Reacting to the last laugh: Probing non-humorous laughter with Furhat robot

Eleni Giannitzi, Vladislav Maraev, Christine Howes

Department of Linguistics, Philosophy and Theory of Science, University of Gothenburg, Sweden

gusgianel@student.gu.se, vladislav.maraev@gu.se, christine.howes@gu.se

Abstract

This paper explores how laughter can be added into a negotiative dialogue with a Furhat, the socially interactive agent (SIA). We present a proof-of-concept method of manipulating robot responses generated by a large language model through the interventions of a human operator. We conduct a pilot experiment focussing on people's perception of Furhat in case of two intervention types: laughter reciprocation and laughter clarification requests.

Index Terms: laughter, human-computer interaction, social robotics

1. Introduction

In recent years, the domain of human-machine interaction has undergone significant advancements, particularly with the development of socially interactive agents (SIAs). While the primary focus remains on dialogue improvement, there is a noticeable shift towards incorporating multimodality in both 2D and 3D SIAs to emulate more realistic, human-like interaction. Notably, facial expressions, body language, and hand gestures have emerged as critical components in achieving a heightened level of authenticity.

Current SIAs are typically task focused, and do not make full use of non-verbal cues from their interlocutor, e.g. gaze, which can be a turn-taking cue, or rapid blinking whilst processing information. In addition, social robots often exhibit unrealistic non-verbal behaviour, such as maintaining eye contact with a user, which can contribute to feelings of uneasiness, awkwardness, and/or embarrassment. Laughter behaviour is also typically unrelated to human expectations. Almost all studies of SIAs consider laughter as pleasant feedback (e.g. in reaction to a joke), whereas in human communication laughter is also used to manage social aspects of interaction. This study presents a first step in extending social laughter to SIAs.

In recent years dialogue systems are increasingly implemented using generative AI based on large language models (LLMs). However, because LLMs are trained on mostly text data that is very different from spoken dialogue, some human behaviours might not be available in the models.[1] This is particularly problematic if used to underpin a social robot, which should express emotions and communicative intentions through the visual modality (gaze and facial gestures).

In this paper we focus on accommodating user laughter into a negotiative dialogue with a social robot. We do so by employing a intervention technique, where the experimenter who observes the interaction can modify the behaviour of the robot in real time. Our contributions are:

- A proof-of-concept method where LLM-generated robot responses can be manipulated in real time by a human operator.
- A pilot experiment in which we analyse how laughter reciprocation and laughter clarification requests (CRs) can alter people's perception of a SIA.

2. Related work

Mazzocconi *et al.* [2] present a taxonomy of laughter functions from the perspective of the intended effect on the context, distinguishing between e.g. pleasant laughter (perhaps in response to a joke) and social laughter (such as smoothing embarrassment). Laughter is well known to have important social effects, being crucial for bonding and managing relationships, while also being immensely influenced by social context [3].

Several models have been proposed to generate laughter and to decide when a SIA should laugh. For example, Ding *et al.* [4] created a laughter behaviour controller to generate face and body motions from laughter audio. Haddad *et al.* [5] created a listening agent that predicts when to smile or to laugh depending on its interlocutor's behaviours.

Several recent studies have investigated laughter synthesis [6]–[8]. Despite this, laughter synthesis in different contexts serving different pragmatic functions not been closely examined. For instance, how can social laughter which accompanies an apology be synthesised with speech in a naturalistic way? Laughter detection is a more developed topic compared to synthesis. State-of-the-art laughter detection is based on machine learning techniques: from support vector machines to deep learning approaches [9]–[11].

Mazzocconi *et al.* [12] differentiate two dimensions of laughter meaning: the argument it is predicating (the *laughable*) and the level of arousal. In a corpus study, they show how each of these dimensions can be a the object of a CR. For instance, a question "What's funny?" assumes funniness – intuitive understanding of laughter, whereas "What you laughing at?" can mean a more general question about the laughable.

A number of studies have introduced interventions in text-based chat in relation to laughter use. Mills *et al.* [13] assessed laughter mimicry and the interrelation between laughter and emotional contagion. Maraev *et al.* [14] inserted spoof contributions such as additional laughs and CRs ("lol?", "lol what", "what's funny" etc.) which appeared to come from the dialogue participants in online text-chat. This study extends this real-time intervention technique to interactions with a SIA.

