2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
October 25-29, 2020, Las Vegas, NV, USA (Virtual)

Read the Room, Robot! Exploring Audiovisual Methods to Improve the Effectiveness of Robotic Comedians

Carson Gray[1], Paris Myers[1], and Naomi T. Fitter[1]

***Abstract*— Social robots with the ability to read the room and work the crowd hold potential for more effective and engaging interactions with people, but such human performer-inspired abilities for robots are currently limited. We propose that audiovisual data and comedian-inspired decision-making can enhance the performance of social robots in settings from live performance to healthcare. The present paper describes a preliminary robotic system for comedy performance, which includes integration with OpenFace facial behavior analysis software for real-time detection of Facial Action Units and gaze information. In parallel to the robotic system creation and initial validation, we are planning to survey human comedy performers to learn more about the audience cues they use to assess joke success and decide on performance adaptations. Preliminary results show promise for OpenFace to aid with the proposed real-time recognition, and future steps will more thoroughly assess proposed behavior recognition models and robot performance adaptation approaches. The products of this work can inform social roboticists who wish to create more compelling and entertaining systems.**

I. Motivation

Humor is one of the most authentic, raw forms of human communication. In interactions between people, humor is a socially desirable trait and an important factor in physical and metal health [1]. Accordingly, we propose that as robots take on more socially interactive roles, they will need improved abilities to understand and adapt to their surroundings using a playful and comedic lens. Multimodal sensing strategies can inform these adaptations, but the infrastructure for this technique requires substantial design and evaluation effort. Our present work aims to begin incorporating these robotic skills in the comedy performance space, a natural arena for early and experimental robot humor evaluations. Ultimately, we envision these skills to be essential in social robot applications such as service and healthcare.

Although early robot comedian prototypes have successfully performed stand-up [2] and improvisational comedy [3], robot abilities to read the room and work the crowd using multimodal information remains limited. Specific past approaches to help robotic comedians adapt include audio-only laughter detection [4], as well as computer vision-centric audience feedback detection scaffolded with colored “voting paddles” meant to help the viewers convey like or dislike of jokes [5]. The most related past effort to our current work was the laughter audio- and computer vision-informed robotic comedian of Katevas et al. [2]. Compared to these past examples, we plan to use multimodal sensing in a more comprehensive way; for example, to note and capitalize on audience behaviors in the comedy routine, as envisioned in Fig. 1. We will additionally test the resulting robotic comedian in a wide variety and large number of performances. The final products of this work can inform the design of successful social robotic systems.

[1]Collaborative Robotics and Intelligent Systems (CoRIS) Institute, School of Mechanical, Industrial, and Manufacturing Engineering, Oregon State University at Corvallis, OR 97331, USA. Email: {graycar, myerspar, naomi.fitter}@oregonstate.edu

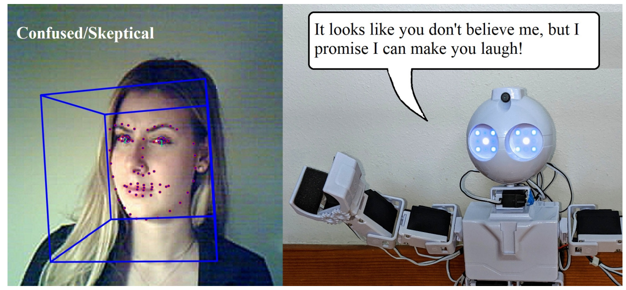

Fig. 1: An example case of a robot reading the room and working the audience. In the proposed interaction, the JD Humanoid identifies a skeptical facial expression and offers a playful tailored response that addresses the user’s demeanor.

II. Methods

To build on past work in robot comedy and advance the state of the field, we selected a proven robotic platform, implemented initial computer vision strategies, and created an experimental protocol for gathering ethnographic information from human comedy performers.

A. System Design

Our proposed robotic comedian system consists of a physically embodied humanoid robot and locally networked computer for sensor data processing.

Robotic Hardware: The robotic comedian in the system is the JD Humanoid by EZ Robot. This small, modular humanoid has built-in cameras and microphones to support the proposed audiovisual sensing.

Mini PC: The robot will be locally networked with a mini PC to allow for fast real-time processing of camera and microphone data. The mini PC will use the OpenFace facial behavior analysis toolkit [6] and the audio processing algorithm included in [4] to generate state estimations of interest during robotic comedy performance.

Facial Expression Analysis: Using the Facial Action Unit (AU) recognition abilities of OpenFace [7], we can build from fundamental psychology research on human emotion to recognize particular facial expressions in real time [8].

In addition to recognizing fundamental human emotions described by past literature (e.g., happiness, surprise, disgust), we propose to use AU and gaze features to train models that can recognize additional behaviors of interest. Specific behaviors will center on human comedian-inspired practices gathered from efforts described in the following subsection.

B. *Ethnographic Data Collection*

Successful work in robotic comedy often arises from human comedian-inspired strategies and collaborations with comedians. For example, one author of [4] is a semi-professional stand-up comedian, and both authors of [3] are experienced improvisers. To capitalize and build on this approach, we plan to gather ethnographic information from human comedians during our interaction design process. This work is approved by the Oregon State University Institutional Review Board under protocol #IRB-2019-0172.

