Rochester Institute of Technology

RIT Digital Institutional Repository

7-2022

A Physiological Computing System to Improve Human-Robot Collaboration by Using Human Comfort Index

Celal Savur
cs1323@rit.edu

A Physiological Computing System to Improve Human-Robot Collaboration by Using Human Comfort Index

by

Celal Savur

A dissertation submitted in partial fulfillment of the requirements for the Degree of Doctor of Philosophy in Electrical and Computer Engineering

Supervised by

Dr. Ferat Sahin

Department of Electrical and Microelectronic Engineering
Kate Gleason College of Engineering

Rochester Institute of Technology
Rochester, New York
July 2022

Approved By:

_____
Dr. Ferat Sahin
Advisor, Department of Electrical and Microelectronic Engineering

_____
Dr. Gill R Tsouri
Committee Member, Department of Electrical and Microelectronic Engineering

_____
Dr. Mujdat Cetin
Committee Member, Department of Electrical and Computer Engineering
at University of Rochester

_____
Dr. Jamison Heard
Committee Member, Department of Electrical and Microelectronic Engineering

_____
Dr. Serdar Ciftci
Honorary Committee Member, Department of Computer Engineering
at Harran University

Biographical Note

Celal Savur works primarily in the field of human-robot collaboration and physiological computing. His research is about the estimation of human comfort levels during human-robot collaboration tasks. He works with human-in-the-loop systems, where he develops machine learning models to increase fluency between human-robot and the safety of humans sharing and working in the robot workspace. Before this, he worked on a project titled American Sign Language letters recognition using surface EMG signal and studied machine learning techniques and their applications in Assistive Technologies. While obtaining his Ph.D., he worked on an NIH-funded project as a lead software engineer where he developed an Android-based app that collects data and estimates the resting heart rate of humans from a front-facing camera. He took multiple roles at the IEEE System of Systems Engineering Conference, including co-program chair and co-publication chair. He also taught K-12 students about robotics in RoboCamp@RIT during the summer. Celal's research interests are human-robot collaboration, machine learning, physiological computing, and deep learning.

Celal Savur received his bachelor in Computer Engineering from Harran University, Sanliurfa, Türkiye. There he majored in software engineering. After completing his B.S. degree, he worked as a Software Engineer for a year. He was awarded a Turkish government scholarship to pursue his master's and Ph.D. degrees. He received his M.S. in Electrical

Engineering in the field of controls and robotics from Rochester Institute of Technology (RIT).

Celal went on to pursue his Ph.D. in Electrical and Computer Engineering at RIT, where he co-leads the development of the CMCR collaboration lab at RIT under the supervision of his advisor Dr. Ferat Sahin. As the Lead Researcher and Manager of the CM Collaborative Robotics Laboratory (CMCRL), he worked with and mentored a team of graduate students on research addressing challenges of experiment design, data collection system, and programming robots for HRC tasks. He has worked on multiple sponsored projects relating to automation using collaborative robots in the industry. He has published his research in leading conferences such as the IEEE International Conference on Machine Learning and Applications (ICMLA), IEEE Systems, Man, and Cybernetics (SMC), IEEE System of Systems Engineering (SoSE), and IEEE Computer Vision and Pattern Recognition (CVPR).

✉ c.savur@gmail.com   👤 celalsavur.com   🏛 mabl.rit.edu/cmcr   ⭕ github.com/csavur

in linkedin.com/in/csavur   ⓘⅅ orcid.org/0000-0001-8767-3598

Abstract

Kate Gleason College of Engineering

Rochester Institute of Technology

Degree: Doctor of Philosophy

Program: ECE PhD

Author: Celal Savur

Advisor: Dr. Ferat Sahin

Dissertation Title: A Physiological Computing System to Improve Human-Robot Collaboration by Using Human Comfort Index

Fluent human-robot collaboration requires a robot teammate to understand, learn, and adapt to the human's psycho-physiological state. Such collaborations require a physiological computing system that monitors human biological signals during human-robot collaboration (HRC) to quantitatively estimate a human's level of comfort, which we have termed in this research as *comfortability index* (CI) and *uncomfortability index* (UnCI). We proposed a human comfort index estimation system (CIES) that uses biological signals and subjective metrics. Subjective metrics (*surprise*, *anxiety*, *boredom*, *calmness*, and *comfortability*) and physiological signals were collected during a human-robot collaboration experiment that varied the robot's behavior. The emotion circumplex model is adapted to calculate the CI from the participant's quantitative data as well as physiological data. This thesis developed a physiological computing system that estimates human comfort levels from physiological by using the circumplex model approach. The data was collected from multiple experiments and machine learning models trained, and their performance was evaluated. As a result, a subject-independent model was tested to determine the robot behavior based on human comfort level. The results from multiple experiments indicate that the proposed CIES model improves human comfort by providing feedback to the robot. In conclusion, physiological signals can be used for personalized robots, and it has the potential to improve safety for humans and increase the fluency of collaboration.

To my late father and mother...
(10/7/2021 ∼ 10/8/2021)

Acknowledgments

I would like to thank my advisor, Dr. Ferat Sahin, who over the past years has been a guide for many aspects of my life including academic, career, and personal affairs. None of this work would have been possible without him. Dr. Sahin, thank you so much and may our personal and professional relationship continue to grow stronger.

I would also like to thank:

- My parents, for their love and support for all my life. You have influenced me in so many ways.

- My wife, Yagmur Akarsu Savur for being so patient during the numerous long nights in the lab, and keeping me positive during times of adversity.

- My brothers/sisters and friends for their continued support and encouragement. It has been quite the journey, but your support has kept me on track.

- Shitij Kumar, for being an excellent colleague, mentor, and great friend. Brainstorming and discussing human-robot collaboration was priceless.

- Kamil Bukum, for being an excellent friend and co-worker.

- My committee, for reviewing my work and suggesting improvements.

- The Multi-Agent BioRobotics Laboratory and Collaborative Robotics Research Lab (CMCR) research members, for the wonderful lab atmosphere, hours of proofreading, and late-night collective brainstorming sessions (*Karthik Subramanian*, *Anmol Modur*, *Sarthak Arora*, *Justin Kon*, *Xavier Tarr*, *Saurav Sigh*, *Nick Nguyen*, *Joseph Armstrong*, *James May* and others that I did not mention).

- Abhiraj Patil, for his countless hours of data collection.

- Electrical and Microelectronic Engineering, and Ph.D. in Electrical and Computer Engineering Program' faculty and staff for their assistance. I am especially grateful to *James Stefano*, *Ken Snyder*, *Patricia Vicari*, *Rebecca Ziebarth*, *Sara Gaudioso*, and *Vincent Antonicelli*.

- For the rest not mentioned, I thank you all.

List of Contributions

Highlights of the Work

- Proposed a novel approach to estimate human comfort index from physiological signals.

- Integration of CIES with human-robot collaboration experiment to improve comfortability, safety, and fluency.

- Two datasets are going to be publicly available for the robotic community.

- A framework and implementation of a system for collecting human physiological signals during HRC tasks and using it for the development of an affective computing system.

- Multiple experiments were conducted and proposed model performances were evaluated.

Publications

- **Savur, C.**, & Sahin, F. (2022). "Adaptive Human-Robot Collaboration Using Human Comfort Index". (TBD)

- **Savur, C.**, & Sahin, F. (2022). "Exploration of Human Comfort Index Estimation in HRC". (TBD)

- **Savur, C.**, Heard, J., & Sahin, F. (2022). "Human Comfortability Index Estimation in Industrial Human-Robot Collaboration". Transactions On Human-Machine Systems. (Under Review)

- **Savur, C.**, & Sahin, F. (2022). "Survey on Physiological Computing in Human-Robot Collaboration". Transactions On Human-Machine Systems. (Under Review)

- **C. Savur**, R. Dautov, K. Bukum, X. Xia, J. -P. Couderc and G. R. Tsouri, "Monitoring Pulse Rate in the Background using Front Facing Cameras of Mobile Devices," in IEEE Journal of Biomedical and Health Informatics, 2022, doi: 10.1109/JBHI.2022.3197076.

- K. Bukum, **C. Savur**, Gill R. Tsouri; "Deep Learning Classifier for Advancing Video Monitoring of Atrial Fibrillation". (CVPR) Workshops, 2022, pp. 2211-2219

- M. Sahin and **C. Savur**, "Evaluation of Human Perceived Safety during HRC Task using Multiple Data Collection Methods," 2022 17th Annual System of Systems Engineering Conference (SOSE), 2022, pp. 465-470,
doi: 10.1109/SOSE55472.2022.9812693.

- N. Deshmukh, **C. Savur** and F. Sahin, "iOS Based Pose Estimation and Gesture Recognition for Robot Manipulation," 2022 17th Annual System of Systems Engineering Conference (SOSE), 2022, pp. 438-443,
doi: 10.1109/SOSE55472.2022.9812685.

- A. Ambrose, **C. Savur** and F. Sahin, "Low Cost Real Time Location Tracking with Ultra-Wideband," 2022 17th Annual System of Systems Engineering Conference (SOSE), 2022, pp. 445-450, doi: 10.1109/SOSE55472.2022.9812679.

- S. P. Kandikere, **C. Savur** and F. Sahin, "System Integration of a Tour Guide Robot," 2022 17th Annual System of Systems Engineering Conference (SOSE), 2022, pp. 431-436, doi: 10.1109/SOSE55472.2022.9812680.

- S. Kumar, **C. Savur** and F. Sahin, "Survey of Human–Robot Collaboration in Industrial Settings: Awareness, Intelligence, and Compliance," in IEEE Transactions on Systems, Man, and Cybernetics: Systems, vol. 51, no. 1, pp. 280-297, Jan. 2021, doi: 10.1109/TSMC.2020.3041231.

- Couderc, J.-P. Y., Tsouri, G., **Savur, C.**, Xiajuan Xia, J. D., Betty Mykins, K. B., & Hall, B. W. (2021). "Monitoring Atrial Fibrillation Patients Using Active Contactless Videoplesthymography Implemented on a Personal Tablet". Heart Rhythm, 18(8), S265. https://doi.org/10.1016/j.hrthm.2021.06.659

- S. Kumar, **C. Savur** and F. Sahin, "Survey of Human–Robot Collaboration in Industrial Settings: Awareness, Intelligence, and Compliance," in IEEE Transactions on Systems, Man, and Cybernetics: Systems, vol. 51, no. 1, pp. 280-297, Jan. 2021, doi: 10.1109/TSMC.2020.3041231.