[1] Similarly, Liesenfeld *et al.* [1] showed that this is true for commercial speech recognition systems which lack in support of features prevalent in human conversation.

3. Methods and materials

3.1. Dialogue design

A dialogue management back-end which interacts with a Furhat robot [15] via Furhat Remote API was implemented in TypeScript using XState[2] library which employs the *statecharts* formalism [16] for designing interactive systems.

The initial phase involved crafting facial expressions and movements essential for interaction, including synchronising head and mouth movements indicative of laughter, and nuanced adjustments in the eyes, nose, and eyebrows. For the inserted laughter, we used a sound sample representing a slight chuckle, lasting one second (0.01) each.

Furhat's speech was generated by ChatGPT4 (see Figure 2 for example prompt). The topics for negotiation were: movies versus books, indoor versus outdoor activities, and selecting a gift for a friend. ChatGPT4 responses were limited to 50 tokens to maintain control, and the transcript of the dialogue so far was fed back into the system to minimise the likelihood of repetition. Furhat initiated each interaction by introducing itself and asking the user's name.

3.2. The experiment

The experiment included 14 participants (5 males, 6 females, and 3 "other/prefer not to say"), aged between 20 and 45, who were instructed to engage in negotiation with Furhat. Participants were assigned to one of three experimental conditions:

- **Laugh CR:** Responds to user's laughter with a CR such as "Why are you laughing?" or "Did I say something funny?".
- **Laugh back:** Responds to user's laughter with laughter.
- **Control:** No interactive responses initiated.

For the experimental conditions, the experimenter triggered the response when user laughter was observed. Additionally, for all three groups, Furhat produced three random laughs per conversation topic.

Following the experiment, participants completed a questionnaire divided into two sections. The first, inspired by Inoue *et al.* [17], evaluates the perceived naturalness, understanding, human-likeness, and empathy of the robot on a scale from 1 to 10. The second, based on Becker-Asano *et al.* [18], made use of the Geneva Wheel of Emotions (GWE) [19] allowing us to view the perceived emotions from the participants' points of view towards Furhat, throughout the entire experiment. The GWE consists of 20 emotions with five intensity levels (0 to 5).

4. Results and Discussion

Table 1 shows the questionnaire results. Due to the small sample size of data, we managed to find only a few trends, namely, the higher level of naturalness and human-likeness in comparison to control group in both intervention conditions.

Qualitative analysis of some of the interventions shows that the laughter CRs, e.g., after line 4 in excerpt (a) are taken to be confrontational. In the example shown below, we further note the particpant's self-awareness of the fact that her laughter was in fact a social laughter, reciprocating one of Furhat's randomly generated laughs. The laugh back condition was rather more subtle, with a potential issue being the delay between the user laughter and Furhat's laughter, since ChatGPT is not incremental, and the laughter was typically inserted after the speech turn in progress when it was triggered.

[2] https://stately.ai/docs/xstate

Group type	N	Empathy (1-10)	Naturalness (1-10)	Human-likeness (1-10)	Understanding (1-10)	Interest (1-5)	Amusement (1-5)	Joy (1-5)	Admiration (1-5)
Control	4	4.25	3.25	3.25	4.50	3.25	3.00	2.75	1.75
Laugh back	5	4.40	5.00	5.40	5.20	3.80	3.80	3.00	2.40
Laugh CR	5	4.40	4.40	4.20	6.00	3.60	3.60	2.60	1.60
Total	14	4.36	4.29	4.36	5.29	3.57	3.50	2.79	1.93

Table 1: *Questionnaire results by intervention type. Only four emotions with higher intensity levels are shown.*

(a)
1 **Participant** [...] there is a film festival in Gothenburg recently
2 **Furhat** [<laughter>][3]
3 **Participant** [yeah] <laughter>
4 **Furhat** yes, the Gothenburg film festival is an annual event in Gothenburg showcasing the international and domestic films, fostering cinema culture.
5 **Furhat** *(human-mediated turn)* Did you just laugh at me?
6 **Participant** No no no, I just want to make the atmosphere more ... uhm [friendly]
7 **Furhat** [Oh I get it now] Allow me to repeat myself. Yes, the Gothenburg film festival ...

