quite unreliable. Children are known people-pleasers and therefore it can be hard to rely on their answers [3].

When interaction designers provoke the user to make a choice during the interaction that would reflect their trust in the robot, this is considered an active objective measure. A classic example of such a measure is the investment game [4]. In this game a truster invests money (or tokens) in a trustee (this investment is the measure of trust), and they later lose, earn or keep the money they invested. While these measures are not dependent on the self-report of users, and thus more objective [14], they constrain the design of the interaction, since they must include these behavior-provoking scenarios.

With a passive objective measure, we can monitor certain behaviors that are correlated with trust, without necessarily being constrained by certain behavior-provoking scenarios. Examples of such passive objective measures are the user's heart rate [8], the distance between the robot and the participant [1], and the words shared by the robot and the user [21]. Some of these could be measured by video observation, which, if automated, would still be an objective measure. However, if humans are involved in the annotation, it could become a less objective measure. We are especially interested in the automated types of measures, because these are not constraining the interaction design, and can be used real-time by the robot in an autonomous way during the interaction.

3 SPEECH AS A MEASURE OF TRUST

Since many human-robot interactions use speech to communicate, it would be useful if the robot could use this to assess the trust level. We propose to look at dialog cues (what people say), interaction cues (when people say it) and vocal cues (how they say it).

First, we can use what people say to measure trust. Scissors et al. [21] already discovered that parties who trust each other often use the same **words**. It could also be interesting to look at **speech acts** that people use, since these have been previously associated with attitudes towards the conversation partner [2]. Speech act classification can also be automated [15].

Second, it would be interesting to look at when the user speaks. Elkins et al. [5] describe a model where the agent takes the user's demographics and voice, and the **duration of the response** into account as a perception of trust. Furthermore, interpersonal attitudes, such as trust, can be reflected in the **turn-taking system** of the conversation [18].

Third, we know that how people say things can show how they feel [11]. Waber et al. [25] found that trust is reflected in the amount of emphasis (a combination of **pitch**, **intensity** and **speech rate**) on words the truster uses, although it could be this is only reflected during first encounters or after trust violations.

We will look at these cues simultaneously as we expect that they complement and relate to each other, which will give us a more complete and better understanding of how trust is expressed in speech.

4 SENSE-THINK-ACT

Measuring trust real-time and objectively would not only be beneficial to HRI researchers because of its objectiveness, but also to designers of HRI, that can incorporate this in a sense-think-act cycle, such as in Fig. 1. An interaction should aim for appropriate trust, since overtrust can lead to complacency and dangerous situations, while undertrust can lead to underuse and suboptimal use of the robot's capabilities [7, 13]. This creates a spectrum of trust levels, where appropriate trust is found in the middle as the balance between overtrust and undertrust. When a user is trusting the robot too much, this could lead to dangerous situations, since the user would believe the robot to always act to the benefit of the user, while this could not be the case. Undertrust, on the other hand, could mean that the user would discard the robot, while it could have been trying to help the user. If the robot could *sense* the trust a person has in it, it could *think* about what the robot should do with this information (to try to lower or raise trust), and it could *act* to reach appropriate trust balance [13]. If under- and overtrust could be avoided by calibrating towards appropriate trust, this could lead to more responsible interactions: the user needs to remain critical towards the robot, without discarding it entirely. Hence, a real-time measure of trust could play a part in creating more responsible human-robot interaction.

5 ETHICAL IMPLICATIONS

Although using speech as a measure of trust has many advantages, as discussed above, its ethical implications should be addressed as well. In order to be able to extract trust from speech, the robot needs to record and analyze the voice of its users. Voice characteristics and speech can give insight in a user's personal information, such as their identity, personality and emotions [12], and hence, is a privacy-sensitive source. Consequently, before a user interacts with the robot, consent needs to be given, where the interaction designers can promise that the voice is only used for the benefit of this interaction. Moreover, the recorded data needs to be stored only when absolutely necessary (e.g., to improve the model) and with the utmost care by securing it with encryption and limited access [16, 17].

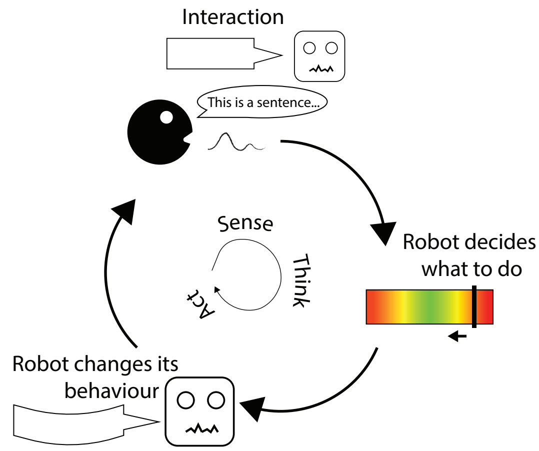

Figure 1: Using speech to integrate a sense-think-act cycle in the interaction

If our trust measure is applied within the sense-think-act cycle, more possible dangers come to mind. If trust can be calibrated towards appropriate trust, it could also be used to actually move away from appropriate trust. This way, robots could intentionally create overtrust, with all its consequences. For example, a company selling dubious products installs a robot to interact with potential buyers, and the robot increases their trust in it, so that when the robot says "You'll definitely need this product", the customer will buy it without thinking much about the possible risks [24]. Therefore, we should monitor that trust calibration is used towards responsible interaction. Furthermore, matching certain behavior of the robot to a trust level could induce stereotypes [9]. For example, voice assistants use female voices because it is deemed more trustworthy, but this could be problematic due to a voice assistant always saying yes (giving the image of a submissive woman).