- K. Subramanian, **C. Savur** and F. Sahin,"Using Photoplethysmography for Simple Hand Gesture Recognition," 2020 IEEE 15th International Conference of System of Systems Engineering (SoSE), 2020, pp. 307-312,
doi: 10.1109/SoSE50414.2020.9130489.

- **C. Savur**, S. Kumar and F. Sahin,"A Framework for Monitoring Human Physiological Response during Human Robot Collaborative Task," 2019 IEEE International

Conference on Systems, Man and Cybernetics (SMC), Bari, Italy, 2019, pp. 385-390. doi: 10.1109/SMC.2019.8914593

- **C. Savur**, S. Kumar, S. Arora, T. Hazbar and F. Sahin, "HRC-SoS: Human Robot Collaboration Experimentation Platform as System of Systems," 2019 14th Annual Conference System of Systems Engineering (SoSE), Anchorage, AK, USA, 2019, pp. 206-211. doi: 10.1109/SYSOSE.2019.8753881

- S. Kumar, **C. Savur** and F. Sahin, "Dynamic Awareness of an Industrial Robotic Arm Using Time-of-Flight Laser-Ranging Sensors," 2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Miyazaki, Japan, 2018, pp. 2850-2857. doi: 10.1109/SMC.2018.00485

- R. Dautov, **C. Savur** and G. Tsouri, "On the Effect of Face Detection on Heart Rate Estimation in Videoplethysmography," 2018 IEEE Western New York Image and Signal Processing Workshop (WNYISPW), Rochester, NY, 2018, pp. 1-5. doi: 10.1109/WNYIPW.2018.8576439

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In this chapter, an overview of Physiological Computing and its application in Human-Robot collaboration (HRC) is presented.

1.1 Physiological Computing in HRC

Physiological computing is an interdisciplinary field that focuses on human physiological signals to simulate a human's psycho-physiological state representation that can recognize, interpret, and process, to dynamically change user's psycho-physiological state. Human-computer interaction, brain-computer interaction, and affective computing are part of physiological computing [5]. The motivation of physiological computing is to simulate the psycho-physiological state of humans, and allow a program/robot to change its behavior accordingly; act appropriately, and respond smartly.

Physiological computing affects many fields, such as Human-Computer Interaction, E-learning, Automotive, Healthcare, Neuroscience, Marketing, and Robotics [5]. As an example of E-learning, physiological computing can help the tutor to modify the presentation style based on students' affective states such as interest, boredom, and frustration. In the automotive industry, HCI can be used as an alert system to alert unattentive drivers about the danger surrounding their vehicles. In social robotics, physiological computing can help robotic pets to understand human physiological states and improve their interaction.

According to NSF Research Statement for Cyber Human Systems (2018-2019), "*improve the intelligence of increasingly autonomous systems that require varying levels of supervisory control by the human; this includes a more symbiotic relationship between human and machine through the development of systems that can sense and learn the human's cognitive and physical states while possessing the ability to sense, learn, and adapt in their environments*" [6]. Thus, to ensure trust and safety between the human and the robot interactively, the robot should sense a human's cognitive and physical state, which will help build trust.

In a human-robot interaction setup, a change in a robot's motion can affect the human physiological state. Experiments such as [7] and [8] revealed that the robot's trajectory has an effect on human skin conductivity. The literature review in [9] highlights the use of the 'psycho-physiological' method to evaluate human response and behavior during human-robot interaction. In our opinion, continuous monitoring of physiological signals during human-robot tasks is the first step in quantifying human trust in automation. The inferences from these signals and incorporating them in real-time to adapt robot motion can enhance human-robot interaction. Such a system capable of 'physiological computing' will result in a closed human-in-the-loop (also known as 'biocybernetics loop') system where both human and robot in an HRC setup are monitored, and information is shared. This approach could significantly improve robot interaction, which would improve trust in automation and increase productivity.

According to Fairclough, physiological computing can be divided into two categories. The first category is a system of sensory-motor function, which is related to extending body schema [10]. In this category, the subject is aware that they are in control. For example, an Electromyography (EMG) sensor placed on the forearm can be used as an alternative method for typing [11], or it can control a prosthetic arm. Similarly, Brain-Computer Interaction (BCI) provides an alternative way to type via Electroencephalogram (EEG) headset.

The second category is about creating a representation of physiological state through monitoring and responding to simultaneous data originating from psycho-physiological interaction in the central nervous system [10]. This category is also known as biocybernetics adaptation. The biocybernetics adaptation needs to detect spontaneous changes in the user's physiological state. Thus, the system can respond to this change. The biocybernetics adaptation has many applications, such as emotional detection, anxiety detection, and mental workload estimation. For example, based on mental workload, the amount of data displayed can be filtered to reduce the workload in flight simulation. A computer game can change difficulty levels based on the player's anxiety levels.

1.2 Human's Comfort Index Estimation System (CIES)

There are various ways to tackle safety. The state-of-the-art technique is to use safety zones (thresholds) such as laser ranging and pressure pads. Although these approaches provide safety, it is not efficient in terms of productivity. Because the robot needs to stop at any time a human enters the robot's workspace.

In an alternative approach, Kumar et al. [12] proposed a Dynamic Speed and Separation (DSS) that uses minimum speed and separation monitoring. The DSS has proven that it is better than the state-of-the-art methods in terms of the efficiency of the robot. However, since this method allows the robot to get close to the human, it may trigger discomfort in the human due to its speed, acceleration, and size. The approaches discussed above can detect a human when it enters its workspace; however, they cannot detect human comfortability.

The human comfort index estimation system (CIES) aims to estimate the human comfort index via physiological signals to control a robot's speed and separation through comfort index. Fig. 1.1 shows the overall human in the loop system. This approach is also called implicit communication between the human and the robot.
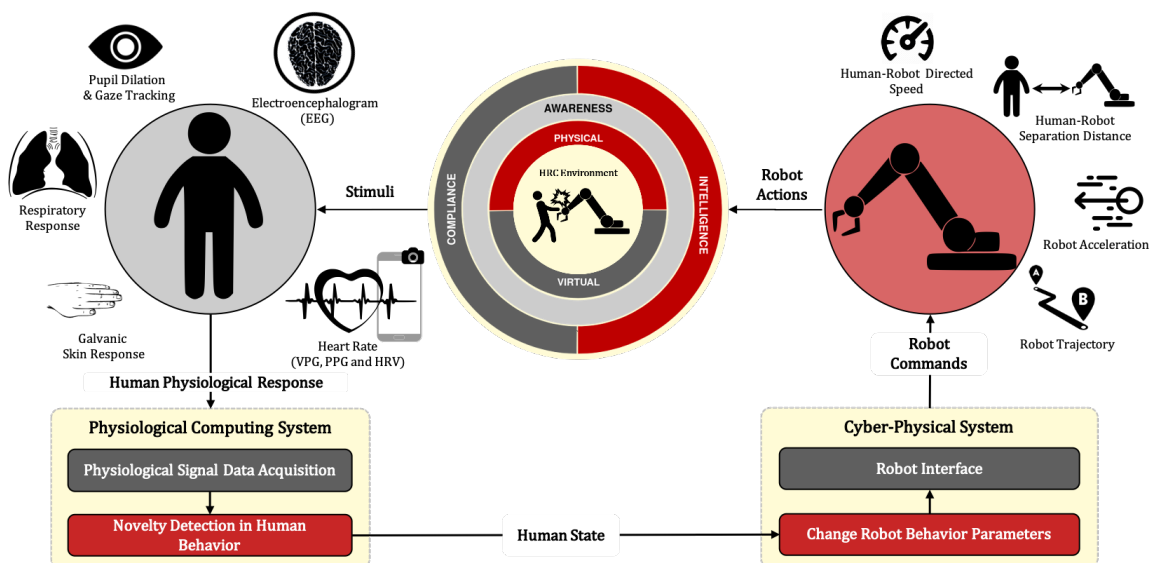
Figure 1.1: The figure shows the human in the loop system where the human and the robot share the same workspace.

As seen in Fig. 1.1, the robot can control its speed, acceleration, trajectory, and distance between human and robot (limited control). Any sudden change in these actions may trigger discomfort in a human. Especially if the person does not have any prior experience with robots. CIES is a physiological computing system that acquires physiological signals from humans in real-time and processes these signals to estimate the comfort index, which will provide a robot intelligence system to change its behavior (*adaptation*) to make people comfortable.

This thesis aims to develop a physiological computing system that monitors human physiological responses in real-time during a human-robot collaboration task to estimate a human comfort index and provide this information to the robot as actionable feedback. In addition, the model has to be adaptive because as human gets used to a task ("learning curve [13]") they feel more comfortable thus, such a system should not trigger a false alarm. This research also highlights the aspects and challenges of collecting human-physiological signals during a human-robot experiment. It underscores the importance of a controlled HRC experiment design, event marker generation related to both the human and the robot, and the synchronization of data collected. Detecting stress, anxiety, and cognitive load is not in the scope of this research.

Chapter 2

Background and Related Work

This chapter provides background information and a literature review of related research on physiological computing for HRC in the industry. A brief summary of physiological signals and their characteristics, data collection methods, and data labeling are presented. Following these sections, a detailed categorization of related research in Human-Robot Collaboration based on physiological computing is presented. Finally, dynamic speed and separation algorithms are presented.

2.1   Physiological Signals

Representation of a human's psycho-physiological state requires a complex analysis of physiological signals. Hence, to estimate the psycho-physiological state a variety of physiological signals were used such as Electrocardiogram (ECG), Photoplethysmography (PPG), Galvanic Skin Response (GSR), Electroencephalography (EEG), Electromyography (EMG), respiration rate (RSP), and Pupil dilation, as shown in Fig. 2.1.
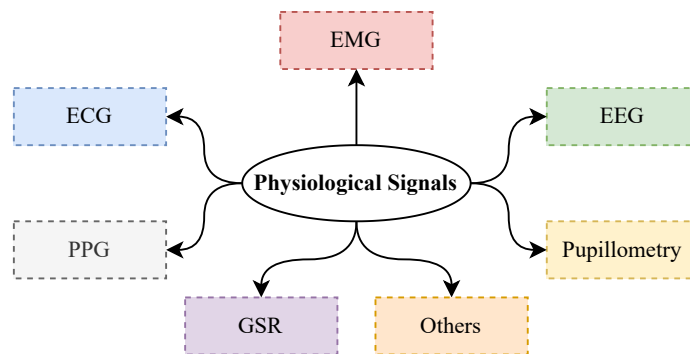
Figure 2.1: Physiological signals

2.1.1 Electroencephalogram

Electroencephalogram (EEG) is a method to measure the electric activity of the neurons in the brain. EEG signal is a complex signal; thus, extensive research is presently being conducted in the field of neuroscience psychology. The EEG signal can be collected using invasive or non-invasive methods. The non-invasive method measures signal over the skin which is widely used to collect the brain's activity. The invasive method measures signal under the skin (in some cases under the skull [14]) and it becomes more available and its results are promising these days.

