Human Stress Response and Perceived Safety during Encounters with Quadruped Robots

Ryan Gupta[*,1], Hyonyoung Shin[*,1,2], Emily Norman[3], Keri K. Stephens[3], Nanshu Lu[1,2] and Luis Sentis[1]

Abstract— Despite the rise of mobile robot deployments in home and work settings, perceived safety of users and bystanders is understudied in the human-robot interaction (HRI) literature. To address this, we present a study designed to identify elements of a human-robot encounter that correlate with observed stress response. Stress is a key component of perceived safety and is strongly associated with human physiological response. In this study a Boston Dynamics Spot and a Unitree Go1 navigate autonomously through a shared environment occupied by human participants wearing multimodal physiological sensors to track their electrocardiography (ECG) and electrodermal activity (EDA). The encounters are varied through several trials and participants self-rate their stress levels after each encounter. The study resulted in a multidimensional dataset archiving various objective and subjective aspects of a human-robot encounter, containing insights for understanding perceived safety in such encounters. To this end, acute stress responses were decoded from the human participants' ECG and EDA and compared across different human-robot encounter conditions. Statistical analysis of data indicate that on average (1) participants feel more stress during encounters compared to baselines, (2) participants feel more stress encountering multiple robots compared to a single robot and (3) participants stress increases during navigation behavior compared with search behavior.

I. INTRODUCTION

Autonomous mobile robot research has been ongoing for decades for a wide range of real-world scenarios like delivery [1], search [2], surveillance [3] and mapping [4]. As mobile robots proliferate in communities, designers must consider the impacts these systems have on the users, onlookers and places they encounter. It becomes increasingly necessary to study situations where humans and robots coexist in common spaces, even if they are not directly interacting.

Subjective questionnaires have been popularly employed in literature to measure preceived safety, preference and human mental state during HRI [5, 6, 7, 8]. However, questionnaires have poor temporal resolution and are subject to many sources of bias including moderacy response bias and memory effects induced by repetitive administration [9]. Questionnaires also impact HRI, requiring researchers to adopt rapidly administered questionnaires which have limited detail and coverage or, in some cases, limit the choice of robots to be studied to those like the Pepper robot with

*These authors contributed equally to this work.

[1]Department of Aerospace Engineering and Engineering Mechanics, University of Texas at Austin, Austin, TX 78712 USA ryan.gupta@utexas.edu

[2]Department of Electrical and Computer Engineering, University of Texas at Austin, Austin, TX 78712 USA

[3]Department of Communication Studies, University of Texas at Austin, Austin, TX 78712 USA
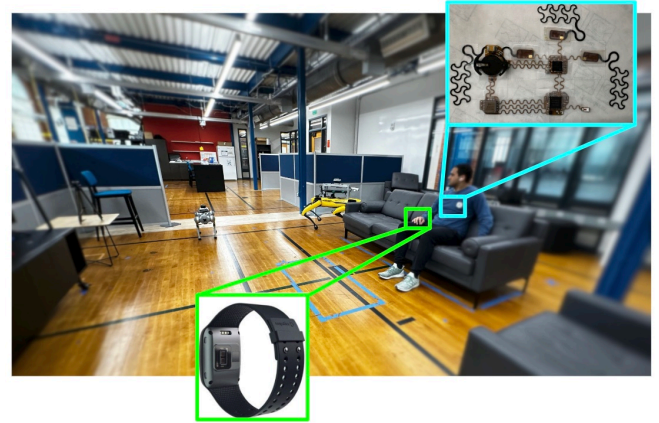
Fig. 1. An overview of the experiments. The Unitree Go1 and Boston Dynamics Spot perform various mobile behaviors in shared spaces with UT community members. Physiological measures are collected from the E4 wristwatch (green) [17] and chest e-tattoo (cyan) [18].

a user interface for questionnaires [10, 11]. To overcome these limitations, physiological measures have also been considered in HRI [12, 13, 14, 15]. Non-invasive physiological sensing enables the tracking of the human internal state in real-time, for example to evaluate stress during HRI [16].

We hypothesize these sensors allow tracking perceived safety during human-robot encounters by decoding acute stress levels from known biomarkers in ECG and EDA data. This is significant in two ways: (1) To the best of our knowledge, this is the first analysis of physiological response to mobile robots navigating freely [11]. Our study deploys two quadruped robots in a realistic apartment setting (Fig. 1). (2) We characterize *human-robot encounters*. This subset of human-robot interactions involve humans and robots coexisting in a shared space with awareness of one another, who are not actively interacting. Human-robot encounters represent situations that are likely to dominate total interaction time if mobile robots are deployed in home or other indoor environments. Our results provide a few insights, based on physiological signal analysis, regarding the elements of human-robot encounters that affect the perceived safety of the robots by humans. These insights can be used in future design where perceived safety is a concern and pave the way for an online pipeline whereby the signals from a user's sensors could be decoded to inform robot behavior in real-time [19, 20].

II. BACKGROUND

A. Stress Response and Comfort in HRI

Perceived safety during encounters with mobile robots remains understudied in many domains [6, 10]. The first group of works we find leverage more objective
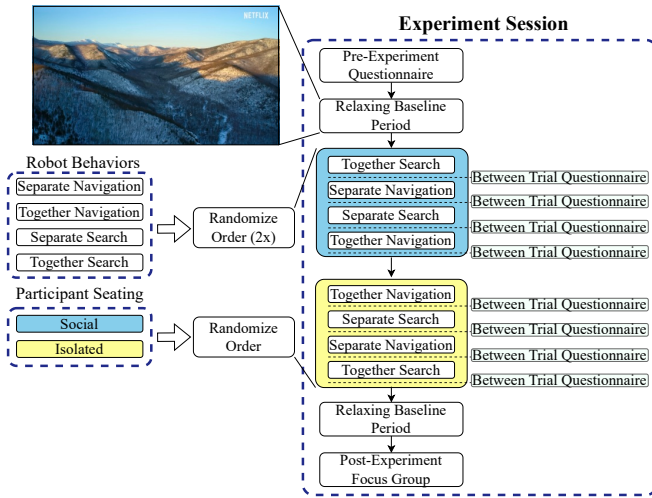
Fig. 2. Overview of an example experiment session designed to gain insights into human responses to encounters with robots. On the left are the list of 4 robot behaviors and 2 participant seating options. The order of each are randomized and then inserted into the experiment session as shown on the right side. The first step is receiving consent from the participants. The nature video is repeated before and after the experimental session as a relaxed baseline phsyiological state. After sensors are removed, participants take part in an interview.