6 DISCUSSION

In this short paper, we proposed to use speech as a real-time measure of trust. Trust is valuable to measure in an interaction, because it acts as a prerequisite for a fruitful interaction [19] and can be used to calibrate trust levels in real-time. Moreover, using speech cues as a measure would be an objective behavioral measure, without constraining the interaction. However, detecting trust from speech cues brings its ethical challenges. It would need privacy policies and, if implemented in a sense-think-act cycle, could have societal consequences. We hope this paper opens a discussion on how we should balance these two sides of the same coin. Furthermore, the generalizability of speech as a measure of trust should be studied (e.g., does it also apply to human-human interaction?). We also want to acknowledge that trust might also be measurable by other modalities such as facial expressions, body posture, hand gestures, and eye gaze, however, speech is most often the main modality used and could potentially be of more wider use to different kinds of robots (also including voice assistants). Future research could investigate the role of other modalities and ways to measure trust objectively and real-time, possibly in combination with speech.

REFERENCES

[1] Franziska Babel, Johannes Kraus, Linda Miller, Matthias Kraus, Nicolas Wagner, Wolfgang Minker, and Martin Baumann. 2021. Small Talk with a Robot? The Impact of Dialog Content, Talk Initiative, and Gaze Behavior of a Social Robot on Trust, Acceptance, and Proximity. *International Journal of Social Robotics* (1 2021), 1–14. https://doi.org/10.1007/s12369-020-00730-0

[2] Valentin Barriere. 2017. Hybrid Models for Opinion Analysis in Speech Interactions. In *Proceedings of the 19th ACM International Conference on Multimodal Interaction (ICMI '17)*. Association for Computing Machinery, New York, NY, USA, 647–651. https://doi.org/10.1145/3136755.3137035

[3] Tony Belpaeme, Paul Baxter, Joachim De Greeff, James Kennedy, Robin Read, Rosemarijn Looije, Mark Neerincx, Ilaria Baroni, and Mattia Coti Zelati. 2013. Child-Robot Interaction: Perspectives and Challenges. In *International Conference on Social Robotics*. Springer, 452–459. https://doi.org/10.1007/978-3-319-02675-6{_}45

[4] Joyce Berg, John Dickhaut, and Kevin McCabe. 1995. Trust, Reciprocity, and Social History. *Games and Economic Behavior* 10 (1995), 122–142.

[5] Aaron C. Elkins, Douglas C. Derrick, Judee K. Burgoon, and Jay F. Nunamaker. 2012. Predicting users' perceived trust in Embodied Conversational Agents using vocal dynamics. In *Proceedings of the Annual Hawaii International Conference on System Sciences*. IEEE Computer Society, 579–588. https://doi.org/10.1109/HICSS.2012.483

[6] Raul Fernandez and Rosalind W Picard. 2005. Classical and Novel Discriminant Features for Affect Recognition from Speech. In *Ninth European Conference on Speech Communication and Technology*.

[7] Amos Freedy, Ewart DeVisser, Gershon Weltman, and Nicole Coeyman. 2007. Measurement of trust in human-robot collaboration. In *Proceedings of the 2007 International Symposium on Collaborative Technologies and Systems, CTS*. 106–114. https://doi.org/10.1109/CTS.2007.4621745

[8] Kunal Gupta, Ryo Hajika, Yun Suen Pai, Andreas Duenser, Martin Lochner, and Mark Billinghurst. 2019. In AI We Trust: Investigating the Relationship between Biosignals, Trust and Cognitive Load in VR. In *25th ACM Symposium on Virtual Reality Software and Technology (VRST '19)*. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3359996.3364276

[9] I Hwang, Y Lee, C Yoo, C Min, D Yim, and J Kim. 2019. Towards Interpersonal Assistants: Next-Generation Conversational Agents. *IEEE Pervasive Computing* 18, 2 (2019), 21–31. https://doi.org/10.1109/MPRV.2019.2922907

[10] Jiun-Yin Jian, Ann M. Bisantz, and Colin G. Drury. 2000. Foundations for an Empirically Determined Scale of Trust in Automated Systems. *International Journal of Cognitive Ergonomics* 4, 1 (3 2000), 53–71. https://doi.org/10.1207/s15327566ijce0401{_}04

[11] Jaebok Kim, Gwenn Englebienne, Khiet P. Truong, and Vanessa Evers. 2017. Towards Speech Emotion Recognition "in the wild" using Aggregated Corpora and Deep Multi-Task Learning. *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH* 2017-August (8 2017), 1113–1117. http://arxiv.org/abs/1708.03920