The researchers categorized EEG signals based on the frequency band: delta band (1-4 Hz), theta band (4-8 Hz), the alpha band (8-12 Hz), the beta band (13-25 Hz), and gamma band (> 25 Hz). Results showed that the delta band has been used in several studies such as sleeping [15]. The theta band is related to brain processes, mostly mental workload [16, 17]. It has been shown that alpha waves are associated with relaxed wakefulness [18], and beta waves are associated with focused attention or anxious thinking [19].

It can be argued that wearing an EEG cap while working can be uncomfortable. However, it must be noted that in industry, workers are required to wear a helmet or hat. With the advent of IoT systems and wireless communication, small-scale EEG sensors can be embedded into the headband [20].

2.1.2 Electrocardiogram

Electrocardiogram (ECG) is the method that records the electrical signal generated by the heart. ECG is a non-invasive method that was developed in 1902 by Dr. Willem Einthoven [21]. ECG signals can provide useful information about the heart and its diseases such as Atrial Fibrillation, Ischemia, and Arrhythmia.

ECG signal is a repeated heartbeat that is defined by a QRS complex. The QRS is the most visible part of the signal, and it lasts 0.06 to 0.10 seconds for adults [22]. Heart Rate (HR) is the number of repeated R peaks in a minute. The ECG is not the only way to determine the HR rate; however, it is the most accurate and trusted signal to calculate since it is directly calculated from the polarization and depolarization of the heart. The other valuable signal that is extracted from ECG is Heart Rate Variability (HRV). It is defined

by the time elapsed between two consecutive R peaks. The HRV helps to indicate heart disease such as Atrial Fibrillation (AF). The HRV varies based on human activity state; for example, during exercise, HRV is low; in the contrast, it is high during resting state. However, it is more important to pay attention to a sudden change in HRV since this can be an indication of either heart disease or emotional change. Chio et al. showed that there is a significant positive correlation between emotion and HRV; thus, they suggest HRV can be used in emotional detection [23].

2.1.3 Photoplethysmography

Photoplethysmography (PPG) is a convenient and inexpensive method that is an alternative to the ECG approach that measures heart rate and heart rate variability. PPG technology uses a light source and photon detector on the human's skin to measure the amount of light reflection (volumetric variations of blood circulation). Unlike ECG signal PPG signal uses inter-beat-interval (IBI) for heart rate and HRV calculation. ECG electrode placement is complicated and is prone to noise from movement, Lu et al. showed that the PPG method can be used for heart rate and heart rate variability [24].

2.1.4 Galvanic Skin Response/Electrodermal Activity

Galvanic Skin Response (GSR) or Electrodermal Activity (EDA) is a physiological signal obtained by measuring skin conductivity. The conductivity of skin changes whenever sweat glands are triggered. This phenomenon is an unconscious process controlled by the sympathetic division of the autonomic nervous system. The sympathetic division is activated when exposed to emotional moments (fear, happiness, joy) or undesirable situations [25]. Hence, it triggers the sweat glands, heart, lungs, and other organs; as a result, one's hands become sweaty, heart rate increases, and one begins breathing excessively.

The GSR signal is used in various fields such as physiological research, consumer neuroscience, marketing, media, and usability testing. The GSR signal is a non-invasive method that uses two electrodes to detect emotional arousal. It is commonly placed on the palms of the hands, fingers, or foot soles.

The GSR signal has two components: tonic level, Skin Conductance Level (SCL), and

phasic response, known as Skin Conductance Response (SCR). The tonic level changes and varies slowly. It also may vary between individuals and their skin moisture level. Thus, it does not provide valuable information about the sympathetic division. Unlike the tonic level, the phasic response changes and alternates faster. These changes and deviations are directly related to reactions coming from the sympathetic division under the autonomic nervous system. The phasic response is sensitive to emotional and mental load [26, 27] and has a delay of 1 to 5 seconds after the onset of event stimuli. The GSR signal is usually used with other signals such as eye-tracking, heart rate, or respiration rate to interpret the autonomic nervous system better [25]. Thus, the phasic response provides essential information about the physiological state.

The GSR signal provides valuable information about the strength of arousal, whether it is decreasing or increasing. However, positive and negative events (moments) may have similar GSR signal output. Therefore, the GSR signal should be used with another sensor such as EEG, ECG, EMG, or Pupil dilation [25].

2.1.5 Pupil Dilation/Gaze Tracking

Human visual attention can be detected by eye movement, and this information can be used in many studies such as neuromarketing and psychological studies [28]. The gaze tracking provides information about the subject's field of view. This information also can be used in other fields, such as robotics. For example, if the robot knows a person is not paying attention to a sensitive operation, the collaborative robot can take some action to notify the person.

The eyes do not only provide information about the direction of focus, but they also provide information about pupil dilation, which is a measurement of change in pupil diameter. Although pupil dilation can be caused by ambient or other light intensity changes in the environment, it has been shown that it can dilate from emotional change as well [29, 30, 31].

2.1.6 Electromyography

Electromyography (EMG) is a non-invasive method that measures electrical activity generated by muscles. EMG has been used in biocybernetics loop applications as a control input for a system or robot [32]. Another example of EMG is sudden emotional changes or reactions [33, 34] from facial muscles.

2.1.7 Physiological Signal Features

Unlike deep learning models, classical Machine Learning (ML) algorithms usually need features for training. Although deep learning algorithms can learn from the raw data, they require large datasets, which are typically difficult to obtain. Classical ML algorithms require features to be extracted from signals. There are different ways of extracting features from a signal: time domain, frequency, and non-linear features. In this section, commonly used features are listed in Table 2.1 based on signal types. Researchers also use discrete wavelet transforms for feature extraction. Al-Qerem et al. provided extensive work that discussed the usage of wavelet transforms for EEG signal [35].

2.2 Data Collection Methods

Data collection is the most important and time-consuming part of physiological computing systems; high signal quality will be beneficial for extracting desired information from signals. There are a variety of ways to collect physiological signals. In this section, we cover the most commonly used methods which are as shown in Fig. 2.2.

2.2.1 Baseline

The objective of the baseline method is to define what the normal is. The baseline is collected before the experiment is started. Thus, the baseline information can be used during the experiment for comparison or to create a rule that would capture target situations. For example, anxiety estimations during computer games [36, 37]. The subject's physiological signals are recorded before the experiment that was marked as a baseline, then these signals

Table 2.1: Commonly used physiological metrics extracted from the ECG, GSR, Pupillometry, and EEG signals

Signal Type	Feature	Description
ECG	MeanNN	The mean of the RR intervals.
	SDNN	The standard deviation of the RR intervals.
	RMSSD	The square root of the mean of the sum of successive differences between adjacent RR intervals.
	SDSD	The standard deviation of the successive differences between RR intervals.
	pNN50	The proportion of RR intervals greater than 50ms, out of the total number of RR intervals.
	pNN20	The proportion of RR intervals greater than 20ms, out of the total number of RR intervals.
	LF	The spectral power of low frequencies.
	HF	The spectral power of high frequencies.
GSR	Amp. Mean	Mean value of peak amplitude
	Amp. Std	Standard deviation of peak amplitude
	Phasic Mean	Mean value of phasic signal
	Phasic Std	Standard deviation of phasic signal
	Tonic Mean	Mean value of tonic signal
	Tonic Std	Standard deviation of tonic signal
	Onset Rate	Number of onset per minute
Pupillometry	Pupil Mean	Mean value of pupil signal
	Pupil Std	Standard deviation of pupil signal
EEG	MAV	Mean Absolute Value
	ZC	Zero crossing
	SSC	Slope Sign Changes
	SKE	Skewness of EEG signal
	Kurtosis	Kurtosis of EEG signal
	Entropy	Entropy of EEG signal
	SEntropy	Spectral entropy of EEG signal

are used to create thresholds and make decisions during the game. For example, the level of game difficulty will change automatically according to the anxiety felt by the players. When the player feels anxiety, the game difficulty was reduced to make the experience more desirable. This method is mostly used in biocybernetics adaptation applications.

2.2.2 Pre-Trial

Unlike the baseline data collection method, pre-trial data is collected for comparison before each trial. This approach captures a physiological state before the trial. As an example, Dobbins et al. [38] collected a questionnaire and physiological signals before-trial and
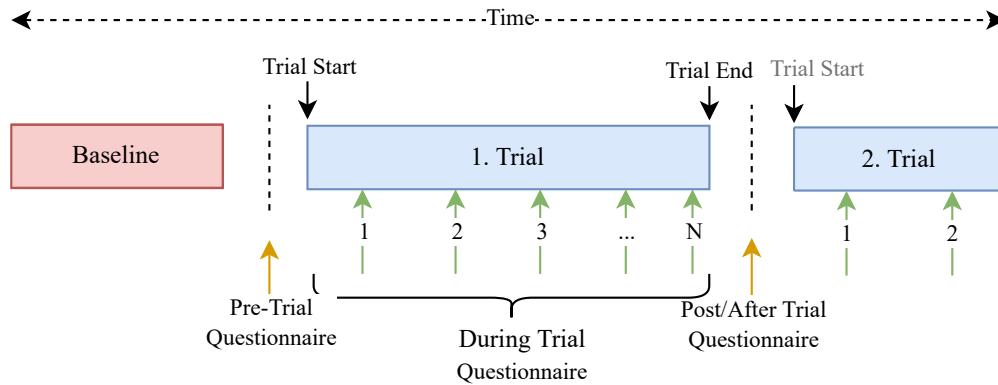
Figure 2.2: Four data collection methods: Baseline, Pre, During, and Post/After trial.

after-trial for five working days from the participant to measure the stress level of the driver during their commute. This allows the investigator to capture the difference between before and after the trial. Hence, this information would provide useful information about the daily ride. Although this approach helps in comparison, it requires data collection before and after each trial. It is important to mention that this method can be overwhelming for the participants to answer the same questions multiple times.

