HRI-SENSE: A Multimodal Dataset on Social and Emotional Responses to Robot Behaviour

Balint Gucsi, Nguyen Tan Viet Tuyen, Bing Chu, Danesh Tarapore
School of Electronics and Computer Science,
University of Southampton, UK
{bg1u17,tuyen.nguyen,b.chu,d.s.tarapore}@soton.ac.uk

Long Tran-Thanh
Department of Computer Science,
University of Warwick, UK
long.tran-thanh@warwick.ac.uk

***Abstract*—We introduce HRI-SENSE, a multimodal dataset of Human-Robot Interactions (HRI) studying users' social, physical (e.g. facial expressions, body movements) and emotional, psychological (e.g. frustration, satisfaction) responses to robot behaviour. The dataset captures participants collaborating with a TIAGo humanoid robot following various behaviour models on a manipulation-based "Burger Assembly" task, eliciting different user reactions. HRI-SENSE contains over 6 hours of verbal and physical interactions taking place over 146 sessions with 18 participants, recording multiple modalities captured simultaneously by RGB and Depth cameras from three angles and one microphone. The time-synchronized multimodal data include non-verbal behaviours (e.g. facial landmarks, expressions, pose landmarks), explicit feedback signals (e.g. verbal interactions), robot movements and self-assessed questionnaires on sociodemographics and user impressions (e.g. frustration, satisfaction) on robot interactions. HRI-SENSE is expected to facilitate further research into modelling non-verbal behaviour and advancing the development of user-aware interaction models in HRI domain.**

***Index Terms*—Multimodal interaction dataset; Human-robot interaction; Assistive robotics; Social signals; Implicit feedback; Explicit feedback**

I. Introduction

Recent studies in HRI [1], [2], [3] highlight that the successful deployment of robotic systems in social settings is critically dependent on the acceptance of interacting users. User acceptance is heavily affected by their experience of working with robots, relying as much (or even more [4], [5]) on subjective aspects (e.g. experiencing helpfulness, success or lack of frustration) as objective metrics (e.g. speed of task completion). Consequently, the analysis of subjective user impressions in human-robot interactions may enable the development of systems that act in a user-aware manner, resulting in enhanced user acceptance.

Previous studies have shown that in addition to explicit feedback signs (e.g. expression of preference [6] or evaluative feedback [7]) and psychological behaviour patterns [8], [9], implicit feedback signals expressed through non-verbal behaviours [10] such as facial expressions or body movements can serve as an indication of similar user characteristics (e.g. engagement [11], [12], or stress [13]) with most of these works focusing on verbal interaction scenarios [14], [11], [12], [15]. However, little has been explored about how non-verbal social signals can infer characteristics of user impressions (e.g. interaction success, user frustration) in HRI, especially in interactions containing collaborative manipulation elements.

As a result, we introduce HRI-SENSE, a time-synchronized multimodal Human-Robot Interactions dataset on Social and Emotional respoNseS to robot bEhaviour, collected from users interacting with different user-aware autonomous robot behaviour models in a manipulation based "Burger Assembly" HRI task. Our dataset contains multiple non-verbal modalities simultaneously captured by two RGB cameras, a RGBD camera and a microphone (see Fig. 1 for data samples), as well as self-assessed questionnaires reporting user impressions. This makes it, to the best of our knowledge, one of the few HRI manipulation datasets capturing implicit and explicit feedback signals, with an additional focus on user impressions.

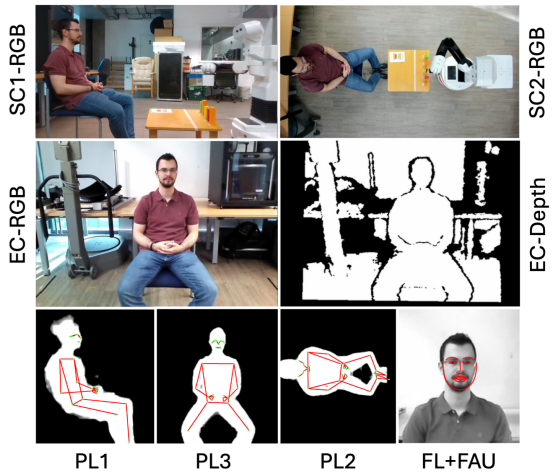

Fig. 1. Multimodal data samples recorded for the HRI-SENSE dataset. (SC: Static Camera, EC: Egocentric Camera, PL: Pose Landmarks, FL: Face Landmarks, FAU: Face Action Units)

