Editorial: Failures and Repairs in Human-Robot Communication

Frank Förster [1,*] **and Patrick Holthaus** [1]

[1] *Robotics Research Group, School of Physics, Engineering and Computer Science, University of Hertfordshire, Hatfield, United Kingdom*

Correspondence*:
Corresponding Author
f.foerster@herts.ac.uk

1 INTRODUCTION

This research topic arose on the back of the WTF workshop series (Förster et al., 2022, 2023a) that brought together an interdisciplinary group of researchers ranging from roboticists and computational linguists to conversation analysts and cognitive scientists to openly and frankly discuss failures of (robotic) speech interfaces they experienced when deploying these in their studies. Some of the issues discussed in the workshops are elaborated in the contributed articles below, more pointers can be found in the workshop summary article by Förster et al. (2023b).

This research topic contributes towards two main objectives: Firstly, we provide a platform for reporting commonly occurring communicative failures in human-robot interaction (HRI). Secondly, this topic aims to highlight the opportunity of potential multi-modal repair mechanisms to render robotic speech interfaces more resilient concerning conversational breakdowns. Hence, we include several articles documenting and analysing such failures to shed light on what is largely an unreported issue experienced by many robotics practitioners. Moreover, this topic also contains articles reporting existing research on conversational repair in HRI and position papers outlining the potential of such mechanisms.

2 CONTRIBUTED ARTICLES

Addlesee and Papaioannou (2025) point out a number of practical issues linked to spoken dialogue systems (SDS) based on both their own experience as well as existing literature when deploying social robots in real-world settings. They report evidence of people struggling to understand robots due to an insufficient volume of robots' voices either due to noise in the environment, limited hearing on the part of the human interlocutors or a combination of both. A second, and in some sense symmetrical issue is that robots frequently cannot hear the human interlocutor. This is typically caused by an insufficient number of built-in microphones or a suboptimal placement of these, e.g. microphones being located behind covering materials. Addlesee and Papaioannou further discuss the related problem of ego-noise, that is, noise that is generated by the robot itself, negatively impacting the speech recognition capability of the robot. As the authors emphasize, all of the highlighted issues could be fixed in a relatively straightforward manner if social robots were designed – from the very start – under consideration of their prospective speech capabilities, rather than microphones and speech-related design decisions being integrated and made at a comparatively late design stage.

Galbraith (2024) investigate how virtual assistants deal with the interactionally highly relevant and frequent '*huh?*', an other-initiated, and likely universal repair marker (cf. Dingemanse et al., 2015). They further investigate what repair strategies these assistants utilise when encountering unintelligible speech, and how native speakers judge these different repair strategies. In their study, two different virtual assistants, Google Assistant and Apple's Siri, are compared across two different languages (English and Spanish). Galbraith finds that neither assistant actively produces '*huh?*' but rather employs more specific repair strategies when confronted with unintelligible speech. The assistants frequently have trouble dealing with a '*huh?*' produced by human users, and some of the repair strategies employed by the two assistants were rated negatively by human judges. While these insights were gained by interacting with virtual assistants, we expect some of these to apply to SDS more generally (cf. Lopez et al., 2022).

Tisserand et al. (2024) present a conversation analysis of sequential failures they observed in a large HRI corpus gathered via an in-the-wild study with the Pepper robot that was placed at the entrance of a university library. The failures they found fell frequently into one of four categories: (1) the inability of Pepper's SDS to distinguish different types of conversational actions involving identical key words, here words associated with greetings; (2) the inability to detect when the human interlocutor takes back the initiative, leading to the robot talking over the human; (3) the failure to detect turn-holding devices; and (4) the SDS' inability to detect when two conversational actions are produced within the same turn. Tisserand et al. subsequently outline the requirements for future dialogue systems that would need to be fulfilled to avoid these types of failures and review the current state of the technical literature with respect to these requirements. This paper illustrates how conversation analysis can be used to provide concrete guidance for future technical developments.

One work within this topic (Frijns et al., 2024) investigates mistakes in a robot's knowledge base, in particular those kinds of mistakes that a robot is *not* aware of. The authors present a user study that leverages the human interaction partner to help a robotic system identify and correct its own misconceptions. For that, they initially compare people's preference for speech or visual communication about the robot's knowledge base in a sorting scenario finding that a combination thereof is being preferred by participants. Moreover, unplanned mistakes that occurred during this study have been found to not be covered by existing failure taxonomies in the field of human-robot interaction. As a consequence, the authors introduce the concept of a *productive failure* and argue that failures often occur as a result of multiple, intertwined causes. The study further highlighted that mistakes can play an important role for users when familiarising themselves with a robotic system where they frequently test the robot's limits to better understand its operating principles.

3 CONCLUSION

The articles collated under this research topic highlight frequently occurring failures in robotic speech interfaces when these are deployed in-the-wild, many of which may not be observed when these SDS are assessed via benchmark datasets. They provide several concrete recommendations on how to improve both robot and SDS design to reduce the latter's propensity for failure, and we hope that they will help to guide research efforts to render robotic speech interfaces more resilient when deployed outside of laboratory settings.

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

AUTHOR CONTRIBUTIONS

FF: main author, contributed to writing and proofreading all parts of the paper PH: contributed to writing the introduction and description of contributed articles and proofreading of the whole paper.

REFERENCES

Addlesee, A. and Papaioannou, I. (2025). Building for speech: designing the next-generation of social robots for audio interaction. *Frontiers in Robotics and AI* 11. doi:10.3389/frobt.2024.1356477

Dingemanse, M., Roberts, S. G., Baranova, J., Blythe, J., Drew, P., Floyd, S., et al. (2015). Universal Principles in the Repair of Communication Problems. *PLOS ONE* 10, 1–15. doi:10.1371/journal.pone.0136100

Frijns, H. A., Hirschmanner, M., Sienkiewicz, B., Hönig, P., Indurkhya, B., and Vincze, M. (2024). Human-in-the-loop error detection in an object organization task with a social robot. *Frontiers in Robotics and AI* 11. doi:10.3389/frobt.2024.1356827

Förster, F., Holthaus, P., Dondrup, C., Fischer, J., Romeo, M., and Wood, L. (2022). WTF 2022: Working with Trouble and Failures in conversation between humans and robots. https://sites.google.com/view/wtfworkshop2022/overview

Förster, F., Romeo, M., Holthaus, P., Trigo, M. J. G., Fischer, J. E., Nesset, B., et al. (2023a). Working with Trouble and Failures in Conversation between Humans and Robots (WTF 2023) & Is CUI Design Ready Yet? https://arxiv.org/abs/2401.04108

Förster, F., Romeo, M., Holthaus, P., Wood, L. J., Dondrup, C., Fischer, J. E., et al. (2023b). Working with troubles and failures in conversation between humans and robots: workshop report. *Frontiers in Robotics and AI* 10. doi:10.3389/frobt.2023.1202306

Galbraith, M. (2024). An analysis of dialogue repair in virtual assistants. *Frontiers in Robotics and AI* 11. doi:10.3389/frobt.2024.1356847

Lopez, A., Liesenfeld, A., and Dingemanse, M. (2022). Evaluation of Automatic Speech Recognition for Conversational Speech in Dutch, English and German: What Goes Missing? In *Proceedings of the 18th Conference on Natural Language Processing (KONVENS 2022)*. 135–143

Tisserand, L., Stephenson, B., Baldauf-Quilliatre, H., Lefort, M., and Armetta, F. (2024). Unraveling the thread: understanding and addressing sequential failures in human-robot interaction. *Frontiers in Robotics and AI* 11, 1359782