2.2.3 Post/After-Trial

The post/after trial data collection is the most used technique for data collection where a visual stimulus is shown to the subject. The subject then evaluates the stimuli by answering a post/after trial questionnaire. For example, data is collected while the subject performs an action after the instruction was displayed on the screen [32].

2.2.4 During-Trial

During-trial is a method where the participant is asked the same question throughout the ongoing trial as shown in Fig. 2.2. This type of data collection is useful when during-trial progress is important. For example, Sahin et al. [4] collected perceived safety in two methods: during-trial and after-trial. Their analysis showed that during-trial data collection provides additional information compared to the after-trial method.

There are two important aspects of the during-trial data collection method. The first

one is to limit the number of questions. As the trial is still ongoing, asking many questions may break the integrity of the experiment. The second is making data input methods easier. Instead of using paper and pen to collect data from the participant, it would be easier to have an application that allows participants to answer questions using a tablet. As an alternative to using a tablet app for data collection, recording participant audio feedback during trials may improve the data collection for the during-trial approach.

In conclusion, although collecting data during trials provide additional information, it may result in participant boredom if there are many questions.

2.3 Data Labeling

After data collection, physiological signals need to be labeled. In some cases, the labeling can be cumbersome, especially in biocybernetics adaptation. This section will discuss commonly used data labeling techniques that are shown in Fig. 2.3.
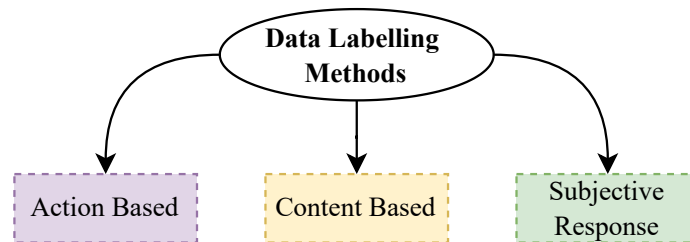
Figure 2.3: Data Labeling Methods

2.3.1 Content/Action Related

Content/Action related labeling is commonly used in a visual stimuli type of experiment [39, 40, 41]. In the visual experiment, the exact time of the shown image or video is known. Thus, the physiological signal can easily be labeled with a corresponding label. Similarly, in the action-related experiment, the time the subject repeats the gesture/action is known; thus, a window that captures the gesture can be labeled accordingly [11]. Savur et al. [11] talked about the critical aspects of data collection and labeling in HRC settings. They provided case studies for a human-robot collaboration experiment that has built signal synchronization and automatic event generation.

Content/action labeling is the simplest way of labeling, and it can be done during the data collection process. Thus, this method is widely adopted in different fields including physiological study, marketing, emotion detection, and other related factors.

2.3.2   Subjective Labeling

The questionnaire is very common in quantitative research; thus, it is widely used in HRC studies. In human-robot collaboration research, questionnaires play a key role. For example, Kumar et al. [12] used subjective response to evaluate their proposed speed and separation monitoring methods with state-of-the-art methods. Similarly, in emotion detection research, one scene may trigger different emotions; thus, most of the researchers use it for subjective evaluation [42]. Dobbins et al. [38] asked the subject to answer pre and post-survey, where the subject answered a survey before and after each experiment. Then both surveys were used quantitatively to subtract from each other, then based on the result, the trial was marked as positive, negative, or neutral.

The questionnaire helps to quantify the subject's preferences and evaluate the proposed methodology. Although it is common to ask a questionnaire, there is not a common set of questions that researchers follow [43]. In general, researchers create their own set of questions or expand upon an existing one that targets their underlying hypothesis. Below commonly used questionnaires in HRC are provided.

- **Godspeed** was designed to standardize measurement tools for HRI by Bartneck et al. [44]. The Godspeed focused on five measurements: anthropomorphism, adaptiveness, intelligence, safety, and likability. The Godspeed is commonly used, and it has been translated into different languages.

- **NASA TLX** was designed to measure subjective workload assessment. It is widely used in cognitive experiments. The NASA TLX measures six metrics: mental demand, physical demand, temporal demand, performance, effort, and frustration [45].

- **BEHAVE-II** was developed for the assessment of robot behavior [46]. It measures the following metrics: anthropomorphism, attitude towards technology, attractiveness, likability, and trust.

- **Multidimensional Robot Attitude Scale (MRAS)** is a 12-dimensional question-naire developed by Ninomiya et al. [47]. The MRAS measures a variety of metrics such as familiarity, ease of use, interest, appearance, and social support.

- **Self-Assessment Manikin Instrument (SAM)** consists of 18 questions that measure three metrics of pleasure, arousal, dominance [26]. Unlike most surveys, SAM uses a binary selection of two opposite emotions: calm vs. excited, unhappy vs. happy, etc.

- **Negative Attitude toward Robots Scale (NARS)** was developed to measure the negative attitude toward robots in terms of a negative interaction with robots, social influence, and emotions. In addition, NARS measures discomfort, anxiety, trust, etc. [48].

- **Robot Social Attributes Scale (RoSAS)** is a survey that targets to extract metrics of social perception of a robot such as warmth, competence, and discomfort [49].

- **STAXI-2** consists of 44 questions that measure state anger, trait anger, and anger expression[50].

2.4   Related Works

Table 2.2 shows some of the related works that are categorized in terms of stimuli type, data collection method, data labeling technique, and the machine learning algorithm used in the field of physiological computing.

Table 2.2: Related research tabulated based on bio-signal, number of subjects, data collection, and labeling type

Ref.	Bio-Signals	#Sub.	Stimuli	Data Col. Type	Label Type	Algorithm	Target
Kulic et al. [7]	SC, HR, EMG	36	Robot trajectory	After Trial	Subjective (Custom)	Fuzzy Inference	Arousal, Valence
Kulic et al. [51]	SC, HR, EMG	36	Robot trajectory	After Trial	Subjective (Custom)	HMM	Arousal, Valence
Nomura et al. [48]	None	240	Interaction with robot	After Trial	NARS	Statistical analysis	Negative attitude
Villania et al. [37]	Control Robot arm	21	Interaction with robot	Baseline	Subjective (Custom)	Thresholding	Stress
Landi et al. [52]	HRV (Smartwatch)	21	Teleoperation	Baseline	Subjective (Custom)	Thresholding	Stress
Rani et al. [53]	ECG, EDA, EMG	NA	Control Mobile robot	Baseline	Subjective (Custom)	Fuzzy Inference	Affective state
Lui et al. [54]	ECG, EDA, EMG	14	Control Robot arm	Baseline	Subjective (Custom)	Regression Tree model	Affective cues
Rani et al. [36]	ECG, EDA, EMG	15	Game	Baseline	Subjective (Custom)	KNN, Bayesian	Compare Learning, Algorithm
Hu et al. [55]	EEG, GSR	31	Car Simulation	Baseline	Subjective (Custom)	LDA, LinearSVM, LR, QDA, KNN	Measuring Trust
Rani et al. [56]	ECG, ICG, PPG, Heart Sound, GSR, and EMG	15	Game	Baseline	NASA TLX	Regression Tree	Affective state
Erebak et al. [57]	None	102	Robot's Appearance	After Trial	Subjective (Custom)	Statistical Analysis	Anthropomorphism of robot
Butler et al. [58]	None	40	Mobile Robot Behavior	After Trial	Subjective (Custom)	Statistical Analysis	Psychological aspect
Rahim et al. [59]	EEG, IBI, GSR	15	Wheelchair	Baseline	STAI	ANOVA, and LDA, SVM, and SLR	Stress estimation
Dobbins et al. [60]	ECG, PPG	21	Commute (Car)	Before/After Trial	STAXI-2, UMACL	LDA, DT, kNN	Negative emotion
Ferrez et al. [61]	EEG	3	HRI	After Trial	Subjective (Custom)	Gaussian Classifiers	Error Related Potential
Ehrlich et al. [62]	EEG	6	HRI	After Trial	Subjective (Custom)	SVM	Error Related Potential
Val-Calvo et al. [63]	EEG, GSR, PPG	18	Visual	After Trial	Subjective (Custom)	Ada-Boost, Bayesian, and QDA	Arousal, Valence
Mower et al. [64]	GSR	26	HRI	-	-	KNN	User State Estimation
Novak et al. [65]	ECG, GSR, RPS, Skin Temp., EEG, and Eye tracking	10	HRI	After Trial	NASA TLX	RF	Workload
Iturrate et al. [66]	EEG	12	HRI	After Trial	NASA TLX	Reinforcement Learning	Error Signal
Ehrlich et al. [67]	EEG	13	HRI	-	Action (Key Press)	LDA	Error Signal
Salazar-Gomez et al. [68]	EEG	12	HRI	After Trial	Subjective (Custom)	LDA	Error Signal
Dehais et al. [69]	GSR, Pupil, Gaze	12	HRI (*Hand-over Task*)	After Trial	Subjective (Custom)	Statistical analysis	metrics
Sahin et al. [4]	GSR, Pupil, ECG	20	HRI	During and After Trial	Subjective (Custom)	Statistical analysis	Perceived Safety

There are many studies that focus on estimating a person's emotional state, stress level, and cognitive state through physiological signals [7, 36, 70]. In addition, there are other researchers investigating the psychological aspects of robot behavior [58].

Kulic et al. [7] present a method to calculate the danger index by using distance and relative velocity between a human and robot, and the inertia of the closest point to the human as suggested in [71]. Then the real-time calculated danger index is used to control the robot's trajectory on a real robot. Similarly, the same authors [51] tried to detect anxiety triggered by two trajectory planners. Biological signals and subjective responses were collected from subjects during the experiment. The result of the subjective responses from the experiment showed that the subjects felt less anxiety during a safe planner than the classical planner. Moreover, the researcher found that the corrugator EMG signal did not help to estimate arousal and valence. However, they have found a strong positive correlation between anxiety and speed, surprise and speed, and a negative correlation between calm and speed. Kulic and Croft in [72, 33] is an extension work of the previous study [51] showing that the Hidden Markov Model (HMM) outperforms Fuzzy inference on estimating arousal and valence from physiological signals.

Nomura et al. [48] investigated negative attitudes toward robots and developed a measurement scale called "Negative Attitude towards Robot Scale" (NARS). An interesting result from this study shows that male students have fewer negative attitudes toward interactions with robots than female students in Japan. However, the authors propose a physiological experiment to be conducted to understand the human's mental state during human-robot interaction. Also, the authors mention that this result may differ for other cultures since proxemics preferences [73] are different from culture to culture.