II. Background

Although various multimodal interaction databases are available, only a limited number of them focuses on HRI. Recent research investigations into the utilisation of implicit user feedback in HRI generated collections of multimodal data on the interactions. The most relevant datasets incorporating aspects of implicit human reactions and communicatory signals (i.e. non-verbal cues such as facial expressions, gaze or body movements) to robots are summarised in Table I. They mostly consider HRI in verbal settings where interactions are recorded [14], [11], [12], [16] and often additional self-assessed feedback questionnaires provide explicit feedback and ground truth values [11], [12], [16]. This allows for examining both implicit and explicit evaluative feedback during interactions [16] or modelling user engagement [12], [11]. However, little is known about how implicit reactions reflect user impressions and generated emotions (e.g. frustration, satisfaction) on robot behaviour in HRI scenarios involving

Dataset	Subjects	Interaction Sessions	Duration (hours:minutes)	Robot Behaviour	Recorded Modalities and Published Features											
					V	A	D	MC	PL	HP	EG	FL	FAU	DT	RB	Other
EMPATHIC [16]	14	98	<20 sec per interaction	autonomous	✓	X	X	X	X	✓	X	✓	✓	X	✓	X
REACT-Nao [15]	72	432	14:24	autonomous	✓	X	X	X	X	✓	✓	✓	✓	X	✓	X
Errors in HRI (HRC-A) [17]	12	23	2:31	autonomous	✓	X	X	X	X	X	X	X	✓	X	X	X
Errors in HRI (HRC-C) [17]	33	65	4:17	autonomous	✓	X	X	X	X	X	X	X	✓	X	X	X
Errors in HRI (PbD) [17]	28	38	0:48	Wizard-of-Oz	✓	X	X	X	X	X	X	X	✓	X	X	X
UE-HRI [11]	54	54	4-15 min per interaction	autonomous	✓	✓	✓	X	✓	✓	✓	✓	X	X	✓	X
MHHRI [12]	18	48	6:00	autonomous, Wizard-of-Oz	✓	✓	✓	✓	✓	✓	✓	X	X	X	X	EDA
Vernissage Corpus [14]	26	13	2:23	Wizard-of-Oz	✓	✓	X	✓	X	✓	X	X	X	✓	✓	X
HARMONIC [18]	24	480	5:48	teleoperated	✓	X	X	X	✓	✓	✓	✓	X	X	✓	EMG
HRI-SENSE	18	146	6:03	autonomous	✓	✓	✓	X	✓	✓	X	✓	✓	✓	✓	X

TABLE I
COMPARISON OF PUBLICLY AVAILABLE RELATED MULTIMODAL HRI DATASETS. (V: RGB VIDEO, A: AUDIO, D: DEPTH FOOTAGE, MC: MOTION CAPTURE, PL: POSE LANDMARKS, HP: HEAD POSE, EG: EYE GAZE, FL: FACE LANDMARKS, FAU: FACE ACTION UNITS, DT: DIALOGUE TRANSCRIPT, RB: ROBOT BEHAVIOUR, EDA: ELECTRODERMAL ACTIVITY, EMG: ELECTROMYOGRAPHY)

manipulation. Existing datasets involving manipulation capture users' reactions to a robot making manipulation errors [17], to suboptimal robot behaviour [16] and when teleoperating a robot arm in a shared autonomy setting [18]. Doing so, they focus on identifying specific behaviour instances (e.g. robot error [17]) or on learning task statistics from implicit feedback [16], rather than capturing and investigating user impressions and emotional responses to robot behaviour.

Addressing these limitations, our data corpus, HRI-SENSE, is designed to contain both implicit reaction data of various modalities and explicit verbal communication collected from users participating in a manipulation based HRI task interacting with different robot behaviour models, complemented by self-assessed questionnaires on user impressions.