[12] Jacob Leon Kröger, Otto Hans Martin Lutz, and Philip Raschke. 2020. Privacy implications of voice and speech analysis – information disclosure by inference. In *IFIP Advances in Information and Communication Technology*, Vol. 576 LNCS. Springer, 242–258. https://doi.org/10.1007/978-3-030-42504-3{_}16

[13] John D. Lee and Katrina A. See. 2004. Trust in automation: Designing for appropriate reliance. *Human Factors* 46, 1 (1 2004), 50–80. https://doi.org/10.1518/hfes.46.1.50{_}30392

[14] Michael Lewis, Katia Sycara, and Phillip Walker. 2018. The Role of Trust in Human-Robot Interaction. In *Studies in Systems, Decision and Control*. Vol. 117. Springer International Publishing, 135–159. https://doi.org/10.1007/978-3-319-64816-3{_}8

[15] Yang Liu, Kun Han, Zhao Tan, and Yun Lei. 2017. Using Context Information for Dialog Act Classification in DNN Framework. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*. 2170–2178. https://www.aclweb.org/anthology/D17-1231.pdf

[16] Andreas Nautsch, Catherine Jasserand, Els Kindt, Massimiliano Todisco, Isabel Trancoso, and Nicholas Evans. 2019. The GDPR & Speech Data: Reflections of Legal and Technology Communities, First Steps towards a Common Understanding. *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH* 2019-September (7 2019), 3695–3699. http://arxiv.org/abs/1907.03458

[17] Andreas Nautsch, Abelino Jiménez, Amos Treiber, Jascha Kolberg, Catherine Jasserand, Els Kindt, Héctor Delgado, Massimiliano Todisco, Mohamed Amine Hmani, Aymen Mtibaa, Mohammed Ahmed Abdelraheem, Alberto Abad, Francisco Teixeira, Driss Matrouf, Marta Gomez-Barrero, Dijana Petrovska-Delacrétaz, Gérard Chollet, Nicholas Evans, Thomas Schneider, Jean François Bonastre, Bhiksha Raj, Isabel Trancoso, and Christoph Busch. 2019. Preserving privacy in speaker and speech characterisation. *Computer Speech and Language* 58 (11 2019), 441–480. https://doi.org/10.1016/j.csl.2019.06.001

[18] Brian Ravenet, Angelo Cafaro, Beatrice Biancardi, Magalie Ochs, and Catherine Pelachaud. 2015. Conversational behavior reflecting interpersonal attitudes in small group interactions. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Vol. 9238. Springer Verlag, 375–388. https://doi.org/10.1007/978-3-319-21996-7{_}41

[19] Samuel Ronfard and Jonathan D. Lane. 2018. Preschoolers Continually Adjust Their Epistemic Trust Based on an Informant's Ongoing Accuracy. *Child Development* 89, 2 (3 2018), 414–429. https://doi.org/10.1111/cdev.12720

[20] 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*, R. Mittu (Ed.). Springer, Chapter 10, 191–218. https://doi.org/10.1007/978-1-4899-7668-0{_}10

[21] Lauren E. Scissors, Alastair J. Gill, and Darren Gergle. 2008. Linguistic mimicry and trust in text-based CMC. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work, CSCW*. ACM Press, New York, New York, USA, 277–280. https://doi.org/10.1145/1460563.1460608

[22] Ju Hyun Song, Tyler Colasante, and Tina Malti. 2018. Helping yourself helps others: Linking children's emotion regulation to prosocial behavior through sympathy and trust. *Emotion* 18, 4 (6 2018), 518–527. https://doi.org/10.1037/emo0000332

[23] Mark Ter Maat, Khiet P. Truong, and Dirk Heylen. 2010. How turn-taking strategies influence users' impressions of an agent. In *Lecture Notes in Computer Science*, Vol. 6356 LNAI. Springer, Berlin, Heidelberg, 441–453. https://doi.org/10.1007/978-3-642-15892-6{_}48

[24] Hans van der Heijden, Tibert Verhagen, and Marcel Creemers. 2003. Understanding online purchase intentions: contributions from technology and trust perspectives. *European Journal of Information Systems* 12, 1 (3 2003), 41–48. https://doi.org/10.1057/palgrave.ejis.3000445

[25] Benjamin Waber, Michele Williams, John Carroll, and Alex Pentland. 2015. A Voice is Worth a Thousand Words: The Implications of the Micro-Coding of Social Signals in Speech for Trust Research. In *Handbook of research methods on trust*. Edward Elgar Publishing, Chapter 26, 302–312. https://digitalcommons.ilr.cornell.edu/articles/905

[26] Lawrence R Wheeless and Janis Grotz. 1977. The Measurement of Trust and Its Relationship to Self-Disclosure. *Human Communication Research* 3, 3 (1977).

[27] Rosemarie E. Yagoda and Douglas J. Gillan. 2012. You Want Me to Trust a ROBOT? The Development of a Human-Robot Interaction Trust Scale. *International Journal of Social Robotics* 4, 3 (8 2012), 235–248. https://doi.org/10.1007/s12369-012-0144-0