Villani et al. [37] introduced a framework that takes human mental health into account to simplify a task in an industrial setup to improve the interaction between humans and robots. The authors used a smartwatch to measure heart rate variability (HRV) to estimate stress by applying a threshold. In order to find the thresholds for stress and resting, a subject's HRV signal was collected just before the experiment started. In the experiment, the subject task was to navigate a mobile robot using hand gestures provided by the smartwatch IMU (Roll, Pitch, Yaw). While the subject was controlling the robot, his/her mental state was measured, and the robot's speed halved when the subject was stressed. Reducing the robot's speed extended task completion time and, as a result, reduced efficiency.

The authors of [52] analyzed the mental workload of an operator for an industrial setup where the operator teleoperated the task. The authors collected HRV signals for 2.5 minutes of resting state and 2.5 minutes of stress state ("creating stress by listening to loud music and counting numbers"). The model was used in a teleoperated task where virtual fixtures appear on the screen for the operator only based on the subject's stress level. Their system predicts stress every 2.5 minutes, which is a drawback of the system.

Villani et al. [37] tried to develop a system that estimates the affective state of a human through wearable sensors. The authors used fuzzy inference, which uses features extracted from ECG, EDA, and EMG signals. They also designed an experiment that allowed a user to communicate with a mobile robot implicitly. As a result, they found that cardiac activity is a strong indicator of anxiety.

Liu et al. [54] designed a system that used effective cues to improve human-robot interaction. The authors used various biometrics such as ECG, EDA, and EMG, where multiple features were extracted. The study consisted of two phases. The first phase was developing a regression tree model for the classification of anxiety. In the second phase, the model was used to set the game's difficulty based on the player's anxiety. As a result, the participant reported a 71% increase in satisfaction while their model was active during the task.

Rani et al. [74] compared the four most common learning algorithms, such as K-Nearest Neighbor, Regression Tree (RT), Bayesian Network, and Support Vector Machine (SVM), on biological signals to detect affect recognition. The result of the experiment shows that SVM outperforms other algorithms with an 85.81% accuracy in classification. However, RT was the fastest algorithm which makes it more suitable for real-time applications. Tan et al. [75] defined important aspects of human-robot collaboration in factory settings. In addition, two experiments were proposed as a case study. The first case study investigated the effect of robot motion speed. The second was conducted to see the effect of the human-robot distance. Results of the experiment show that mental workload is directly proportional to the robot's speed and inversely proportional to distance.

Arai et al. [76] investigated the effects of the distance from a subject and the robot's speed on the subject's mental state by using only skin conductivity sensors. As a result, research suggests that the distance between an operator and a robot should be more than two meters, and the speed of the end effector to be no more than 500 mm/s. In addition, Arai

et al. also discovered that notifying the operator about robot speed reduces an operator's mental strain.

Schirner et al. [77] discussed the future of human-in-the-loop cyber-physical systems. They gave possible applications and explained the framework they are working on. The purpose of the framework is to receive biometrics and estimate human intention from signals where this kind of system is helpful for locked-in individuals. Hu et al. [55] experimented on an estimated human trust index model using EEG and GSR sensors in real-time. In their experiment, they asked users to evaluate a virtual sensor reading in simulation. Based on sensor accuracy and the subject's response, the result showed that using physiological signals to estimate human trust level was promising.

The authors of [59] introduced a multi-modal emotional state detector using multiple devices. Their experiment focused on short-term GSR and heart rate, short-term GSR and EEG, and long-term GSR and heart rate characterization. Rani et al. [56, 78] tried to analyze anxiety-based affective implicit communication between the human and the robot. The researcher used ECG, EDA, EMG, and temperature signals in the regression tree and a fuzzy inference engine. Their results showed that the detection of anxiety using physiological signals was promising and might show better results in the future.

As computer games became more popular, Rani et al. [36] were trying to keep computer games more engaging by utilizing physiological signals. The authors estimated a gamer's effective state in real-time and altered game difficulty. The results showed that performance improves and creates lower anxiety during gameplay.

Erebak et al. [57] conducted an experiment among caregivers. Their result showed that human-like robots and typical robots did not have a difference in terms of appearance. There was a moderate correlation between trust in automation and intention to work with robots. There was a weak positive correlation between trust in automation and preference for automation.

Dobbins et al. [38] used a wristband that recorded GSR signal during the day. Their study was to estimate negative emotions, such as stress. The GSR signal from 6 subjects was collected during the day, with two daily surveys to be used for the labeling.

Ferrez et al. used EEG signals to detect error-related potentials (ErrP) when the subject makes a mistake. Their results showed that estimation of ErrP was promising and could be used in HRI [61]. Ehrlich et al. investigated the usage of EEG signals to detect the

intention to initiate eye contact when a robot needs to engage with humans [62]. Val-Calvo et al. proposed a framework that uses multiple signals of EEG, HR, and GSR to estimate Emotions while the subject is watching TV that can be used in HRI applications [63]. Mower et al. used physiological signals to estimate engagement implicitly. KNN with GSR and Skin temperature signals were used to estimate user engagement with an accuracy of 84.73%. The author suggested that implicit cues could be used in HRI applications [64].

Novak et al. used ECG, GSR, RPS, Skin Temperature, EEG, and Eye tracking signals to estimate a human's workload and effort. During trials, subjects were asked to fill a NASA-TLX questionnaire [65]. Iturrate et al. [66] used physiological signals to detect ErrP from EEG signals; they used both simulation and actual robots in their experiments. Their result showed that the brain-computer interface could be used as a continuous adaptation when there was no explicit information about the goal. Ehrlich et al. tried to validate robot action implicitly by using EEG signals. In their proposed approach, they focused on ErrPs, and their result showed a classification accuracy of 69.0% for the detection of an incorrect robot's actions [67]. Similarly, Salazar-Gome et al. [68] used EEG signals with error-related potential to fix the robot's mistake during the task. In the experiment, they used a Baxter robot to make decisions based on EEG signals in real-time. Dehais et al. [69] focused on hand-over tasks where a robot hand-over an item to a human. They collected the physiological signals during the task. Their results showed that the physiological response varies between the robot's motions.

2.5   Dynamic Speed and Separation

Speed and Separation Monitoring (SSM) is one of the safety methods for a human operator in industrial standards [79]. The SSM uses the minimum separation distance and the relative velocity between a human and a robot to control the robot to toggle between multiple modes such as slow and stop to avoid a collision. The conventional industrial application uses SSM with a LIDAR which is fixed at the base of the robot and measures the distance between a human and the robot. Then it uses this distance information to change the mode, as shown in Fig. 2.4. Although this approach increases safety, it is not productive when humans enter and exit many times during the task. As a result, the robot will stop many times, which reduces productivity. In order to address this issue, an alternative approach called

Dynamic Speed Separation Monitoring is implemented by Kumar et al. [1, 80, 81]. Unlike the fixed threshold by static LIDAR approach, multiple time of flight (ToF) sensor rings are placed on the robot to dynamically measure the distance based on robot pose and motion. Kumar et al. [1] presented three flavors of the SSM algorithm: TriSSM-Vo, TriSSM-Vr, and TriSM. These flavors toggle between three modes: normal, reduced, and stop. Hence, they are called Tri-Modal.
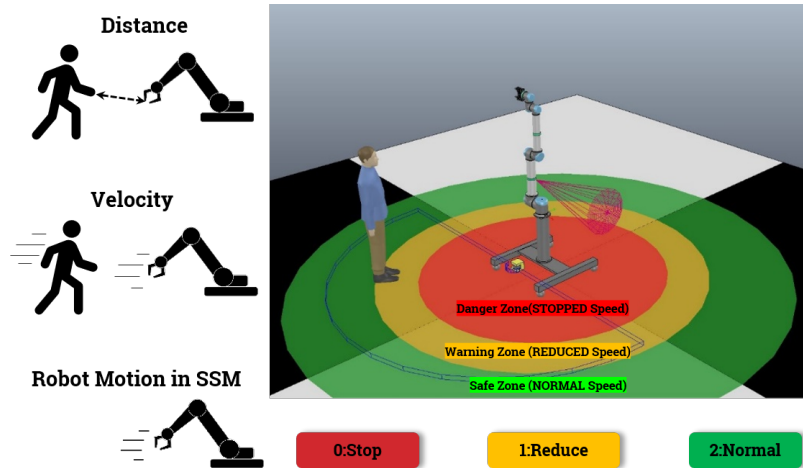
Figure 2.4: A basic SSM setup in a simulation representing static 2D safety zones around the robot. There are three safety zones, *Danger*, *Warning* and *Safe*, where the robot motion *stops*, *reduces* speed (slows down), or moves *normally*, respectively, to ensure the safety of the human [1]

- TriSSM-Vo is an SSM safety approach that takes into consideration the human/agent and robot-directed speeds for calculating the Protective Safety Distance (PSD) and Reduce Safety Distance (RSD) thresholds according to the SSM formulations in [82]. In the TriSSM-Vo, the human-agent/object is considered non-stationary. The directed speed is calculated as the projection of the relative velocity of the robot link and the human-agent/object onto the minimum distance vector [1].

- TriSSM-Vr is another dynamic SSM safety approach that takes into consideration the robot-directed speed and distance for calculating the PSD and RSD thresholds according to the SSM formulations. TriSSM-Vr considers the human agent/object is stationary, and the directed speed is calculated in the same way as TriSSM-Vo [1].

- Tri-Modal Separation Monitoring (TriSM) is an approach that does not consider the directed speeds. It uses constant thresholds for PSD and RSD. This is a form of dynamic SSM approach when used with on-robot ToF sensor arrays and a static SSM while using a 2D LiDAR [1], as shown in Fig. 2.4.

2.6 Summary

This chapter presents a compilation of state-of-the-art methods for physiological computing, commonly used physiological signals, data collection techniques, data labeling techniques, and various questionnaire methods. There might be other methods that are not included in this thesis for their scope and unpopularity.

In HRC application, safety and fluency are critical in the industrial standard. Although current standards address various aspects of safety, these are not sufficient. New personal and adaptive methods need to be implemented. Physiological computing is a way to implement personalized, safer, and fluent HRC applications. Therefore, in the near future, multiple variations of physiological computing will improve human-robot applications.

