How Task Complexity and Voice Type Shape Prosodic and Physiologic Cues of Engagement in Human–Machine Dialogue

Océane Granier
Aix-Marseille Univ, CNRS, LPL
Aix-en-Provence, France
oceane.granier@univ-amu.fr

Roxane Bertrand
Aix-Marseille Univ, CNRS, LPL
Aix-en-Provence, France

Kévin Gravouil
Airudit
Bordeaux, France

Laurent Prévot
Aix-Marseille Univ, CNRS, LPL
Aix-en-Provence, France

Abstract

This study investigates prosodic cues of user engagement in task-oriented interactions with non embodied conversational assistants. We hypothesize that task complexity and the type of assistant voice (synthetic vs. human) influence user engagement. We measure both vocal and physiological parameters expected to correlate with engagement. We focus on prosodic features such as fundamental frequency, speech rate and intensity, and also explore physiological indicators, including electrodermal activity and heart rate. While we did not observe significant variations in physiological parameters, our results suggest that both voice type and task complexity influence prosodic markers of engagement.

1 Introduction

Agents used in a professional context —particularly in industrial environments— have to meet additional constraints. A non-embodied artificial agent is preferred (Feng et al., 2020), as it eliminates visual distractions and allows the user to focus on their primary task. Consequently, the type of artificial agent most commonly used in this context is a conversational assistant (CA). The use of CAs enhances tool efficiency and reduces users' cognitive load (Quigley et al., 2004). Despite these advantages, professionals remain reluctant to use CA. Whether due to fear of being replaced by machines or a rejection of recent technologies, they do not use CA in the long term (Cai et al., 2022). To assess this acceptability, we hypothesise that making a machine more engaging could foster the relationship between user and machine.

2 Background

In this study we define engagement as the degree of sustained and goal-directed attention between two interactants over the course of an interaction Sidner and Dzikovska (2002). Engagement can be modulated according to different parameters, such as the type of task we perform, which modifies our level of interest (Berger et al., 2023; Peters et al., 2005). Engagement is optimal when skills match the level of difficulty. A difficulty level perceived as too low leads to boredom (Chanel et al., 2008; Kawada et al., 2023; Scherer, 2003; Westgate, 2020) wich occurs when one is under-stimulated. The voice of our interlocutor may also affect us, especially in the case of a CA that lacks a physical embodiment (Tolmeijer et al., 2021; Éva Székely et al., 2023). Human voices are traditionally preferred over artificial ones and are therefore perceived as more engaging. Jansen (2019) shows that the more an entity resembles a human, the greater our affinity. However, when this resemblance reaches a certain threshold, affinity drops sharply. Uncanny valley is the expression used to describe the feeling of strangeness experienced at that time (Jansen, 2019). Our study is set in a context where the CA is not at the center of the interaction but serves to assist the user in their professional task. In addition to being the most logical choice in an interaction with a voice-based system, prosodic parameters had, to our knowledge, never been studied in human-machine interaction. Physiologic parameters, wich have already been studied in both human-human and human-machine interactions(Perugia et al., 2017; Monkaresi et al., 2016; Rani and Sarkar, 2005), unlike prosodic parameters, will allow us to confirm our experimental measurements.

We investigated whether the type of voice and the complexity of the task influenced participants' engagement by studying the prosodic cues in their voices. We hypothesize that using a CA with a human voice in an industrial setting would lead it to fall into the uncanny valley, due to a mismatch between the CA's vocal capabilities and the associated robotic tools. We argue that participants should be more engaged when interacting with the

Proceedings of the 29th Workshop on the Semantics and Pragmatics of Dialogue, September 3–5, 2025, Bielefeld, Germany.

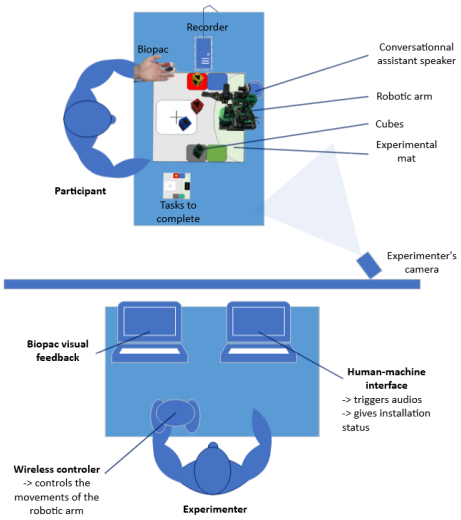

Figure 1: Experimental Set-up

artificial voice rather than the human voice. The second hypothesis is that participants will be more engaged in the interaction if the task complexity is higher, because a task that is too easy may lead to understimulation.

3 Method

The experiment consists of an interaction in French between fluent participants and a robotic arm accompanied by a CA, as illustrated in Figure 1. The arm is operated remotely using a wireless video game controller. The CA is implemented using a Wizard-of-Oz paradigm, meaning that its utterances are triggered remotely to simulate a smooth and natural interaction. Participants are instructed to use their voice to direct the robotic arm to move colored cubes into different designated areas. They are required to complete objective cards by providing the correct movement instructions to the robotic arm.

In order to present more or less engaging conditions to our participants, we defined two levels of task complexity: an easy task, which is supposed to be unstimulating and therefore boring, and a difficult task, which is supposed to be stimulating and therefore engaging. The complexity of the task is adjusted by the complexity of the objective cards. We also tested two different female voices for our CA: a human voice and an artificial voice. Each participant will attend 4 interaction sessions (2 complexity X 2 voices) with the robotic arm accompanied by a CA.