We plan to build upon this study, using scripted dialogues (to avoid issues like ChatGPT's repetitiveness), focusing on practical and collaborative tasks such as following a cooking recipe. Emphasising the significance of gaze coordination in providing contextual cues during such interactions, we will explore how specific gaze patterns aligned with non-humorous laughter [20] influence people's perception of human-machine interaction. More specifically, we plan to study how laughter is perceived depending on the accompanying gaze patterns. According to Becker-Asano *et al.* [18], laughter can also be classified into two main categories: aversive and friendly. In particular, in their experiments, the human-like robot's direct gaze at the participant while laughing led to the perception of the robot's laughter as "laughing at someone" rather than "laughing with someone", which deems Furhat's default mode of gaze-following inappropriate, at least for some laughter types.

The current study is limited by using only one sample for laughter. Even though it wasn't perceived negatively by the participants, it is apparent that the variety of laughter forms is needed. Despite the data on the absence of significant correlation between pragmatic types of low-arousal laughs and their acoustic features [21], it is intriguing to study how different realisations of such laughs are perceived in SIAs. We are planning to use the method of transfering the time-frequency representations of laughter acoustic signal into the spectrotemporal modulation domain previously employed to evaluate the differences between mimicking and non-mimicking laughter [22], [23].

Overall, our intervention method may help understand how SIA's laughter affects human-robot interaction and the interpretation of laughter's pragmatic functions.

[3] Overlapped speech is marked by square brackets.

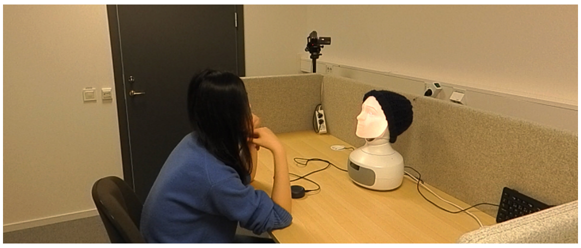

Figure 1: *Experiment setting*

Topic: ${topic name}
Task: Continue the debate by giving an OPPOSITE statement from this following argument (ANSWER WITH ONLY 50 TOKENS)
${argument}

Figure 2: *Example prompt*

5. Acknowledgements

This work was supported by two grants from the Swedish Research Council (VR); 2023-00358 – Social laughter for virtual agents (SocLaVA) and 2014-39 for the establishment of the Centre for Linguistic Theory and Studies in Probability (CLASP) at the University of Gothenburg. It was also supported by ERC Starting Grant 101077927; Divergence and convergence in dialogue: The dynamic management of mismatches (DivCon).

6. References

[1] A. Liesenfeld, A. Lopez, and M. Dingemanse, "The timing bottleneck: Why timing and overlap are mission-critical for conversational user interfaces, speech recognition and dialogue systems," in *Proceedings of the 24th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, S. Stoyanchev, S. Joty, D. Schlangen, O. Dusek, C. Kennington, and M. Alikhani, Eds., Prague, Czechia: Association for Computational Linguistics, 2023, pp. 482–495.

[2] C. Mazzocconi, Y. Tian, and J. Ginzburg, "What's your laughter doing there? a taxonomy of the pragmatic functions of laughter," *IEEE Transactions on Affective Computing*, pp. 1–1, 2020.

[3] R. R. Provine, "Laughing, tickling, and the evolution of speech and self," *Current Directions in Psychological Science*, vol. 13, no. 6, pp. 215–218, 2004.

[4] Y. Ding, J. Huang, and C. Pelachaud, "Audio-driven laughter behavior controller," *IEEE Trans. on Affect. Comp.*, vol. 8, no. 4, pp. 546–558, 2017.

[5] K. E. Haddad, H. Çakmak, E. Gilmartin, S. Dupont, and T. Dutoit, "Towards a listening agent: A system generating audiovisual laughs and smiles to show interest," in *Proceedings of the 18th ACM International Conference on Multimodal Interaction*, ACM, 2016, pp. 248–255.

[6] T. Nagata and H. Mori, "Defining laughter context for laughter synthesis with spontaneous speech corpus," *IEEE Trans. Affect. Comput.*, vol. 11, no. 3, pp. 553–559, 2020.