As wearable technology advances, new sensors are available on the market, and their cost will decrease. Hence, researchers and companies are taking advantage of this opportunity to develop better multi-modal sensor systems to improve human-robot collaboration/interaction. Although data collection is challenging, it is an essential input for physiological computing. The quality of the data is a key factor for physiological computing systems. Therefore, widely used data collection techniques need to be preferred and new innovative methods implemented. One drawback of the HRC application is the lack of open-source datasets. Future researchers need to implement their experiments and collect data. Public datasets will allow researchers to compare and improve existing physiological computing applications while maximizing time and effort at the same time.

Similar to data collection methods, data labeling lacks standard procedure. Presently, there is an absence of any standard protocol for data labeling. Researchers label their data based on the experiment, making it difficult to reproduce and compare the results.

In conclusion, this chapter discussed the commonly used physiological signals, data collection methods, data labeling methods, and questionnaires. We have also categorized

physiological computing research in terms of stimuli and data collection type used during experiments, data labeling methods, and machine learning algorithms. We hope this research will provide a framework for future researchers to select a suitable approach for future physiological computing systems.

Chapter 3

Human Comfort Index Estimation System

In this chapter, the core contribution of this dissertation is presented. The exploration of human comfort index estimation is discussed and the formulation and implementation of a comfortability index estimation system are presented. Then physiological signals, subjective data collection, data labeling, and machine learning algorithms are provided.

3.1 Exploration of Human Comfort Index Estimation in HRC

Physiological computing is an interdisciplinary field that affects many research areas. In this section, we are going to cover how physiological computing was used and how it can be used in the field of human-robot collaboration. One method of physiological computing is to have a dynamic representation of the user-state that is updated through physiological signals. Then the representation of the user-state can be used to inform a system for adaptation [5]. For example, a collaborative robot can reduce its velocity if a person is not comfortable or recover its velocity if the person is starting to feel comfortable. This type of physiological computing is known as 'biocybernetic adaptation'.

As the physiological signals are the fundamentals of physiological computing, usually an experiment is required to monitor the human responses. In section 2.2, the importance of the data collection was explained. For simplicity, in this chapter, we will assume that the physiological signals and subjective responses are collected during the HRC task that triggers a variety of levels of comfortability. By using this data, we will present four biocybernetic adaptation approaches that can be used in HRC.

3.1.1 Machine Learning Approach

Machine learning and Deep learning have become popular in the last decade in various applications, including physiological computing systems. A common approach is to map physiological signals to the desired output, such as emotion. Here, we will provide an overview of how it can be used in HRC applications.

Figure 3.1: Machine Learning Steps for Physiological Computing Applicationa

Fig. 3.1 shows how to use physiological signals in a machine learning application. This approach is widely used in the field of physiological computing systems. The standard ML steps are applied to signals: preprocessing, feature extraction, and based on the label type, a classification or regression algorithm can be used to estimate the target. The signal preprocessing and feature extraction methods are discussed in section 2.1.7.

Once the ML model is trained and its hyper-parameters are tuned, it can be used for HRC applications. For example, an ML model can use extracted features from the signals to estimate comfortability. Then the robot can use the comfortability signal (control signal) to change its behavior, such as its velocity, trajectory, or acceleration. Researchers widely use this approach. For example, Artal-Sevil et al. [40] used this approach to implement a low-cost robotic arm controlled by an EMG signal. Similarly, Savur et al. [11] used surface EMG signal to estimate American Sign Language alphabets.

Unlike classical ML models, deep learning approaches can learn from the raw data. Thus, Fig. 3.1 can be simplified by removing the feature extraction block. It is important to mention that deep learning approaches are data-hungry, and it is difficult to collect data for HRC experiments. The HRC experiments require the programming of a robot, testing, getting permission for human subject experiments from the institution, and calling for people for data collection. This whole process takes a long time.

3.1.2 Anomaly-Based Approach

The novelty-based approach is similar to the ML approach; however, the model learns what the normal is and detects anomaly occurrences during a process. A good example of this type of system would be one-class classification. Bowen et al. [83] presented an online novelty detection system that uses a mixture of Gaussian with a one-class classification method to estimate the stress level of an equine using equine's motion data. Then the system triggers an alert when a novelty is detected.

The machine learning algorithm suffers during training when the dataset is imbalanced since it creates a bias towards a dominant class in the dataset. However, the imbalance in the dataset between classes is not affected by the novelty detection methods. Therefore, novelty detection algorithms are preferable when there is limited data for minority classes.

3.1.3 Rule-Based Approach

The Fuzzy Logic method was developed by Lotfi Zadeh in 1960. The fuzzy logic allows us to define human-readable rules for reasoning between input and output signals via member functions. It has been used in many fields, from control systems to machine learning algorithms [84]. The fuzzy logic does not need a model; hence it does not need to be trained. However, rules for the fuzzy logic system need to be defined.
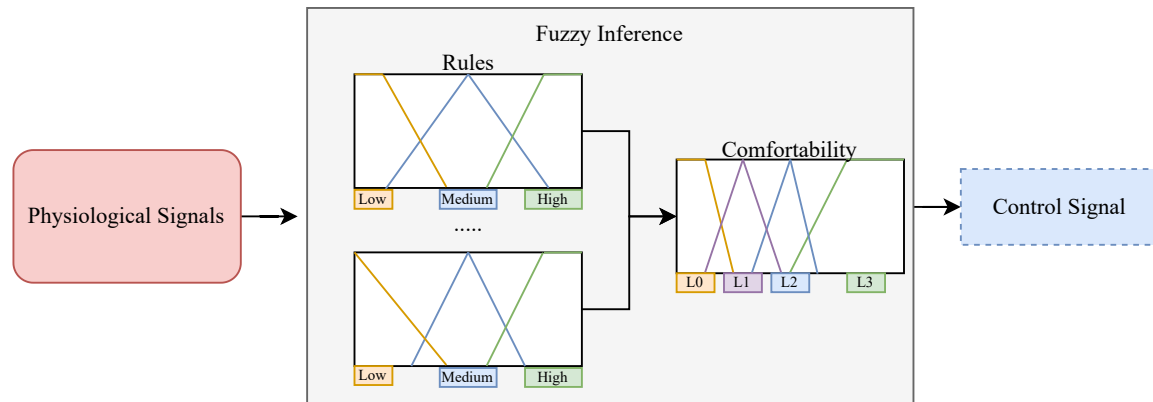
Figure 3.2: A simple Fuzzy logic system that can be used in physiological computing

Fig. 3.2 shows a simple diagram that can estimate comfortability using physiological

signals. In Fig. 3.2, three-member functions are defined for signals and four-level membership functions are defined for comfortability (output). Even though there is no training for such a system, it is challenging to define appropriate rules for the system and requires expert knowledge in the field.

Fuzzy logic algorithms are used in many applications, such as stress level estimation during commute [38], online stress detection during human-robot cooperation task [84], human safety estimation during HRC task [8] and emotion estimation [85]. In addition, researchers have used fuzzy logic for estimating the arousal and valence from physiological signals and then can estimate emotion based on Russel's circumplex model [2].

In general, fuzzy systems are easy to implement, and the system's internal state can be monitored. However, the number of member functions and the type of function that will be used are based on trial and error. In addition, defining rules plays a critical role in fuzzy systems and requires domain-specific knowledge.

3.1.4 Arousal and Valence-Based Approach
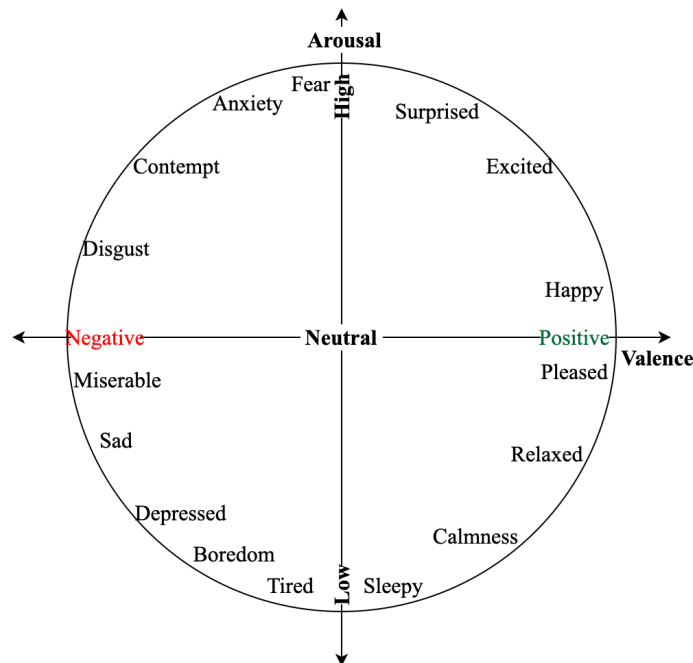
Figure 3.3: Russell's Circumplex model with basic emotions

A widely used circumplex model (aka emotion wheel) for basic emotions' representation was proposed by Russel et al. [2] in 1980. The circumplex model allows each emotion to be represented by *arousal* (y-axis) and *valence* (x-axis). The *arousal* value indicates how calming or exciting the emotion is, and the *valence* value indicates how negative or positive the feeling is. Thus, any emotion can be represented as a function of arousal and valence [2]. The circumplex model can be divided into four quadrants, as shown in Fig. 3.3. For example, the emotion 'happy' is in the first quadrant of the graph where *valence* is positive, and *arousal* is slightly greater than natural. On the other hand, 'sad' is in the 3rd quadrant where *arousal* is slightly greater than neutral, and *valence* is negative. The strength of emotion is measured by the distance from the origin [3, 86], where the origin represents no emotion.

Arousal and valence (AV) domain can be used in two ways: circumplex-based model and distribution-based model.

Circumplex Model-Based

The circumplex model-based approach uses Russel's circumplex approaches. In the Russell model, emotions are in the unit circle, and based on arousal and valence; the emotions are located in the 2D Cartesian systems. Here the origin represents a neutral, and as the emotion goes away from the origin, the strength of the emotion increases [3].

Fig. 3.4 shows the circumplex model, where the y-axis (low to high) represents arousal, and the x-axis (negative to positive) represents valence. To find the region of comfortability on the circumplex model, we will need three things: arousal, valence, and perceived comfortability. If we have sufficient data and we know what the perceived comfortability value is for that AV point, we can fit a line with a margin to the data that has maximized the comfortability, as shown in Fig. 3.4. Once the location of the comfortability axis is detected, we can easily estimate the comfortability level by using only arousal and valence.