Participants Thirty-three participants (24 women and 9 men) were recruited. Each participant received a compensation of €15 for a 90-minute session. Five sessions were excluded from the analysis: one due to improper application of the protocol, four due to faulty recordings.

Participant equipment Participants are equipped with various Biopac measurement tools: an abdominal belt placed below the chest to record respiration; electrodes attached to the second phalanges of the index and middle fingers on the non-dominant hand to measure EDA ; and a photoplethysmograph (PPG) on the same hand to record heart rate. A headset microphone connected to a Zoom H4n Pro is also used to record participants' speech.

4 Results

No significant differences were observed in the physiological measures (t-test and ANOVA) neither by voice type or task complexity. Concerning the prosodic parameters, results show sevreal significant differences. The standard deviation of intensity was significantly ($p<0.05$) higher for the human voice ($\sim$12.06) compared to the artificial voice ($\sim$11.45). Speech rate was faster for the artificial voice, with a rate of around 3.54 syllables per second (SD: $\sim$0.59), compared to around 3.32 syllables per second (SD: $\sim$0.46) for the human voice ($p<0.03$). The speech rate was significatively ($p<0.02$) slower for the difficult task ($\sim$3.30 syllables/second) compared to the easy task ($\sim$3.55 syllables/second). There was a fairly substantial session order effect for EDA.

5 Conclusion

The aim of this study is to find engagement cues in the voice of a CA user. To this end, we selected the user's physiological cues correlated with engagement in human-machine interactions and the prosodic cues of the user's voice correlated with engagement in human-human interactions. We compared these parameters as a function of the CA's voice and the difficulty of the task to be performed. While we did not observe significant variations in physiological parameters, our results suggest that both voice type and task complexity influence prosodic markers of engagement.

References

Jonah Berger, Wendy W Moe, and David A Schweidel. 2023. What holds attention? linguistic drivers of engagement. *Journal of Marketing*, page 00222429231152880.

Danting Cai, Hengyun Li, and Rob Law. 2022. Anthropomorphism and ota chatbot adoption: a mixed methods study. *Journal of Travel & Tourism Marketing*, 39(2):228–255.

Guillaume Chanel, Cyril Rebetez, Mireille Bétrancourt, and Thierry Pun. 2008. Boredom, engagement and anxiety as indicators for adaptation to difficulty in games. In *Proceedings of the 12th international conference on Entertainment and media in the ubiquitous era*, pages 13–17.

Shengjia Feng, Peter Buxmann, et al. 2020. My virtual colleague: A state-of-the-art analysis of conversational agents for the workplace. In *HICSS*, pages 1–10.

Dennis Jansen. 2019. Discovering the uncanny valley for the sound of a voice. *Unpublished master's thesis]. School of Humanities and Digital Sciences Department of Cognitive Science & Artificial Intelligence. Tilburg*.

Michiko Kawada, Akihito Shimazu, Daisuke Miyanaka, Masahito Tokita, Keiko Sakakibara, Naana Mori, Fuad Hamsyah, Lin Yuheng, Kojiro Shojima, and Wilmar B Schaufeli. 2023. Boredom and engagement at work: do they have different antecedents and consequences? *Industrial health*.

Hamed Monkaresi, Nigel Bosch, Rafael A Calvo, and Sidney K D'Mello. 2016. Automated detection of engagement using video-based estimation of facial expressions and heart rate. *IEEE Transactions on Affective Computing*, 8(1):15–28.

Giulia Perugia, Daniel Rodríguez-Martín, Marta Díaz Boladeras, Andreu Català Mallofré, Emilia Barakova, and Matthias Rauterberg. 2017. Electrodermal activity: Explorations in the psychophysiology of engagement with social robots in dementia. In *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pages 1248–1254. IEEE.

Christopher Peters, Catherine Pelachaud, Elisabetta Bevacqua, Maurizio Mancini, Isabella Poggi, and Universita Roma Tre. 2005. Engagement capabilities for ecas. In *AAMAS'05 workshop Creating Bonds with ECAs*.

Morgan Quigley, Michael A Goodrich, and Randal W Beard. 2004. Semi-autonomous human-uav interfaces for fixed-wing mini-uavs. In *2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)(IEEE Cat. No. 04CH37566)*, volume 3, pages 2457–2462. IEEE.

Pramila Rani and Nilanjan Sarkar. 2005. Operator engagement detection and robot behavior adaptation in human-robot interaction. In *Proceedings of the 2005 IEEE International Conference on Robotics and Automation*, pages 2051–2056. IEEE.

Klaus R Scherer. 2003. Vocal communication of emotion: A review of research paradigms. *Speech communication*, 40(1-2):227–256.

Candace Sidner and Myrosia Dzikovska. 2002. Human-robot interaction: Engagement between humans and robots for hosting activities. In *Proceedings. Fourth IEEE International Conference on Multimodal Interfaces*, pages 123–128. IEEE.

Suzanne Tolmeijer, Naim Zierau, Andreas Janson, Jalil Sebastian Wahdatehagh, Jan Marco Marco Leimeister, and Abraham Bernstein. 2021. Female by default?–exploring the effect of voice assistant gender and pitch on trait and trust attribution. In *Extended abstracts of the 2021 CHI conference on human factors in computing systems*, pages 1–7.

Erin C Westgate. 2020. Why boredom is interesting. *Current Directions in Psychological Science*, 29(1):33–40.

Éva Székely, Joakim Gustafson, and Ilaria Torre. 2023. Prosody-controllable gender-ambiguous speech synthesis: A tool for investigating implicit bias in speech perception. *Interspeech*.