[7] H. Luong and J. Yamagishi, "LaughNet: Synthesizing laughter utterances from waveform silhouettes and a single laughter example," *CoRR*, vol. abs/2110.04946, 2021.

[8] N. Kanda, X. Wang, S. E. Eskimez, M. Thakker, H. Yang, Z. Zhu, M. Tang, C. Li, C.-H. Tsai, Z. Xiao, Y. Xia, J. Li, Y. Liu, S. Zhao, and M. Zeng, *Making flow-matching-based zero-shot text-to-speech laugh as you like*, 2024.

[9] J. Gillick, W. Deng, K. Ryokai, and D. Bamman, "Robust Laughter Detection in Noisy Environments," in *Proc. Interspeech 2021*, 2021, pp. 2481–2485.

[10] G. Rennie, O. Perepelkina, and A. Vinciarelli, "Which model is best: Comparing methods and metrics for automatic laughter detection in a naturalistic conversational dataset," in *Interspeech 2022*, ISCA, Sep. 2022.

[11] B. Ludusan, "The usefulness of phonetically-motivated features for automatic laughter detection," in *Disfluency in Spontaneous Speech (DiSS) Workshop 2023*, ISCA, Aug. 2023.

[12] C. Mazzocconi, V. Maraev, and J. Ginzburg, "Laughter repair," in *Proceedings of the 22nd Workshop on the Semantics and Pragmatics of Dialogue - Full Papers*, Aix-en-Provence, France: SEMDIAL, Nov. 2018.

[13] G. Mills, E. Gregoromichelaki, C. Howes, and V. Maraev, "Influencing laughter with ai-mediated communication," *Interaction Studies*, vol. 22, no. 3, pp. 416–463, 2021.

[14] V. Maraev, C. Mazzocconi, G. Mills, and C. Howes, "LOL what?: Empirical study of laughter in chat based dialogues," in *Proceedings of Laughter Workshop 2020*, 2020, pp. 55–58.

[15] S. A. Moubayed, J. Beskow, G. Skantze, and B. Granström, "Furhat: A back-projected human-like robot head for multiparty human-machine interaction," in *Cognitive Behavioural Systems: COST 2102 International Training School, Dresden, Germany, February 21-26, 2011, Revised Selected Papers*, Springer, 2012, pp. 114–130.

[16] D. Harel, "Statecharts: A visual formalism for complex systems," *Sci. Comput. Program.*, vol. 8, no. 3, pp. 231–274, 1987.

[17] K. Inoue, D. Lala, and T. Kawahara, "Can a robot laugh with you?: Shared laughter generation for empathetic spoken dialogue," *Frontiers in Robotics and AI*, vol. 9, 2022.

[18] C. Becker-Asano and H. Ishiguro, "Laughter in social robotics-no laughing matter," in *Intl. Workshop on Social Intelligence Design*, 2009, pp. 287–300.

[19] K. R. Scherer, "What are emotions? and how can they be measured?" *Social science information*, vol. 44, no. 4, pp. 695–729, 2005.

[20] C. Mazzocconi, V. Maraev, V. Somashekarappa, and C. Howes, "Looking for laughs: Gaze interaction with laughter pragmatics and coordination," in *Proceedings of the 2021 International Conference on Multimodal Interaction*, ser. ICMI '21, Montréal, QC, Canada: Association for Computing Machinery, 2021, pp. 636–644.

[21] C. Mazzocconi, "Laughter in interaction: Semantics, pragmatics and child development," Ph.D. dissertation, Université de Paris, 2019.

[22] C. Mazzocconi, B. O'Brien, K. E. Haddad, K. Bodur, and A. Fourtassi, "Differences between mimicking and non-mimicking laughter in child-caregiver conversation: A distributional and acoustic analysis," in *Proceedings of the Annual Meeting of the Cognitive Science Society*, vol. 45, 2023.

[23] C. Mazzocconi, B. O'Brien, and T. Chaminade, "How do you laugh in an fmri scanner? laughter distribution, mimicry and acoustic analysis," in *Disfluency in Spontaneous Speech (DiSS) Workshop 2023*, 2023.