Procedure: We will recruit human comedy performers using performance arts networks, such as social media groups dedicated to stand-up and improvisational comedy. Consenting participants will complete an online survey focused on understanding performer methods for judging joke success and adapting based on crowd responses. Respondents will also be invited to complete a synchronous videoconferencing-based interview with the research team.

Measures: Quantitative and qualitative data are both essential to the design of robotic systems. Accordingly, our measures involve both 1) succinct multiple choice questions to gauge how common particular practices of interest are across comedians and 2) open-ended free-response and interview questions to understand how and why comedians do these things, in addition to catching gaps or omissions in our envisioned robot behaviors.

Analysis: Assessment of response distributions for quantitative multiple choice questions will yield information about what sensing and adaptation strategies are reasonable, in addition to which approaches should be prioritized by our robot. We will use grounded theory methods (similar to the practices in [9]) to understand the open-ended feedback collected from the human performers.

III. Preliminary Results and Future Work

Currently, we have successfully networked out robotic hardware with an external computer and used OpenFace to detect AUs that will be essential to our later efforts. Figure 2 shows example AU outputs from a subset of human behaviors of interest. In the context of a stand-up comedy routine, the serious expression could be used to identify when the audience did not find the joke particularly funny, but were most likely not offended or confused, either. The laughing expression could indicate a positive audience response. Finally, the confused/skeptical facial expression could represent a negative audience response that could be remedied through tailored actions from the robotic comedian.

In future steps, we will use the outputted OpenFace features to train new models that will allow our robotic comedian to recognize key audience behaviors noted by the human comedians. For example, if an audience member nods enthusiastically, the robot may point at that person and say, "This guy knows what I'm talking about!", or if the whole room grimaces at a risky joke, the robot might quip, "What? Too soon?" After completing the design and implementation of robot room-reading and crowd-working behaviors, we plan to conduct an empirical evaluation of differences in audience responses to a robotic comedian with and without multimodal sensing-based adaptations.

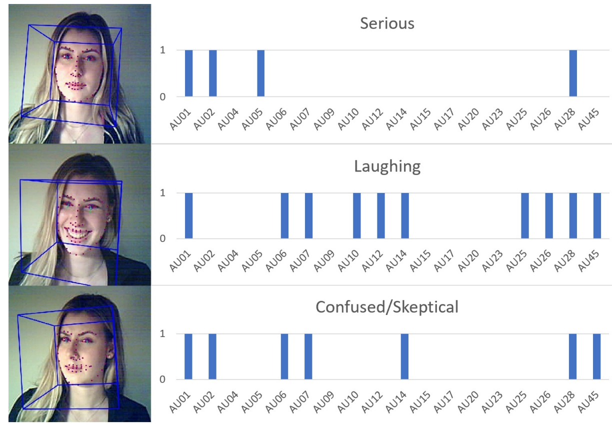

Fig. 2: A subset of our OpenFace-identified AU occurrences for serious, laughing, and confused/skeptical facial expressions. We plan to train models using these AU features and gaze information to identify key audience behaviors.

Key *strengths* of this work include the potential to make social robots more adaptive, effective, and entertaining. At the same time, potential *challenges or limitations* include creating a system that can capably respond to the behavioral cues of diverse system users. We aim to help robots effectively use humor in everyday interactions, beginning in entertainment and eventually extending to service and healthcare.

Acknowledgments

We thank Kory Mathewson and Piotr Mirowski for sharing software tools for the JD Humanoid robot.

References

[1] R. A. Martin and T. Ford, *The psychology of humor: An integrative approach*. Academic Press, 2018.

[2] K. Katevas, P. G. Healey, and M. T. Harris, "Robot comedy lab: Experimenting with the social dynamics of live performance," *Frontiers in Psychology*, vol. 6, p. 1253, 2015.

[3] K. W. Mathewson and P. Mirowski, "Improvised theatre alongside artificial intelligences," in *Proc. of the Thirteenth Artificial Intelligence and Interactive Digital Entertainment Conference*, 2017.

[4] J. Vilk and N. T. Fitter, "Comedians in cafes getting data: Evaluating timing and adaptivity in real-world robot comedy performance," in *Proc. of the ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2020, pp. 223–231.

[5] H. Knight, S. Satkin, V. Ramakrishna, and S. Divvala, "A savvy robot standup comic: Online learning through audience tracking," in *Proc. of TEI'10 (Workshop Paper)*, 2011.

[6] T. Baltrušaitis, A. Zadeh, Y. C. Lim, and L.-P. Morency, "Openface 2.0: Facial behavior analysis toolkit," in *Proc. of the IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, 2018, pp. 59–66.

[7] T. Baltrušaitis, M. Mahmoud, and P. Robinson, "Cross-dataset learning and person-specific normalisation for automatic action unit detection," in *Proc. of the IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, vol. 6, 2015, pp. 1–6.
[8] J. F. Cohn, Z. Ambadar, and P. Ekman, "Observer-based measurement of facial expression with the facial action coding system," *The handbook of emotion elicitation and assessment*, vol. 1, no. 3, pp. 203–221, 2007.
[9] B. Mutlu and J. Forlizzi, "Robots in organizations: The role of workflow, social, and environmental factors in human-robot interaction," in *Proc. of the ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2008, pp. 287–294.