III. SENSOR SUITE

We perform data collection of the human participant's interaction with the TIAGo humanoid robot [19] using a suite of sensors containing multiple RGB cameras placed at different angles, a stereo and depth camera and a microphone (see setup in Fig. 2). Additionally, we record the robot's arm joint parameters representing the arm's position throughout the interaction. One RGB camera is mounted onto a ceiling rail system providing a top-down perspective recording of the interaction and a secondary RGB camera is positioned to capture a side view of the interaction. We utilise the TIAGo robot's built-in stereo depth cameras to focus on capturing the user's reactions and expressions from closer proximity. The interaction's verbal components and any ambient sounds are recorded from a side table adjacent to the interacting parties.

1) RGB Sensors: In addition to the capturing of visual data, our RGB cameras enable the extraction of valuable semantic information from the interactions. We particularly focus on recording participants' non-verbal signals, including gestures, body positions, facial expressions, and their relative position to the robot, since these details may be representative of their impressions or reactions elicited by the robot's actions. We choose a Logitech BRIO camera considering its wide view angle 90 degree, small form factor and ease of use (requiring a single USB cable for real-time data transmission and power supply) as our top-down camera, positioned directly above the table between the participants, attached to the ceiling rail

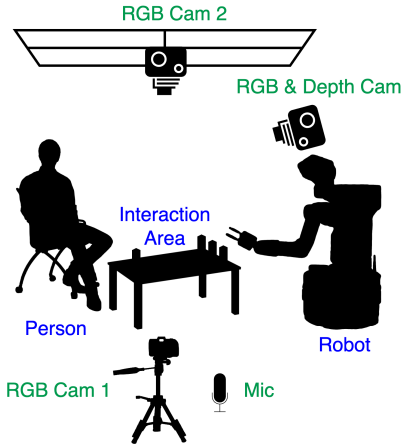

Fig. 2. Data collection setup for the HRI manipulation task with the TIAGo robot. We used two RGB static cameras, one RGB & Depth egocentric camera and one microphone.

system. An Intel RealSense D435i camera was used to create the side view recordings.

2) Stereo and Depth Camera: Many collaborative robots are equipped with an RGBD (stereo and depth) camera system attached to the robot's head facing forward to acquire geometric information about the environment in the form of depth information and to record the current task/interaction enabling the collection of visual feedback. This feedback can serve robot movement verification purposes (e.g. for following manipulation movements) or provide information on the collaborating user throughout the interaction, depending on the camera's pose and orientation. We focus on the latter, by positioning the camera at a fixed angle encompassing the user's face and front, capturing their reactions. Additionally, the depth information recorded from the cameras serves as an indication of the user's position and distance from the robot.

3) Microphone: As human-robot interaction scenarios, such as our interaction task, often include verbal communication elements, we utilise an external Yeti Nano microphone using omnidirectional patterns to record conversations and additional ambient sound during the interactions.

4) Robot Arm Joint Movements: Since our observed interaction involves the robot executing pick-and-place and general manipulation movements with the 7 degree-of-freedom (DoF)

robot arm, it is important to account for the robot arm's positions and movement characteristics (e.g. speed, acceleration, explainability, etc.) that may affect the user's impressions. We record the arm joint states provided by the robot's movement planner engine (MoveIt! [20]) throughout the interaction at 50Hz frequency. The recorded data enables the analysis of both the robot's movement behaviour and the robot joints' position relative to the environment or the human user, extracted from the joint states following the robot's geometry.

IV. The HRI-SENSE Dataset

The data collection was designed to take place as part of a human-robot collaborative task with verbal interactions and manipulations. Complementing the interaction structure of established assembly datasets [21], [22], [23] with HRI elements and user-preference based actions, designed for the collection of continuous (implicit and explicit) feedback of user reactions through various modalities, we introduce the "Burger Assembly" task domain. While collaborating, participants take an active role alongside the TIAGo Steel robot equipped with a 7 DoF arm and parallel grippers, running a completely autonomous system (i.e. without any teleoperation).