physiological measures to identify stress response. These studies integrate robots in human activities with known external stressors like seminar style presentations or cognitive quizzes to evoke discomfort [16, 19, 21]. They ask how replacing humans with robots impacts people's response to situations, however none consider mobile robots moving freely through shared environments. The second group of studies leverage more subjective measures, for example statistical analysis of subjective questionnaires [6, 7, 22, 23, 24, 25] or qualitative evaluation of open form responses [26, 27, 28, 29]. Ultimately, the proposed study differentiates from many works, using a combination of methods to investigate perceived safety during mobile robot encounters.

B. Stress Decoding from Physiological Signals

Non-invasive physiological recordings have been shown to contain a wealth of information about human physical and mental states [30, 31]. We are interested in mental stress, particularly acute stress from short recordings. This is fulfilled by electrodermal activity (EDA) recordings, or galvanic skin response, which measures changes in skin conductance over time on the palm, foot, or wrist [32, 33]. Skin conductance time series consists of two components: tonic and phasic. Tonic refers to the low-frequency changes in the baseline of the signal, while the phasic refers to the sharp spikes followed by recovery to baseline in the signal [34]. The phasic component is known to correspond to psychological arousal events, allowing quantitative determination of mental stress over time. Decoding acute stress from tonic-phasic decompositions of EDA signals has been demonstrated extensively [30, 32, 33, 35].

Another signal to detect mental stress is heart rate variability (HRV) [36], which can be evaluated from any modality for reliable heart beat information. HRV features have been linked to a variety of both acute and chronic stress

in both real life and experimentally induced situations [31]. This is due to the fact that HRV reflects complex changes in sympathetic and parasympathetic nervous systems which manage the homeostasis of psychological arousal. HRV is often evaluated over long recordings; however it can be reliably estimated from ultra-short-term and short-term recordings like our study [37, 38]. Acute stress has been associated with decreases in HRV entropy in both experimentally induced situations [38, 39] and real-life situations [40, 41]. SDNN (Standard Deviation of NN intervals) and RMSSD (Root Mean Square of Successive Differences) are both measures commonly used in the analysis of heart rate variability (HRV), which is the variation in the time intervals between successive heartbeats. Decreases in SDNN and RMSSD and increases in low-frequency bandpower (0.04–0.15 Hz) compared to high-frequency (0.15–0.40 Hz) are also commonly noted features of acute stress [31].

Estimation of mental stress from EDA and HRV is non-trivial due to these signals being indirect measures of certain changes caused by the autonomic nervous system, acting as proxies for mental stress. Therefore, to maximize success, raw signals from participants must be as high-quality and free of noise and motion artifacts as possible, to ensure that no information is lost to such factors. While this incentivizes the use of clinical grade equipment, this conflicts with our study goals where we are interested in realistic human-robot encounters in an apartment setting. We conclude that state-of-the-art wearable devices are more suitable to avoid lengthy sensor set-up, skin preparation and cables that interfere with movement or cause stress on their own. To this end, for ECG (HRV), we used a wireless chest e-tattoo device [18] which is accurate while being minimally intrusive due to its ultrathin and lightweight form factor. Although this device records both electrocerdiography (ECG) and seismocardiography (SCG), only the former was used for analysis in this study. For EDA, we used an Empatica E4 wristwatch [17], which, according to a 2023 review on wearables used to assess perceived stress [42], was used in 56.3% of such studies.

III. EXPERIMENT DESIGN

We conduct a study with 17 total participants over 9 sessions to investigate perceived safety in human robot encounters. This study was independently reviewed and approved by the University of Texas Institutional Review Board (IRB #00004099). An overview of experimental sessions are shown in Fig. 2. The blockwise randomization of the trial configurations ensured that the effect of confounding factors such as time were removed. Fig. 2 shows each configuration was also repeated twice to enhance the repeatability of results from each configuration. The research questions in this work are as follows:

- $Q1$: Do participants have a stress response when robots are present compared to baseline sessions?
- $Q2$: What mobile robot dog behaviors cause human stress response?
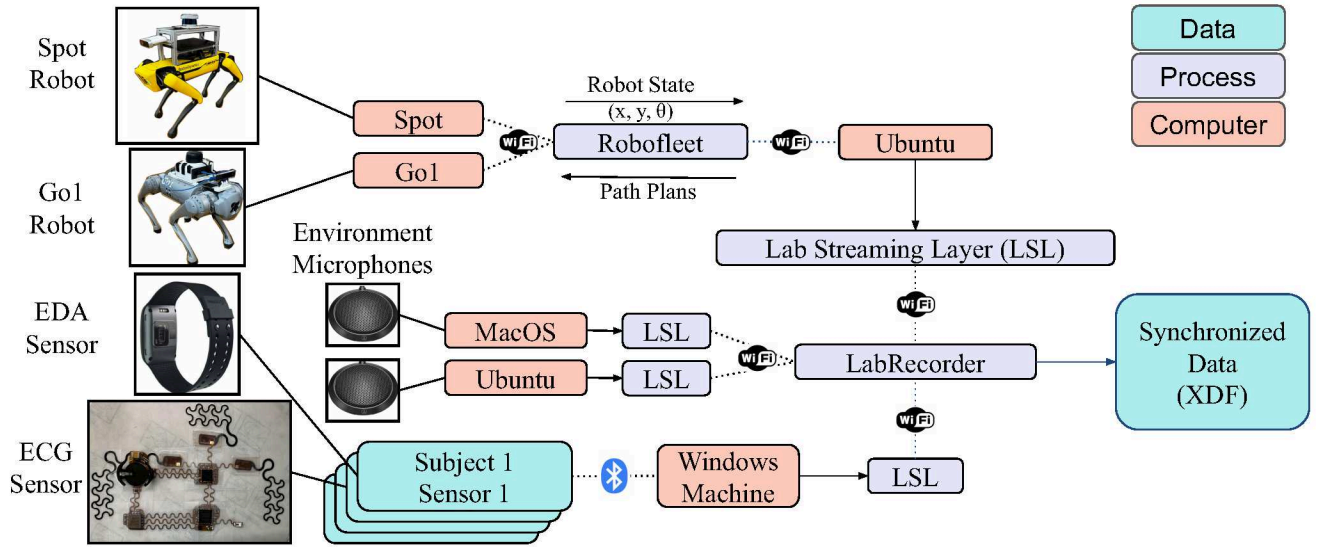- $Q3$: Do participants feel more comfortable when seated together vs. when isolated?

Fig. 3. A figure demonstrating the syncronized data acquisition architecture. The Spot and Go1 robots send state information via Robofleet [43] to an Intel NUC base station. Two omnidirectional microphones connect to a MacBook Pro (MBP) and Ubuntu (knapsack) machines. Finally, the two sensors that was participants wears connect via Bluetooth to designated Android and Windows machines. LabStreamingLayer [44] is used to push timestamped sensor data from all different sources to the network. LabRecorder [45] is used to record all streams into a single synchronized XDF file.

A. Data Acquisition

Synchronized data acquisition from all physiological sensors and the robots enables analysis of response to the varied conditions. We also collect data from two microphones in the environment, one in the small room and one in the large living room near the participant seating locations that record audio intensity streams. From each robot we collect position and orientation, represented as $(x, y, \theta)_t^{Spot}$ and $(x, y, \theta)_t^{Go1}$ recorded at time t. Each participant also wears the chest e-tattoo and the Empatica E4 wristwatch.

Figure 3 shows the architecture for real-time synchronized data acquisition. Robot data is sent as timestamped ROS [46] messages via Robofleet [43] to a base station. The base station and other systems stream robot and microphone data via LabStreamingLayer [44]. EDA and ECG sensors on each participant connect via Bluetooth to a Windows machine for streaming. All data is recorded using LabRecorder [45] and saved as a synchronized Extensible Data Format (XDF) file.

B. Conditions

We control three variables: (1) single v. multi-robot, (2) navigation v. search behavior, and (3) social v. isolated participant seating. Fig. 4 demonstrates the control variables overlaid on a map of the apartment, showing the four robot behaviors. The figure shows the navigation paths Fig. 4(a,b) and the search paths Fig. 4(c,d). In the large room, one or both participants sit together on a couch and then one or no participant on a chair in the small room (Fig. 4(a)). A pre-experiment questionnaire collects data about the personality, baseline stress and demographics of the involved participants.

C. Experiment Procedure

First, participants are presented with a consent form. Upon signing consent, the chest e-tattoo and E4 wristwatch are applied to participants. Participants are seated together on the living room couch to fill out a pre-experiment questionnaire.

The participants then watch a 5 minute relaxing nature video to establish a relaxed baseline state for physiological data. After, the participants' seating locations are updated to isolated if necessary (chosen at random). Next, experimental trials begin, each characterized by (1) a seating position, (2) single or multi-robot and (3) search or navigation behavior. Order effects of the conditions are addressed by counterbalancing. In particular, we randomize the isolated v. participant seating and each of the four robot behaviors within each seating condition. Between trials participants take a 10 question Likert scale questionnaire. It is a combination of surveys focused on stress, comfort and perceived safety used in [16]. After four trials, participant seating is switched. The same 4 behaviors are again repeated in the new seating locations with order again randomized. Participants then watch the same 5 minute nature video, establishing a second known baseline relaxation state to confirm that perturbations in the physiological signals during the 8 trials were induced by the robot and not a confounding factor such as elapsed time in the experiment and fatigue. After the nature video, participants removed the sensors and were debriefed then interviewed.

IV. DATA ANALYSIS METHODS

A. Signal Processing and Feature Extraction

ECG signals were validated first by visually examining the waveform of the average R peak grand average for a clear QRS complex and then by ensuring that the heart rate was reasonable. HRV features were extracted from peak information using Neurokit2 [47], which returns a total of 91 features from both the time and frequency domains as well as non-linear features. From these, 26 features cannot be reliably evaluated from ultra-short-term recordings and were removed. In our case, this consisted of HRV bandpower in the ULF and VLF bands, detrended fluctuation analysis (DFA)-related indices, as well as any time domain features
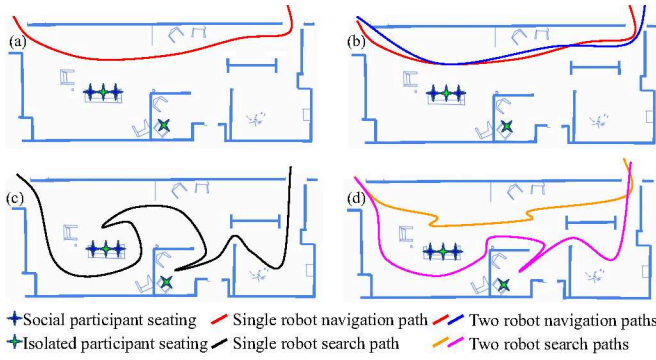
Fig. 4. An overview of the three control variables. Namely, four robot behaviors and two participant seating settings. In (a) the social and isolated participant seating locations are shown with the single robot navigation path. (b) is the two robot navigation paths. In (c) is the single robot search path and (d) shows the two robot search paths.

that required recordings longer than 5 minutes.