The comfortability is estimated based on how much the AV point is away from the origin and by how close it is to the comfortability line (dashed region). The comfortability gets smaller as the AV point gets close to the origin and gets larger when getting closer to the edge of the circumplex model while staying in the cone.

AffectNet is a large dataset of images that are labeled by arousal, valence, and discrete emotions [86]. Similarly, DEAP is another dataset containing physiological signal (EEG)
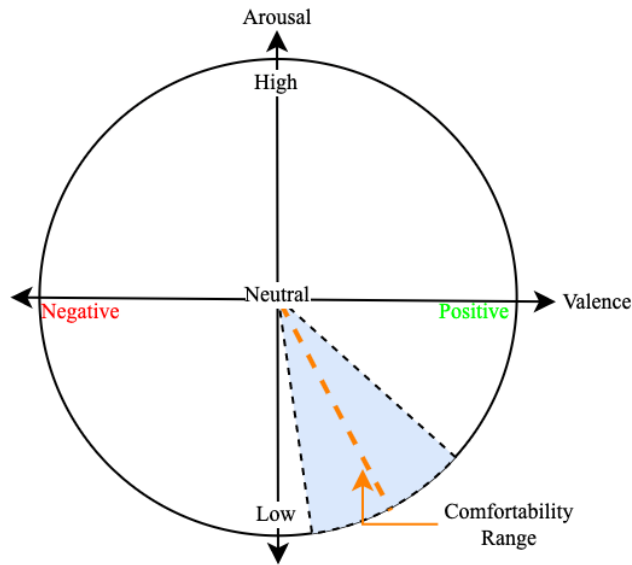
Figure 3.4: The circumplex model estimation of comfortability

data that contains arousal and valence for basic emotions. There were multiple approaches that attempted to use Fuzzy logic to estimate arousal and valence from physiological signals. Recently, Toisoul et al. [3] estimated continuous arousal and valence from the human face and used this information to detect what emotion the person was experiencing. Du et al. [87] examined the effects of emotion on takeover performance using the circumplex model approach with the SAM questionnaire.

Density-Based

The circumplex model assumes that comfortability is an axis within the circumplex model. However, the KDE-based approach does not make this assumption. The KDE-based approach uses arousal and valence space and then fits a 2D distribution based on comfortability. Thus, the shape of the distribution looks like an ellipse, as shown in Fig. 3.5. The KDE-based approach estimates comfortability based on how far the AV point is from the center of the distribution. The comfortability increases as the AV point get closer to the center and decreases when going away from the center of the distribution.

Unlike the fitting of distribution to the AV domain, Galvao et al. [88] divided the AV
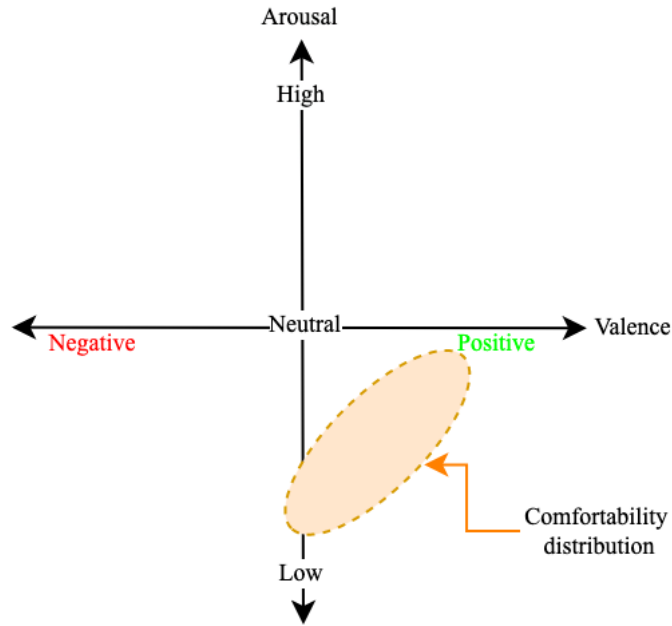
Figure 3.5: Density-based model

domain into four regions and tried to estimate arousal and valence from EEG signals.

This approach is simple and intuitive; however, it needs to know arousal and valence beforehand.

3.2 Proposed Methodology

This thesis used the arousal and valence-based approach to estimate comfortability and uncomfortability. In the proposed research, participants were asked to report their four emotions and comfortability level during trials. The basic emotions are marked in the circumplex model shown in Fig. 3.6 and the emotions used in this research are marked with color (*surprise*, *anxiety*, *boredom* and *calmness*).

Redondo et al. [89] discussed the comfortability in HRI settings. Their experiment results showed that comfortability differs from an emotional or affective state and can be triggered by a combination of multiple emotions/affective states. Therefore, we used *surprise*, *anxiety*, *boredom* and *calmness* emotions that locate in each quadrant in circumplex model to estimate comfortability and uncomfortability.
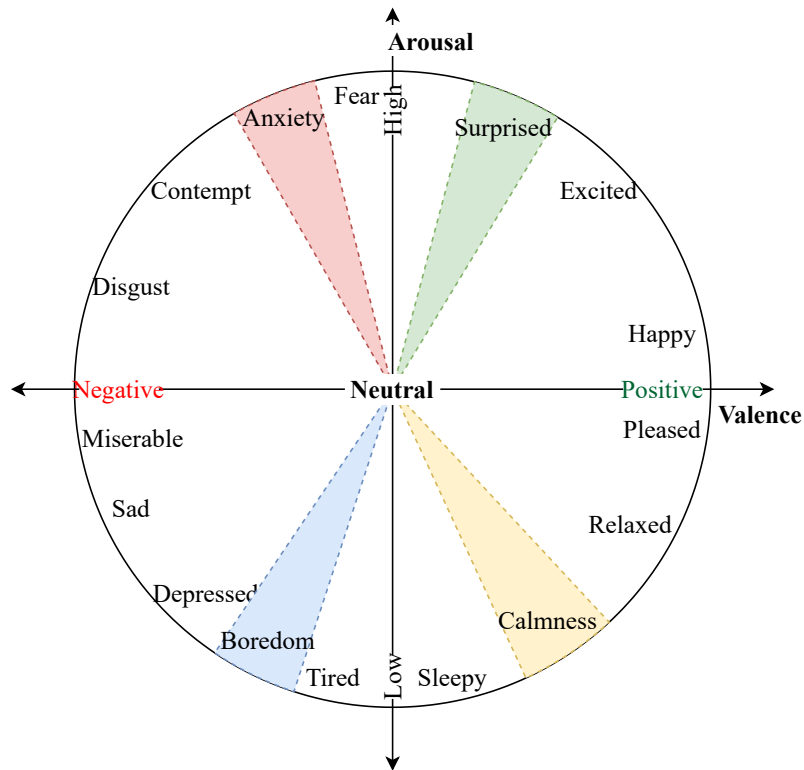
Figure 3.6: Arousal and Valence circumplex model and a few discrete emotional classes and their locations [2, 3].

We proposed two approaches, the first one is a circumplex model, and the second one is a Kernel Density Estimation (KDE) model. Although both approaches use the same formula for the AV Transformation, the estimation in CI/UnCI is different. The first approach estimates the comfortability and uncomfortability axis location (angles) on the circumplex model from perceived emotions reported by the participants. Unlike the first approach, the second approach does not make any assumptions about the circumplex model. It uses AV data points to estimate the distribution of the data points with reported comfortability or uncomfortability values.

In addition to the proposed approaches, we estimate emotion from physiological signals to be used in AV Transformation. Then, we evaluated both approaches with the estimated emotion.

Next, we will explain both approaches and how to estimate emotions from physiological signals.
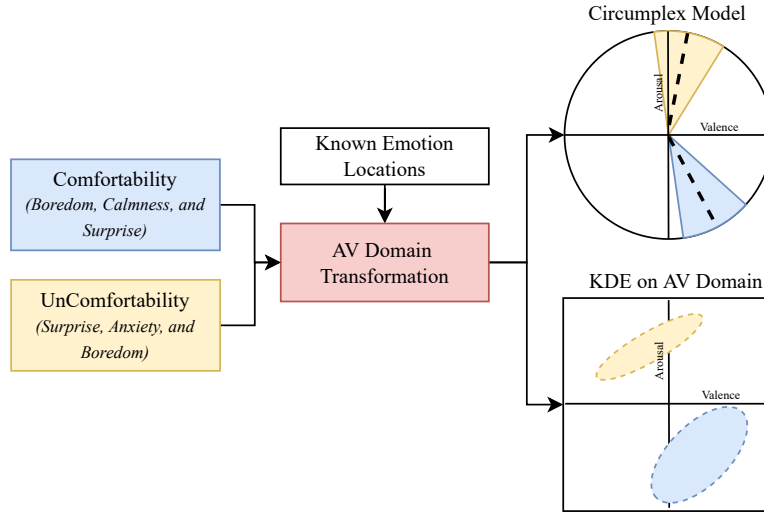
Figure 3.7: The proposed method uses reported emotions and pre-defined emotion locations to estimate arousal and valence. The first approach uses the circumplex model and estimates comfortability and uncomfortability axes. The second approach is only using arousal and valence to estimate the density of the reported comfortability and uncomfortability.

3.3 Approach 1: Circumplex Model-Based

In this research, we estimate comfortability and uncomfortability separately since the emotions that trigger both cases are different. For example, if a person feels anxiety, s/he cannot be comfortable simultaneously. Similarly, a person cannot be uncomfortable and feel calm at the same time. However, a participant can feel *surprised*, *calm*, and *bored* while being comfortable and feel *surprised*, *anxiety*, and *bored* while being uncomfortable.

Here, the input (the perceived emotion reported from the participants) p_e is obtained for both comfortability and uncomfortability as shown in Fig. 3.7. The comfortability estimation uses *calmness*, *surprise*, and *boredom* emotions. The uncomfortability estimation uses *surprise*, *anxiety*, and *boredom* emotions. We did not include the *calmness* emotion in estimation uncomfortability, since a person cannot feel *uncomfortable* and *calm* at the same time. Similarly, we did not include *anxiety* in the *comfortability* estimation.