A. Interaction Scenario: The "Burger Assembly" Task

Context: As part of the collaborative task, the participant and a robot are asked to "cook" a hamburger, assembled from typical ingredients (hamburger buns, meat, salad, tomato) represented by coloured foam blocks placed on a table between the robot and person. The ingredients are referred to by numerical identifiers throughout the interactions. The robot takes an assistant role, handling the preparation of suitable ingredients and handing them to the human user for assembly.

Verbal Interactions: Firstly, collaborators are tasked with preparing a specific burger (of 3 different burger types), and participants receive instruction graphics (an ordered list of ingredients to assemble the burger from), which are unknown to the robot. To learn the suitable ingredient to prepare, the robot may pose inquires to the user, such as confirmation queries (e.g. "Should I pass you ingredient number 1?") or instruction queries (e.g. "What should I do next?). Alternatively, the robot may hand over an object without any prior inquiries as well, if deemed suitable. Over time, the robot learns how various burgers are typically constructed and may attempt to proactively prepare ingredients.

"Burger Assembly": Afterwards, the robot assists the user by "preparing" the ingredient considered suitable: the object is picked up and placed in front of the user on the table for assembly. Finally, the user stacks the received ingredient object on top of any previous ingredients, building the hamburger.

Verbal Feedback: Following each object handover, the robot inquires for a verbal feedback on its previous actions, to which the user responds. Participants have been asked to focus on the robot's intentions over the handover task execution's precision when providing their verbal feedback.

B. Data Collection Procedure

First, the participants are asked to complete a pre-study questionnaire focusing on their demographics, prior experience and personality. Then the Burger Assembly task is performed by the participants in collaboration with the TIAGo humanoid robot multiple times. Each participant interacts with 3 different robot behaviour models, including instruction-based, confirmation-based and hybrid, in which the robot's verbal interaction strategy may vary, aiming to elicit different interaction reactions depending on the context. Participants interact with different robot behaviour models in a random order, with the aim of avoiding bias. For each behaviour model, 3 sequential trials take place per participant, incentivising them to accommodate the robot's behaviour and appreciate the behaviour model's improvement over the trials. All interactions are recorded using our installed sensor suite (see data collection pipeline in Fig. 3). After completing interactions with each behaviour model, participants complete a post-interaction questionnaire rating their experience with the robot.

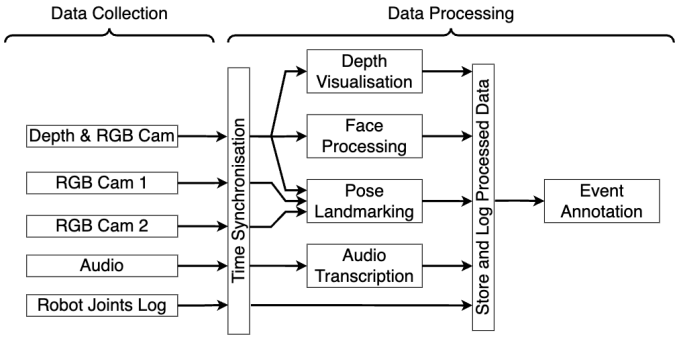

Fig. 3. Data collection and post-processing data extraction steps illustrated.

For each experiment trial, all data from the RGB cameras and microphone is collected onto the laptop's storage, while stereo and depth camera recordings and robot arm joint state logs are collected using the Robot Operating System (ROS) bag functionality. Data collection is started and stopped in a synchronous manner managed by a control module on the investigator's laptop, interacting with the TIAGo robot over the local network using the ZeroMQ communication toolset. The experiments were conducted at the University of Southampton Robotics Laboratory, following the approval of all experiment plans by the University's Ethics Committee.

C. Participants

Our data collection took place through sessions conducted with 18 participants associated with the University of Southampton, mostly graduate and undergraduate students and researchers. Participants came from various fields of expertise (i.e. course of study, degree, profession) and had different technological backgrounds and prior experience with robots. An informed consent was obtained from all participants, each of whom was assigned a numerical ID to anonymise the data.

D. Sensory Data Processing

1) User Poses: RGB videos captured from top-down, side and robot points-of-view are all pipelined into Google's MediaPipe [24] Pose Landmarker, which identifies 33 significant body landmark locations and locate these in a 3-dimensional space. The output matrices represent the user's body position describing various movements from all three recording angles without any privacy issues.