We found that there was a level of noise mixed into the EDA signals in some sessions, possibly due to poor skin-electrode contact during subject motion and the low sampling rate of the device. To mitigate this, a 2nd order Savitzky–Golay filter was used to act as a low-pass filter and smoothen out the noise before passing the pre-processed signal to the cvxEDA algorithm [34]. This algorithm decomposes the 1-dimensional EDA signal into phasic and tonic components via convex optimization. EDA features were then calculated according to the methodology in [34], with the addition of EDA autocorrelation [48], normalized EDA sympathetic index, as well as median and maximum amplitudes of the phasic sudomotor nerve activity (SMNA) peaks for a total of 15 EDA features. Features were normalized by the recording length wherever necessary. EDA and HRV features were then Z-scored and combined to result in a total feature matrix of 17 subjects × 10 trials × 80 features.

B. Exploratory Data Analysis

Next, we used the Chi-squared test to examine the 20 most statically significant features for the Baseline v. Robot condition. We chose a Chi-squared test compared with a principal component analysis (PCA) for interpretability of the results. While a PCA can be performed for a similar purpose, the resulting components cannot be compared with features of acute stress identified from literature as done in this work, which results in less certainty about the applicability of the results. This is due to the fact that there is no common "stress metric" in literature which can be used. In the feature selection via chi-squared test function available in Matlab, which is good to use for continuous data for classification problems, the features are ranked according to negative logs of the p-values. We note that EDA and HRV features were shortlisted in an approximately even proportion (9 EDA, 11 HRV) by the algorithm despite more candidate HRV features evaluated. This suggests that the multimodal approach based on both cardiac and electrodermal activity was informative for the estimation of robot-induced stress. Since exposure to robots is not a well-established stressor condition (in contrast to established psychological or neuroscientific tasks
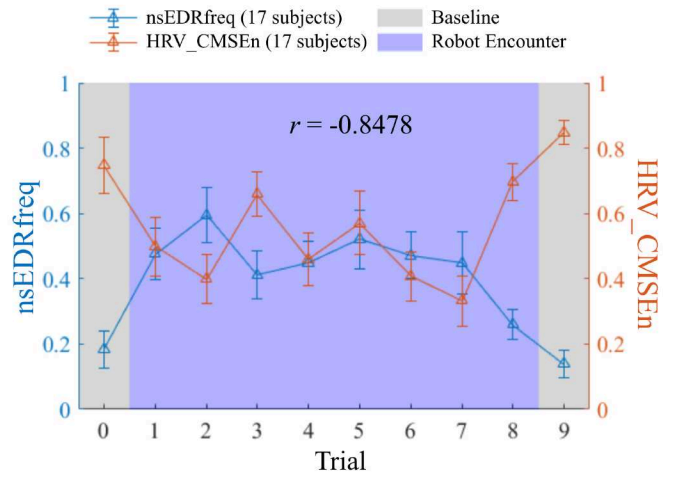

Fig. 5. Normalized features HRV$_{CMSEn}$ and nsEDRfreq over trials i.e. experiment time. Data shows averaged data from 17 subjects and their standard error. The trends show negative correlation as expected ($r = -0.8478$, see also Section IV-B) and show a clear U-shape as values representing the baseline relaxed state are perturbed by the introduction of robots into the environment in trials 1 through 8 before returning to normal (see further analysis in Fig. 6).

such as Stroop and Speech tests), an important step is to: (1) further examine the shortlisted features based on whether they have reasonable physiological background for being predictive of acute stress and arousal, and (2) examine the evolution of these features in the encounter conditions of interest. These steps will help to ensure that the predictive performance of the automatic human-robot encounter classifier in Section IV-C is not a result of overfitting.

We first plot physiological changes induced by robot encounters in Fig. 5, represented by 2 principal features: HRV composite multiscale entropy ("HRV_CMSEn") and the number of EDA phasic SMNA peaks normalized by length of the recording ("nsEDRfreq"). HRV_CMSEn is calculated from the chest e-tattoo's ECG signals using the methodology in [49] and nsEDRfreq is calculated from the E4's EDA signals using the methodology in [30]. HRV_CMSEn and nsEDRfreq showed high negative correlation ($r = -0.8478$) in their values over experiment time ("trials"), which show successful estimation of acute stress over time in which 2 devices with different sensing modalities observed similar stress-indicative signal properties across the experiment. This agrees with literature reporting increases in EDA phasic activity (i.e. nsEDRfreq) and decreases in HRV entropy (i.e. HRV_CMSEn) to be indicative of increased acute stress (see Section II-B). We plotted average feature changes over all participants according to the encounter conditions and conducted pairwise comparisons (Fig. 6) to investigate if the controlled encounter conditions induced physiological changes compared to baseline across all subjects.

