

"A Stubborn Child" - How Robot Sounds are Oriented to in Everyday Situated Interaction at Home

Hannah R. M. Pelikan
hannah.pelikan@liu.se
Linköping University
Linköping, Sweden

ABSTRACT

Humans make sense of robot actions in the situated context that these actions occur in. This paper takes a conversation analytic approach in studying how the social robot Cozmo is received in a family home, focusing on the non-lexical sounds that the robot uses to communicate. Preliminary findings suggest that participants treat the robot similar to a young child or pet and orient to the robot's sounds in the local context of the interaction.

CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in HCI**; **Empirical studies in collaborative and social computing**.

KEYWORDS

social robotics, situated interaction, conversation analysis, robot sounds, non-lexical sounds

1 INTRODUCTION

In the past 12 months, three social robotics companies went bankrupt, and social home robots Kuri, Jibo and Cozmo are no longer produced. Engineers and computer scientists concluded that the public simply was not ready for social home robots, i.e. entertaining companion robots for the home. This overlooks however, that we still have little insights into how people interpret the robots' actions moment-by-moment in the situated interaction, particularly in family homes. The reason for the failure of these social robots may be hard or even impossible to uncover in lab studies and user interviews, which do not take into account the fine-grained patterns that humans follow when interacting with each other.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

MuC'19 Workshops, Hamburg, Deutschland

© Proceedings of the Mensch und Computer 2019 Workshop on Interacting with Robots and Virtual Agents. Copyright held by the owner/author(s).
<https://doi.org/10.18420/muc2019-ws-659>

Designed to be cute and targeting users' emotions, social home robots produce a variety of sounds alongside, or as a replacement for lexical speech. Cozmo for instance communicates almost exclusively through beeps, motor sounds and human- and animal-inspired vocalizations such as "wow" or "quack" (the only lexical items it utters are users' names and it can read out short texts typed by the user when in programming mode). A small body of human-robot interaction work has studied robot sounds. Comparing robot sounds to no acoustic signal, people felt more comfortable when being approached by a robot that produces beeps, preferably with rising intonation [1]. Research rooted in computer science showed that robot sounds can convey affect and people can categorize robotic beeps into representing different emotions [5]. Interestingly, Read and Belpaeme also found that the interpretation of robot sounds strongly depends on the context of the situation [4]. In their lab study, participants interpreted the same sound differently when it was played immediately after the robot was e.g. hit vs. kissed. This finding points to the importance of studying not only speech, but also non-lexical sounds in the specific sequential contexts they occur in, out 'in the wild', in this case in peoples' homes.

Taking a conversation analysis perspective, my research investigates how families interact with a Cozmo robot at home. Drawing on insights from a running study, this paper gives first answers to the questions *How do users engage with a robot that communicates almost exclusively through non-lexical sound?* and *How do participants orient to non-lexical sounds in the local context in which these sounds are produced?*

2 STUDY SETUP

To study how people naturally interact with a social home robot, I am collecting video data of people playing with a Cozmo robot in their homes. I explored the field during a pilot study with 4 pairs of German-speaking adults, each of the pairs interacting with the robot for 10-30 minutes. Currently, I am collecting longitudinal data in Swedish family homes (2 families so far, a visit to a third family is planned). Parents sign informed consent for themselves and their children before the robot and video cameras are switched on. Participants are videotaped by the researcher when she is introducing and removing the robot and record themselves

with a simple camcorder on a tripod when interacting with the robot during the days in between (at least one week).

Participants consist of families with at least one child in the age of 8-14, as the robot is recommended as a toy for children in this age. The entire family is encouraged to interact with the robot, and the data collected by the families themselves often involves more than one participant interacting with the robot, including siblings and parents.

The Robot

Cozmo is a small toy robot inspired by Pixar's Wall-E. The robot does not typically speak but interacts through sounds, movements and its animated eyes. Cozmo recognizes faces and motion and displays different emotions (happiness, sadness, anger, etc.). Participants control the robot through the Cozmo smartphone app, which allows different modes such as letting the robot roam freely, teach it a person's name and face, play games or to make it carry out simple programs.