2) User Facial Expressions: To represent characteristics of the recorded facial expressions and typical movement patterns, we use head poses, facial landmarks and Action Units (AUs) as defined by the established Facial Action Coding System (FACS) [25]. Each RGB recording captured by the robot's camera is processed using the OpenFace toolkit [26], outputting head pose and rotation values, facial landmark coordinates and AU intensity, presence and confidence scores.

3) Audio Transcription: To provide a time-coded log of verbal interactions between the user and the robot during each experiment, we processed all recorded audio data using OpenAI's Whisper transcription model [27] (using English language, large model settings).

4) Depth Information: As recorded from the robot RGBD cameras, the depth footage is stored in the original 800*656 resolution, in 30 FPS in mp4 format. The recording was re-encoded to reduce its size without quality loss.

5) Labelled Annotations of Interaction Events: Significant interaction events corresponding to a list of textual tags that occurred during the experiments have been identified and manually labelled. Depending on the nature of the event, a timestamp or a time interval has been assigned as a label. The textual tags are designed to describe events affecting the outcome of the interaction or the user's impressions.

E. Questionnaire-based Assessments

Our pre-study questionnaire focused on assessing the participants' background information (e.g. demographics, prior experience and expectations with technology and robots) and personal behaviour tendencies in handling non-optimal everyday situations causing frustration, as the displayed response might show similarities to collaborating with a suboptimal robot. Specifically, we used the 5-point Likert scale based assessments on Frustration Discomfort Scale [28] and Frustrative Nonreward Responsiveness Scale [29].

The post-interaction questionnaire containing 17 items (5-point Likert scale format) aims to evaluate the participants' interaction experience with different robot behaviours, focusing on Frustration, Satisfaction, Perceived Usability, Perceived Usefulness and Behavioural Intention to Use, following established prior works [30], [31], [32], [33], [34].

V. Discussion

HRI-SENSE[1] contains features extracted from six hours of recordings collected during 146 human-robot interaction sessions. 18 participants involved in our experiments. Each participant interacted with the robot configured with 3 different decision models over 2–3 sessions per configuration. Each session involved performing 5 verbal interactions and object handovers in the context of the Burger Assembly task.

A. User Reaction Findings

The human-robot interactions recorded in our data contain various elements that may influence the participant's reactions and impressions. Considering the robot's physical performance, object pick-and-place sequences are mostly executed following user expectations. However, errors may occasionally occur, primarily in the form of accidentally toppling the target object attempting the pickup, or failing to grasp the object properly. Occasionally, although the robot's actions are correct in handing over the target object, the robot arm movements may be unexpected or irregular in terms of trajectory, speed, or acceleration. Such irregularities can elicit reactions from participants, ranging from surprise to frustration.

Additionally, considering the robot's intention's correctness, whether the robot hands over (or attempts to hand over) the expected target object or an incorrect one, severely affects the interactions' success and the user's reactions, too.

Factors related to verbal interaction were also found to influence user reactions. Overall, whether a verbal interaction was perceived smooth, complex or even unnecessary is crucial for its evaluation. How users receive different forms of verbal interactions (e.g. confirmation queries, instruction queries, direct actions) depending on the timing, frequency and duration of these may serve as an indication. During verbal interactions with the robot, on multiple occasions the time lag of the used Speech-to-Text model [24] caused the users' responses to be ignored by the robot (in case of rapid user responses), forcing them to repeat their response. This unexpected factor may also influence the user's impressions reflected in the data.

B. Anticipated Use Cases

Our HRI-SENSE collection of multimodal human-robot interaction data focusing on collaborative assembly and verbal interactions eliciting various human reactions to robot behaviour offers opportunities in several research directions.

1) Detection of Errors in Collaborations: Through analysing data collected from various modalities at the time when a robot error of certain type and severity has been made, learning models may be trained to identify or classify such situations to be suitably handled.

2) Implicit Human Feedback and Response Analysis: Continuously observable implicit multimodal human feedback (e.g. facial reactions, body movements or action choices) in HRI-SENSE may serve a valuable role in reasoning about human feedback during a human-robot interaction. Further on, the displayed user responses can contribute to understanding the user's impressions and psychological state (e.g. frustration, satisfaction, confusion, etc.) throughout an interaction.