C. Classification of Human-Robot Encounters

To confirm the insights from the exploratory analysis, we built a machine learning-based classifier of human-robot encounters using features from EDA and ECG signals. To ensure fair evaluation, we report leave-one-subject-out (LOSO) cross-validated results. In each iteration of LOSO cross-validation, the model is trained on data from all subjects
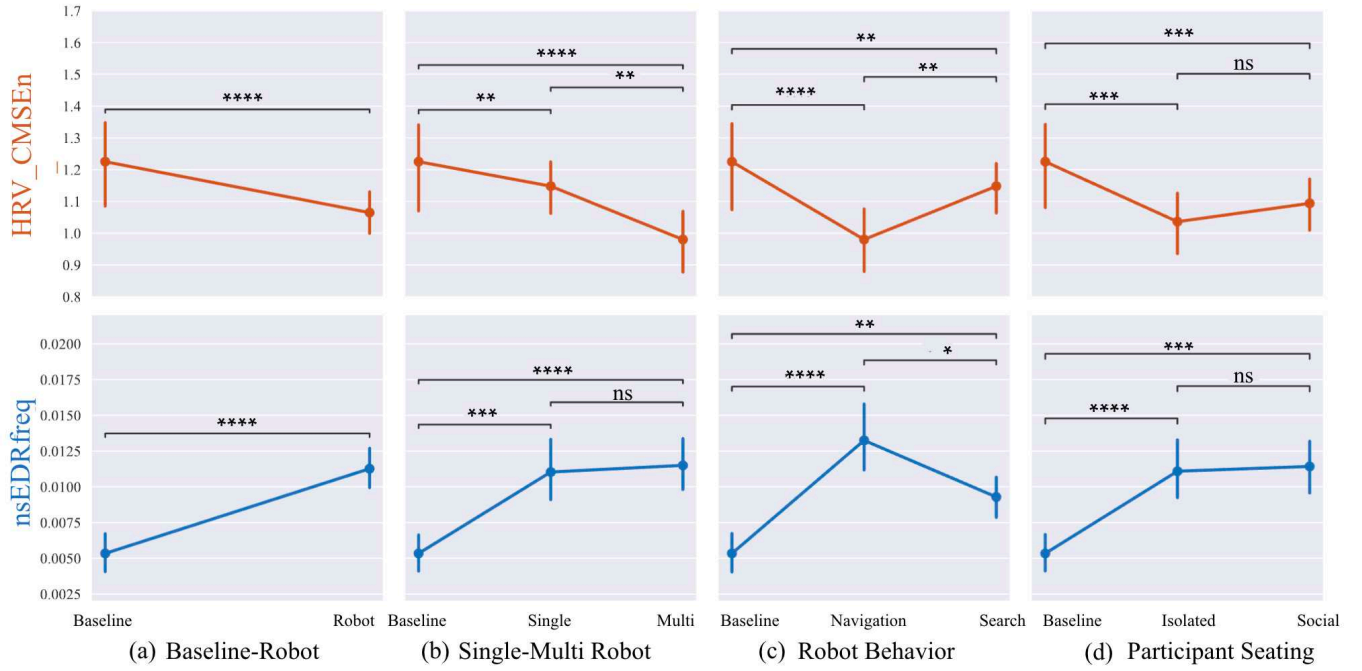
Fig. 6. Statistical comparison for the principal features HRV$_{CMSEn}$ and nsEDRfreq in each possible experimental condition, with all subject data. Error bars show the 95% confidence interval. Asterisks indicate statistical significance after two-sided Mann-Whitney-Wilcoxon tests with Bonferroni correction for multiple comparisons - ns: $0.05 < p$; *: $0.01 < p \leq 0.05$; **: $0.001 < p \leq 0.01$; ***: $0.0001 < p \leq 0.001$; ****: $p \leq 0.0001$.

other than one before prediction is evaluated on data from the unseen subject. This process is repeated the same number of times as the number of subjects to cover all possible train-test splits. LOSO cross-validation therefore ensures that the reported performance accounts for the classifier's ability to predict the type of human-robot encounter experienced by any potential unseen person wearing the same sensors. We decided to use random undersampling boosting (RusBoost) [50] for the binary classification and multi-class adaptive boosting (AdaBoost) [51] for the multi-class classification. For the former, RusBoost was chosen for its known robustness to class imbalances resulting from our study design where subjects spent more time encountering robot(s) rather than not, whereas for the multi-class classification, this was not a concern due to the further categorization of robot encounters into types. The maximum number of splits was fixed at 20.

V. RESULTS

All of the data taken during experiments including from robots, microphones, and physiological sensors is published for open access at [52] and a video of experimental encounters can be found at `https://youtu.be/3xu5AoBYi44`. Quantitative results are also reported in the data repository. Fig. 7 demonstrates screenshots from several encounters with human participants. For data from all subjects combined, the Mann-Whitney-Wilcoxon two-sided test was used with Bonferroni correction for multiple comparisons. On average, we found that subjects were: more stressed whenever the robots were in the shared environment compared to when they were not; more stressed when the robots were travelling together rather than alone; and more stressed when the robots were navigating compared to searching (Fig. 6(a-c)). Whether the subjects

were located alone or together did not induce significant changes in the stress features examined (Fig. 6(d)).

For subject-specific data (Fig. 5), to control for inter-subject variability, repeated measures ANOVA was used to investigate the effect of experiment conditions on the trial-level EDA response (nsEDRfreq). As noted on Table I, we found that the EDA response within subjects was significantly altered across the 10 trials. Specifically, individual subjects had their EDA response change significantly depending on the presence of a robot (Baseline-Robot) and robot behavior.

The predictive value of the extracted HRV and EDA features were validated by classifying human-robot encounters via machine learning according to Section IV-C. The resulting LOSO-validated confusion matrices are presented in Tables II - V. As hinted by the statistical significance of differences in principal features (Fig. 6) between the encounter conditions, it is within expectation that the binary classifier for identifying robot encounters was the most effective, followed by 3-class classifiers for single- vs. multi-robot encounters and encounters with navigating robots vs. searching robots. Classifying isolated- vs. together-seating encounters was the least effective.

In Fig 6, the values are adjusted with Bonferroni correction. The two-sided Mann-Whitney-Wilcoxon tests with Bonferroni correction that yields pairwise across-subjects significance is different from the p-values reported in Table I, which instead refers to the within-subject analysis with repeated measures ANOVA. Specifically, the p-values in Table I are the p-values of the F-statistic, which simply shows how likely is it that the null hypothesis of "no difference among group means" is true. Therefore, for this analysis, to the best of our knowledge, adjustment is not necessary. There

is, however, a procedure for instead getting pairwise (e.g. isolated-baseline, social-baseline, social-isolated) adjusted p-values for within-subjects RMANOVA, but it is very complicated and will result in similar information to Fig. 5, which presents the pairwise comparisons across subjects, which is a more generalizable result than the within-subjects one.