3 DATA AND ANALYTIC METHODOLOGY

The video data collected so far consist of 7 hours of video (4h recorded by researcher, 3h by participants). Data are analyzed taking a conversation analytic approach, and the analysis focuses on participants' understandings as these are displayed in situated interactive contexts. Body and voice are often tightly coordinated in the production of non-lexical sounds in human interaction [2] and participants also seem to orient to robot sounds as coordinated with the robot's body. Therefore, all robot sounds are closely scrutinized together with robot movements in the sequential context. In addition, to include the designer's reasoning in creating the sounds, Cozmo's sound designer was interviewed for 45 minutes.

4 PRELIMINARY FINDINGS

Engaging with a social home robot that communicates through non-lexical sounds, users conceptualize the robot in interesting ways. Participants may categorize the robot as a "stubborn child" (pilot study, see Figure 1, line 9) or as their "buddy" (child in family 2). When interacting with the robot, participants employ simple sentence constructions and sometimes imitate the robot's sounds in their own utterances. Both children and adults seem to test the robot's capabilities in the beginning, asking the robot to do things such as lifting a cube. This way of interacting with the robot resembles interaction with small children and pets (see e.g. [3]). Interestingly, not all sounds get oriented to in the same way. While motor sounds are frequently ignored and talked over, sounds with a strong affective component (such as the animation for sadness in Figure 1, l. 5-7) are almost always responded to. For instance, after an offer to play with the touch-responsive cubes ("boxes") that come with the robot (l. 2-4), a sound with falling intonation in combination with

lowering of the robot's forklift-arms (l. 5-7) is interpreted as a "no" (l. 6) and as a negative response to the offer (l. 6-10).

```

01 COZ •(1.7) dadu?
    coz •forklift up-->
02 HUS (1.6) I thi•nk he wants to play with his bo•[xes]
    coz ----->•drive forward forklift(fl) down•fl up-->
03 COZ [kwa↑ke]
04
    hus +(1.0)•(0.7)+(0.8)•
    coz +put box on fl+
    im ----->•
    im #image
05 COZ •uJa• Jwa::
    coz •-l-• 1)fl down
06 HUS (0.4) N•O?•
    coz •turn away-->
07 COZ =waJw•aJwaJwao::
    coz ---->•
08 HUS o:h nQ
09 WIF hm m:
10 HUS (0.3) that's not what it was (what he wanted)
11 COZ •(0.3)• kekekwā:
    coz ---2--• 2) bang forklift on table
12 HUS (0.4) Jm:
13 WIF J a stubborn child J

```



Figure 1: Husband HUS and wife WIF playing with Cozmo COZ and its cubes (transcript translated from German).

5 CONCLUSION

Taking an interactional and praxeological approach to the study of social home robots yields insights into how users orient to the non-lexical sounds that a robot produces on a moment-by-moment basis. The approach uncovers how sounds are interpreted in the situated context of previous turns and the robot's embodied conduct.

ACKNOWLEDGMENTS

This work is funded by the Swedish Vetenskapsrådet, project no. 2016-00827.

REFERENCES

- [1] Kerstin Fischer, Lars Christian Jensen, and Leon Bodenhagen. 2014. To Beep or Not to Beep Is Not the Whole Question. In *Social Robotics*, Michael Beetz, Benjamin Johnston, and Mary-Anne Williams (Eds.). Springer International Publishing, Cham, 156–165.
- [2] Leelo Keevalik. 2018. What Does Embodied Interaction Tell Us About Grammar? *Research on Language and Social Interaction* 51, 1 (Jan 2018), 1–21. <https://doi.org/10.1080/08351813.2018.1413887>
- [3] Robert W. Mitchell. 2001. Americans' Talk to Dogs: Similarities and Differences With Talk to Infants. *Research on Language and Social Interaction* 34, 2 (2001), 183–210. https://doi.org/10.1207/S15327973RLSI34-2_2
- [4] Robin Read and Tony Belpaeme. 2014. Situational Context Directs How People Affectively Interpret Robotic Non-Linguistic Utterances. In *2014 9th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. 41–48.
- [5] Robin Read and Tony Belpaeme. 2016. People Interpret Robotic Non-linguistic Utterances Categorically. *International Journal of Social Robotics* 8, 1 (01 Jan 2016), 31–50. <https://doi.org/10.1007/s12369-015-0304-0>