3) Learning User-Aware Human-Robot Interactions: Focusing on how robots can accommodate their behaviour to user preferences, our dataset provides a corpus of multimodal human-robot interaction training data that can enable mobile robots to learn user specific desirable interaction behaviours. For instance, the identified helpful and desired robot behaviour patterns may be learnt using Behaviour Cloning, or Imitation Learning methods may be used to learn reward functions or optimal policies for robot decision models.

4) Studying Human-Robot Interaction Patterns: Human-Robot interactions tend to involve a number of modalities, several of which we attempt to capture in our dataset. By investigating correlations and patterns between modalities, general user preferences on robot behaviour may be identified in human-robot interactions, enabling researchers to tune future interaction model developments accordingly.

[1]Publicly available at https://doi.org/10.5281/zenodo.14267885. This work was supported by UK Research and Innovation [EP/S024298/1].

References

[1] A. Meissner, A. Trübswetter, A. S. Conti-Kufner, and J. Schmidtler, "Friend or Foe? Understanding Assembly Workers' Acceptance of Human-robot Collaboration," *ACM Transactions on Human-Robot Interaction*, 2021.

[2] Y. Yan and Y. Jia, "A Review on Human Comfort Factors, Measurements, and Improvements in Human–Robot Collaboration," *Sensors*, 2022.

[3] A. Weidemann and N. Rußwinkel, "The Role of Frustration in Human–Robot Interaction – What Is Needed for a Successful Collaboration?" *Frontiers in Psychology*, 2021.

[4] G. Bansal, B. Nushi, E. Kamar, E. Horvitz, and D. S. Weld, "Is the most accurate AI the best teammate? Optimizing AI for teamwork," *Proceedings of the AAAI Conference on Artificial Intelligence*, 2021.

[5] K. Candon, Z. Hsu, Y. Kim, J. Chen, N. Tsoi, and M. Vázquez, "Perceptions of the helpfulness of unexpected agent assistance," in *Proceedings of the 10th International Conference on Human-Agent Interaction*. Association for Computing Machinery, 2022.

[6] C. Basu, E. Bıyık, Z. He, M. Singhal, and D. Sadigh, "Active learning of reward dynamics from hierarchical queries," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE Press, 2019.

[7] W. B. Knox and P. Stone, "Interactively shaping agents via human reinforcement: The TAMER framework," in *Proceedings of the Fifth International Conference on Knowledge Capture*. Association for Computing Machinery, 2009.

[8] J. Dollard, N. E. Miller, L. W. Doob, O. H. Mowrer, and R. R. Sears, *Frustration and Aggression.* Yale University Press, 1939.

[9] L. Berkowitz, "Frustration-aggression hypothesis: Examination and reformulation." *Psychological Bulletin*, 1989.

[10] A. Vinciarelli, M. Pantic, and H. Bourlard, "Social signal processing: Survey of an emerging domain," *Image and Vision Computing*, 2009.

[11] A. Ben-Youssef, C. Clavel, S. Essid, M. Bilac, M. Chamoux, and A. Lim, "UE-HRI: A new dataset for the study of user engagement in spontaneous human-robot interactions," in *Proceedings of the 19th ACM International Conference on Multimodal Interaction*. Association for Computing Machinery, 2017.

[12] O. Celiktutan, E. Skordos, and H. Gunes, "Multimodal human-human-robot interactions (MHHRI) dataset for studying personality and engagement," *IEEE Transactions on Affective Computing*, 2019.

[13] P. E. Paredes, F. Ordonez, W. Ju, and J. A. Landay, "Fast & furious: Detecting stress with a car steering wheel," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, 2018.

[14] D. B. Jayagopi, S. Sheikhi, D. Klotz, J. Wienke, J.-M. Odobez, S. Wrede, V. Khalidov, L. Nguyen, B. Wrede, and D. Gatica-Perez, "The vernissage corpus: A multimodal human-robot-interaction dataset," 2012.