	F value	p-value
Trials ($df = 9$)	2.763	0.00588**
Baseline-Robot ($df = 1$)	13.477	0.000365***
Participant Seating ($df = 1$)	0.463	0.498
Robot Behavior ($df = 1$)	5.986	0.0159*
Single-Multi Robot ($df = 1$)	2.215	0.139

***p-value ≤ 0.001; **p-value ≤ 0.01; *p-value ≤ 0.05.

TABLE II

ROBOT ENCOUNTERS VS. BASELINE

		Predicted	
		Baseline	Robot
Actual	Baseline	70.0%	30.0%
	Robot	17.5%	82.5%

TABLE III

SINGLE VS. MULTI-ROBOT ENCOUNTERS

		Predicted		
		Baseline	Single	Multi
Actual	Baseline	56.7%	36.7%	6.7%
	Single	20.0%	46.7%	33.3%
	Multi	10.0%	33.3%	56.7%

TABLE IV

ENCOUNTERS WITH NAVIGATING VS. SEARCHING ROBOTS

		Predicted		
		Baseline	Navigation	Search
Actual	Baseline	53.3%	0.0%	43.3%
	Navigation	3.3%	65.0%	31.7%
	Search	16.7%	30.0%	53.3%

TABLE V

ENCOUNTERS IN ISOLATED VS. SOCIAL SEATING

		Predicted		
		Baseline	Isolated	Social
Actual	Baseline	56.7%	33.3%	10.0%
	Isolated	26.8%	46.4%	26.8%
	Social	7.8%	46.9%	45.3%

VI. DISCUSSION

Fig. 6 and Tables II - V suggest that for the average human participant, sharing indoor spaces with mobile quadrupedal robots was associated with statistically significant physiological changes indicative of acute stress, which provides a response to $Q1$. Further, it was possible to distinguish, with higher than chance-level accuracy, specific characteristics (single v. multi robot and navigation v. search) about the human-robot encounters based on the ECG and EDA signals. These statistical correlations provide insight to research questions $Q2$. These observations should be critically examined for potential insights for designers of robots and human-robot interaction.

For instance, we found that the average participant was more stressed by robots on navigation behavior than those on search behavior. One could hypothesize that this was due to robots on search inherently spending more time near the participants, giving them an opportunity to become more familiar with the robot. Indeed, familiarity, which is influenced by interaction duration and frequency, is one of the major factors behind perceived safety [10]. However, to confirm this idea, further analysis of the data may be necessary (e.g. correlate human-robot distance with stress features to decouple navigation/search behavior from distance; interview and questionnaire data collected may contain information about how differently the subjects perceived robots on navigation vs. search). While less statistically significant, we also found that the average participant was more stressed encountering multiple robots travelling together rather than alone. This could suggest that whenever possible, mobile robots should travel alone, at least in similar physical settings as our experiment (i.e. closed indoors space occupied by humans), to maximize perceived safety. We also found that whether the subjects were alone or together was not associated with significant differences in physiological features of stress, responding to $Q3$. This could indicate that the level of stress perecieved by people due to the co-location by mobile robots may not pose significant difference when they are deployed in environments where people work alone vs environments where multiple people work alongside each other.

Fig. 5 and Table I suggest that these changes were consistent when examined on a per-participant basis. This could potentially allow longer-term tracking of individuals' perceived safety of mobile robots when combined with the appropriate wearable sensing technology and wireless data acquisition hardware, as demonstrated in this study. Overall, this study outlined an objective approach to recording human-robot encounters, and investigated factors affecting perceived safety of mobile quadrupeds in indoor environments by analyzing human physiological markers of acute stress.

VII. CONCLUSIONS

We presented results found from experimental human-robot encounters in an indoor apartment setting with members of the University of Texas at Austin. Particularly, we found that participants showed a measurable response to the robots over the baseline, multiple robots compared to a single robot, and under navigation compared with search behavior. Results were computed using statistical analysis of features found from electrodermal activity and heart rate variability recorded during experiments. One can hypothesize that the classifier may generalize to other HRI scenarios where participants are seated in a comfortable indoor environment when robots are present. Further, this classifier may provide a baseline from which to continue to develop an online classifier for acute mental stress, which may be used to control robot behavior. The complete open access database can be found at [52] and includes a plethora of
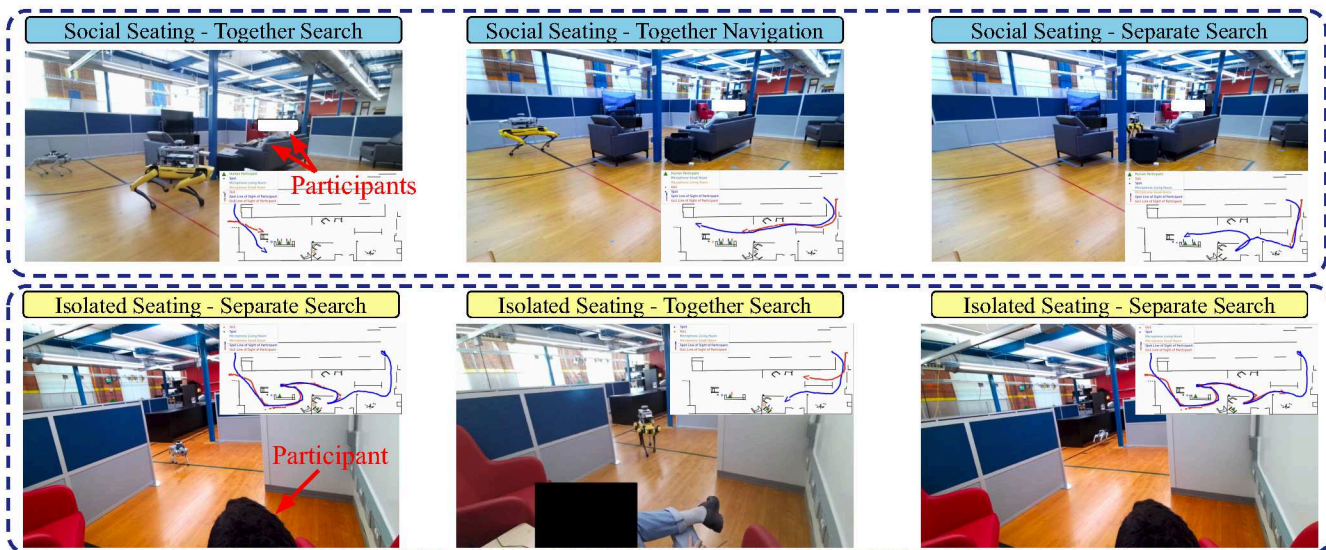
Fig. 7. Moments captured during encounters with real human participants. Overlays are used to protect the identity of participants. Robot path and position in various behaviors is overlaid on a map of the apartment environment at the moment captured. Robot data is visualized using the XDF files.

rich time series data outside of the scope of this manuscript. For instance, one can manipulate the robot pose information to identify when the robots are in the line-of-sight of participants or the distance to participants. This data offer a new context from which to study proxemics. The dataset also includes audio intensity data from participant seating locations, which can further the investigation of the impacts of sound on physiological response. Ongoing work looks to combine the presented results with findings from interviews with participants in order to paint a more complete picture of human response to mobile robot encounters.

VIII. ACKNOWLEDGMENTS

REFERENCES

[1] Daegyu Lee et al. "Assistive delivery robot application for real-world postal services". In: *IEEE Access* 9 (2021), pp. 141981–141998.

[2] Ryan Gupta et al. "Fast LiDAR Informed Visual Search in Unseen Indoor Environments". In: *arXiv preprint arXiv:2309.14150* (2023).

[3] Ben Grocholsky et al. "Cooperative air and ground surveillance". In: *IEEE Robotics & Automation Magazine* 13.3 (2006), pp. 16–25.

[4] Benjamin Charrow et al. "Information-theoretic mapping using cauchy-schwarz quadratic mutual information". In: *2015 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE. 2015, pp. 4791–4798.

[5] Christoph Bartneck et al. "Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots". In: *International journal of social robotics* 1 (2009), pp. 71–81.

[6] Elliott Hauser et al. "" What's That Robot Doing Here?": Perceptions Of Incidental Encounters With Autonomous Quadruped Robots". In: *Proceedings of the First International Symposium on Trustworthy Autonomous Systems*. 2023, pp. 1–15.

[7] Sanne Van Waveren et al. "Increasing perceived safety in motion planning for human-drone interaction". In: *Proceedings of the 2023 ACM/IEEE International Conference on Human-Robot Interaction*. 2023, pp. 446–455.

[8] Ada V Taylor, Ellie Mamantov, and Henny Admoni. "Observer-aware legibility for social navigation". In: *2022 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. IEEE. 2022, pp. 1115–1122.

[9] Bernard CK Choi and Anita WP Pak. "Peer reviewed: a catalog of biases in questionnaires". In: *Preventing chronic disease* 2.1 (2005).

[10] Neziha Akalin et al. "A Taxonomy of Factors Influencing Perceived Safety in Human–Robot Interaction". In: *International Journal of Social Robotics* 15.12 (2023), pp. 1993–2004.

[11] Matteo Rubagotti et al. "Perceived safety in physical human–robot interaction—A survey". In: *Robotics and Autonomous Systems* 151 (2022), p. 104047.

[12] Frédéric Dehais et al. "Physiological and subjective evaluation of a human–robot object hand-over task". In: *Applied ergonomics* 42.6 (2011), pp. 785–791.

[13] Tiffany L Chen et al. "Touched by a robot: An investigation of subjective responses to robot-initiated touch". In: *Proceedings of the 6th international conference on Human-robot interaction*. 2011, pp. 457–464.

[14] Jessi Stark, Roberta RC Mota, and Ehud Sharlin. "Personal space intrusion in human-robot collaboration". In: *Companion of the 2018 ACM/IEEE international conference on human-robot interaction*. 2018, pp. 245–246.

[15] Christian JAM Willemse, Alexander Toet, and Jan BF Van Erp. "Affective and behavioral responses to robot-initiated social touch: toward understanding the opportunities and limitations of physical contact in human–robot interaction". In: *Frontiers in ICT* 4 (2017), p. 12.

[16] Neziha Akalin, Annica Kristoffersson, and Amy Loutfi. "Do you feel safe with your robot? Factors influencing perceived safety in human-robot interaction based on subjective and objective measures". In: *International journal of human-computer studies* 158 (2022), p. 102744.

[17] URL: https://www.empatica.com/research/e4/.

[18] Sarnab Bhattacharya et al. "A Chest-Conformable, Wireless Electro-Mechanical E-Tattoo for Measuring Multiple Cardiac Time Intervals". In: *Advanced Electronic Materials* (2023), p. 2201284.

[19] Mariacarla Staffa and Silvia Rossi. "Enhancing Affective Robotics via Human Internal State Monitoring". In: *2022 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. IEEE. 2022, pp. 884–890.

[20] Pramila Rani et al. "Anxiety detecting robotic system–towards implicit human-robot collaboration". In: *Robotica* 22.1 (2004), pp. 85–95.

[21] Julie M Turner-Cobb et al. "Use of a non-human robot audience to induce stress reactivity in human participants". In: *Computers in Human Behavior* 99 (2019), pp. 76–85.