[15] K. Candon, N. C. Georgiou, H. Zhou, S. Richardson, Q. Zhang, B. Scassellati, and M. Vázquez, "REACT: Two datasets for analyzing both human reactions and evaluative feedback to robots over time," in *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*. Association for Computing Machinery, 2024.

[16] Y. Cui, Q. Zhang, B. Knox, A. Allievi, P. Stone, and S. Niekum, "The EMPATHIC framework for task learning from implicit human feedback," in *Proceedings of the 2020 Conference on Robot Learning*. PMLR, 2021.

[17] M. Stiber, R. H. Taylor, and C.-M. Huang, "On using social signals to enable flexible error-aware HRI," in *Proceedings of the 2023 ACM/IEEE International Conference on Human-Robot Interaction*. Association for Computing Machinery, 2023.

[18] B. A. Newman, R. M. Aronson, S. S. Srinivasa, K. Kitani, and H. Admoni, "HARMONIC: A multimodal dataset of assistive human–robot collaboration," *The International Journal of Robotics Research*, 2022.

[19] J. Pagès, L. Marchionni, and F. Ferro, "TIAGo: The modular robot that adapts to different research needs," 2016.

[20] D. Coleman, I. A. Sucan, S. Chitta, and N. Correll, "Reducing the barrier to entry of complex robotic software: a moveit! case study," vol. abs/1404.3785, 2014.

[21] S. Stein and S. J. McKenna, "Combining embedded accelerometers with computer vision for recognizing food preparation activities," in *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. Association for Computing Machinery, 2013.

[22] Y. Ben-Shabat, X. Yu, F. Saleh, D. Campbell, C. Rodriguez-Opazo, H. Li, and S. Gould, "The IKEA ASM Dataset: Understanding People Assembling Furniture through Actions, Objects and Pose," in *2021 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2021.

[23] F. Sener, D. Chatterjee, D. Shelepov, K. He, D. Singhania, R. Wang, and A. Yao, "Assembly101: A Large-Scale Multi-View Video Dataset for Understanding Procedural Activities," in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022.

[24] C. Lugaresi, J. Tang, H. Nash, C. McClanahan, E. Uboweja, M. Hays, F. Zhang, C.-L. Chang, M. G. Yong, J. Lee, W.-T. Chang, W. Hua, M. Georg, and M. Grundmann, "MediaPipe: A framework for building perception pipelines," 2019.

[25] P. Ekman and W. V. Friesen, "Facial action coding system," *Environmental Psychology & Nonverbal Behavior*, 1978.

[26] T. Baltrušaitis, P. Robinson, and L.-P. Morency, "OpenFace: An open source facial behavior analysis toolkit," in *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2016.

[27] A. Radford, J. W. Kim, T. Xu, G. Brockman, C. Mcleavey, and I. Sutskever, "Robust speech recognition via large-scale weak supervision," in *Proceedings of the 40th International Conference on Machine Learning*. PMLR, 2023.

[28] N. Harrington, "The frustration discomfort scale: Development and psychometric properties," *Clinical Psychology & Psychotherapy*, 2005.

[29] K. A. Wright, D. H. Lam, and R. G. Brown, "Reduced approach motivation following nonreward: Extension of the BIS/BAS scales." *Personality and Individual Differences*, 2009.

[30] L. H. Peters, E. J. O'Connor, and C. J. Rudolf, "The behavioral and affective consequences of performance-relevant situational variables," *Organizational Behavior and Human Performance*, 1980.

[31] P. A. Lasota and J. A. Shah, "Analyzing the effects of human-aware motion planning on close-proximity human-robot collaboration." *Human factors*, 2015.

[32] V. Venkatesh and F. D. Davis, "A theoretical extension of the technology acceptance model: Four longitudinal field studies," *Management Science*, 2000.

[33] F. D. Davis, "Perceived usefulness, perceived ease of use, and user acceptance of information technology," *MIS Quarterly*, 1989.

[34] C. Bröhl, J. Nelles, C. Brandl, A. Mertens, and V. Nitsch, "Human–Robot Collaboration Acceptance Model: Development and Comparison for Germany, Japan, China and the USA," *International Journal of Social Robotics*, 2019.