[22] Pourya Aliasghari et al. "How do we perceive our trainee robots? exploring the impact of robot errors and appearance when performing domestic physical tasks on teachers' trust and evaluations". In: *ACM Transactions on Human-Robot Interaction* 12.3 (2023), pp. 1–41.

[23] Ziming Wang et al. "The effects of natural sounds and proxemic distances on the perception of a noisy domestic flying robot". In: *ACM Transactions on Human-Robot Interaction* 12.4 (2023), pp. 1–32.

[24] Sheldon Cohen, Tom Kamarck, Robin Mermelstein, et al. "Perceived stress scale". In: *Measuring stress: A guide for health and social scientists* 10.2 (1994), pp. 1–2.

[25] Christoforos Mavrogiannis et al. "Social momentum: Design and evaluation of a framework for socially competent robot navigation". In: *ACM Transactions on Human-Robot Interaction (THRI)* 11.2 (2022), pp. 1–37.

[26] Nicola Döring and Sandra Poeschl. "Love and sex with robots: a content analysis of media representations". In: *International Journal of Social Robotics* 11.4 (2019), pp. 665–677.

[27] Leopoldina Fortunati et al. "The rise of the roboid". In: *International Journal of Social Robotics* 13 (2021), pp. 1457–1471.

[28] Helene Joffe. "Thematic analysis". In: *Qualitative research methods in mental health and psychotherapy: A guide for students and practitioners* (2011), pp. 209–223.

[29] Gareth Terry et al. "Thematic analysis". In: *The SAGE handbook of qualitative research in psychology* 2.17-37 (2017), p. 25.

[30] Alberto Greco et al. "Acute stress state classification based on electrodermal activity modeling". In: *IEEE Transactions on Affective Computing* (2021).

[31] Sarah Immanuel et al. "Heart Rate Variability for Evaluating Psychological Stress Changes in Healthy Adults: A Scoping Review". In: *Neuropsychobiology* 82.4 (June 2023), pp. 187–202. ISSN: 0302-282X.

[32] Roberto Zangróniz et al. "Electrodermal activity sensor for classification of calm/distress condition". In: *Sensors* 17.10 (2017), p. 2324.

[33] Hugo D Critchley. "Electrodermal responses: what happens in the brain". In: *The Neuroscientist* 8.2 (2002), pp. 132–142.

[34] Alberto Greco et al. "cvxEDA: A convex optimization approach to electrodermal activity processing". In: *IEEE trans-actions on biomedical engineering* 63.4 (2015), pp. 797–804.

[35] Delphine Caruelle et al. "The use of electrodermal activity (EDA) measurement to understand consumer emotions–A literature review and a call for action". In: *Journal of Business Research* 104 (2019), pp. 146–160.

[36] Hye-Geum Kim et al. "Stress and heart rate variability: a meta-analysis and review of the literature". In: *Psychiatry investigation* 15.3 (2018), p. 235.

[37] Lizawati Salahuddin et al. "Ultra short term analysis of heart rate variability for monitoring mental stress in Mobile settings". In: *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (Aug. 2007).

[38] Seungjae Lee et al. "Mental stress assessment using ultra short term HRV analysis based on non-linear method". In: *Biosensors* 12.7 (June 2022), p. 465.

[39] Agostino Brugnera et al. "Heart rate variability during acute psychosocial stress: A randomized cross-over trial of verbal and non-verbal laboratory stressors". In: *International Journal of Psychophysiology* 127 (2018), pp. 17–25. ISSN: 0167-8760.

[40] Dimitriy A. Dimitriev, Elena V. Saperova, and Aleksey D. Dimitriev. "State anxiety and nonlinear dynamics of Heart Rate Variability in students". In: *PLOS ONE* 11.1 (Jan. 2016).

[41] Paolo Melillo, Marcello Bracale, and Leandro Pecchia. "Nonlinear heart rate variability features for real-life stress detection. case study: Students under stress due to university examination". In: *BioMedical Engineering OnLine* 10.1 (Nov. 2011), p. 96.

[42] Agata Klimek et al. "Wearables measuring electrodermal activity to assess perceived stress in care: a scoping review". In: *Acta Neuropsychiatrica* (2023), pp. 1–11.

[43] Kavan Singh Sikand et al. "Robofleet: Open source communication and management for fleets of autonomous robots". In: *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE. 2021, pp. 406–412.

[44] URL: https://github.com/sccn/labstreaminglayer.

[45] URL: https://github.com/labstreaminglayer/App-LabRecorder.

[46] URL: http://wiki.ros.org/costmap_2d.

[47] Dominique Makowski et al. "NeuroKit2: A Python toolbox for neurophysiological signal processing". In: *Behavior research methods* (2021), pp. 1–8.

[48] Sjoerd van Halem et al. "Moments That Matter? On the Complexity of Using Triggers Based on Skin Conductance to Sample Arousing Events Within an Experience Sampling Framework". In: *European Journal of Personality* 34.5 (2020), pp. 794–807.

[49] Shuen-De Wu et al. "Time Series Analysis Using Composite Multiscale Entropy". In: *Entropy* 15.3 (2013), pp. 1069–1084. ISSN: 1099-4300. URL: https://www.mdpi.com/1099-4300/15/3/1069.

[50] Chris Seiffert et al. "RUSBoost: Improving classification performance when training data is skewed". In: *2008 19th International Conference on Pattern Recognition*. 2008, pp. 1–4.

[51] Trevor Hastie et al. "Multi-class adaboost". In: *Statistics and its Interface* 2.3 (2009), pp. 349–360.

[52] Ryan Gupta et al. *Community Embedded Robotics: A Multimodal Dataset on Perceived Safety during Indoor Mobile Robot Encounters*. Version V1. 2024. DOI: 10.18738/T8/FT9VYS. URL: https://doi.org/10.18738/T8/FT9VYS.