Next, we selected the location of the *surprise*, *anxiety*, *boredom*, and *calm* emotions from other research [2, 86, 3, 87] (*surprise* = 60°, *anxiety* = 110°, *boredom* = 240°, *calmness* = 290°) as shown in Fig. 3.6. Then, ±5 degree noise was added to each emotion

axis to create a range instead of a line, as shown in Fig. 3.6. An emotion location θ_e vector is obtained using pre-defined emotion locations. Afterward, the arousal and valence location AV_{loc} is computed as

$$d_{valence} = \frac{\boldsymbol{p_e} \cdot cos(\boldsymbol{\theta_e})}{\sum \boldsymbol{p_e}} \tag{3.1}$$

$$d_{arousal} = \frac{\boldsymbol{p_e} \cdot sin(\boldsymbol{\theta_e})}{\sum \boldsymbol{p_e}} \tag{3.2}$$

$$AV_{loc} = (d_{valence}, d_{arousal}) \tag{3.3}$$

where the $d_{valence}$ indicates the *valence* component of the AV_{loc}, and $d_{arousal}$ indicates the *arousal* component of the AV_{loc}.
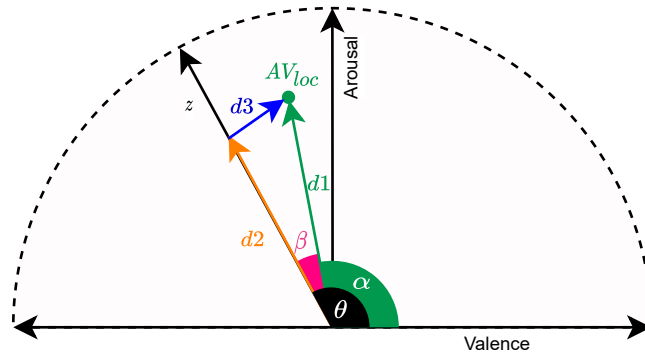
Figure 3.8: Given a z axis on the circumplex model, how CI or UnCI is calculated.

Once the perceived emotion is transferred in the AV domain, there are multiple ways to estimate CI/UnCI level. One way is to follow the circumplex model approach. In the circumplex model, the origin represents a natural/no feeling zone, and emotions get stronger as they get close to the edge of the circumplex circle [3, 86]. Fig. 3.8 shows an z axis and AV_{loc}. By following the circumplex approach and given an axis location and an AV_{loc}, we estimate the strength of the CI/UnCI by using $g(\theta, AV_{loc})$ as

$$g(\theta, AV_{loc}) = \begin{cases} d1 & if \quad |\beta| \le 5°, \\ d2 * (1 - \frac{d3}{2})) & otherwise \end{cases} \tag{3.4}$$

where β is an angle that is calculated with absolute difference between z axis' angle (θ) and AV_{loc}'s angle (α) as defined $|\theta - \alpha|$, $d1$ is the length of AV_{loc} from the origin, $d2$ is the length of projection of AV_{loc} on axis from the origin, and $d3$ is the distance between AV_{loc} and the axis.

To find the optimal z axis that represents the CI and UnCI axes, we optimize θ the angle that minimizes the mean squared error (MSE) shown in 3.5.

$$minimize = f_0(\theta)$$
$$= \frac{1}{n} * \sum_{i=1}^{n} (y_i - g(\theta, a_i)))^2 \tag{3.5}$$

where $a_i \in R^2$ represents one AV_{loc} that's estimated in 3.3, y_i is the corresponding comfort/uncomfort value that was reported by the participant.

In summary, we used perceived subjective emotion from the participant to estimate arousal and valence (AV domain) for both comfortability and uncomfortability. Then we used the AV locations to fit the best axis that fit for comfortability and uncomfortability separately by using 3.5. These two axes will be used to estimate the level of comfort and uncomfort in the following section.

3.4 Approach 2: Kernel Density Estimation (KDE)

The underlying assumption for the circumplex model was that an emotion is an axis where the emotion gets stronger as it moves to the edge of the circumplex circle. An alternative method can be to fit a distribution to the data points in the AV domain as shown in Fig. 3.7. Here, we used non-parametric kernel density estimation (KDE) to fit the data points in the AV domain by using the *scipy.stats.gaussian_kde* method with the *weight* parameter set to the reported comfortability values [90]:

$$kernel = gaussian\_kde(AV_{loc}s, weight = w) \tag{3.6}$$
$$g(AV_{loc}) = kernel(AV_{loc})/kernel_{max} \tag{3.7}$$

where *kernel* is a fitted Gaussian (distribution obtained by KDE), *w* is the weight vector of perceived comfort or uncomfort levels, and *kernel$_{max}$* is the maximum value of *kernel*. We divide the result by the maximum value in 3.7 to have a comfortability/uncomfortability value that ranges [0, 1]. Fig. 3.9 shows the KDE fitted to both the comfortability and uncomfortability levels.
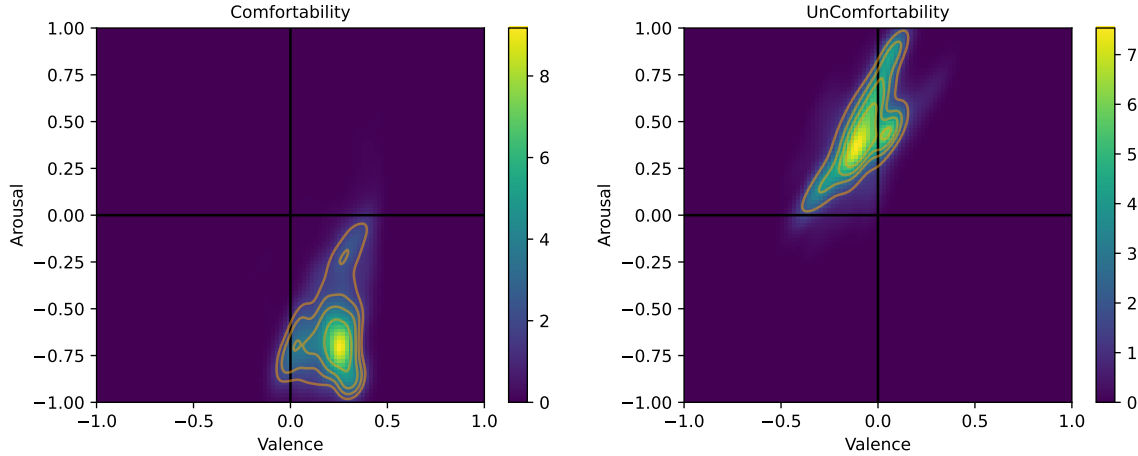
Figure 3.9: (a) Comfortability KDE plot, (b) UnComfortability KDE plot.

In summary, this approach does not make any assumptions and fits a non-parametric KDE distribution to the reported data. We then used the fitted kernel to estimate the likelihood of arousal and valence to estimate CI/UnCI.

3.5 Physiological Signals

In this research, we obtained ECG, GSR, and Pupillometry signals, and they were collected during multiple trials. Brief information about the signals that were used is as follows:

- **Electrocardiogram** was first preprocessed to remove high-frequency noise; a Butterworth low pass filter (cutoff=0.5, and order=5) was applied. Then power line noise was removed from the ECG signal. After that 'neurokit' algorithm was applied to extract RR intervals [91]. Then the time-domain metrics, shown in Table 2.1, were used to quantify the ECG signal. Commonly used time-domain metrics are the mean of heart-rate (mean HR), mean of the RR (Mean RR), the standard deviation of the

NN intervals (SDNN), the root-mean-square of the difference of consecutive NN intervals (RMSSD), and the proportion of the number of successive NN intervals that differ by more than 50 msec divided by the total number of NN intervals.

- **Galvanic Skin Response** is used in the proposed system. A low pass filter with *order* = 4 and *Wn* = 0.078 was applied to the GSR signal, and a 'boxzen' filter was used to smoothen the signal [92]. Subsequently, the signal is decomposed into two components known as phasic (SCR) and tonic response. The phasic signal was used to detect onsets, peaks, peak amplitude, and recovery time. Gamboa's algorithm [93] was used to detect the onset in the phasic signal. Common features extracted from the GSR signal are listed in Table 2.1. These features are the mean and standard deviation of the tonic and phasic response, rate of onset event from the phasic response, mean of the peak amplitude, mean of the rise time, and mean of the recovery time.

- **Pupillometry** signals were recorded from both eyes. A Butterworth low pass filter (cutoff=3, and order=2) was applied to the pupillometry signal to remove high-frequency noise. Then, time-domain features (see Table 2.1) were extracted.

3.5.1 Extracted Features

There are multiple ways of extracting features from signals: time domain, frequency, and non-linear features. In this research, the features listed in Table 3.1 were used.

Table 3.1: Physiological metrics extracted from the ECG, GSR, and Pupillometry signals

Type	Metric	Unit	Description
ECG	Mean HR	bpm/min	Mean of Heart rate
	Mean RR	ms	Mean of RR/IBI intervals
	SDNN	ms	Standard deviation of RR/IBI intervals
	RMSSD	ms	The root-mean-square of the difference of consecutive RR/IBI intervals
	pNN50	%	Percentage of successive RR/IBI intervals that differ by more than 50 ms
GSR	Tonic Mean	Micro-siemens	Mean of tonic component of GSR signal
	Tonic Std.	Micro-siemens	Standard deviation of tonic component of GSR
	Phasic Mean	Micro-siemens	Mean of phasic component of GSR signal
	Phasic Std.	Micro-siemens	Standard deviation of tonic component of GSR
	Onset Rate	onset/sec	SCR onset rate per second
	Peak Amp. Mean	Micro-siemens	Mean of Peak amplitude (SCR)
	Rise Time Mean	ms	Mean of rise time (SCR)
	Recovery Time Mean	ms	Mean of recovery time (SCR)
Pupil	Pupil Mean	pixel	Mean of pupil size
	Pupil Std.	pixel	Standard deviation of pupil size

3.6 Framework for Monitoring Human Physiological Signals during HRC Task

The block diagram of the proposed framework is shown in Fig. 3.10. The proposed framework is a solution for concurrently and continuously monitoring the human and robot state during an HRC task. The framework from a systems perspective can be conceptually categorized further into three sub-modules: *Awareness*, *Intelligence* and *Compliance* [94]. The communication layer between these sub-modules is equally important as it is responsible for data transformation and synchronization.

The sub-module *Awareness* is the perception of a system that is generated from physical world sensors and digitally represented in the virtual world. The physical world is responsible for sensing the environment through sensor information such as PPG sensor, GSR sensor, camera, motion capture system, etc. On the other hand, a virtual world is a digital-twin representation of the physical world that mimics the environment of the HRC task as well as the movements and behavior of the robot and human agents [95]. The digital twin can be used to calculate metrics such as human-robot minimum distance, directed human-robot speeds, possible collisions, and changes in trajectory [96][97]. The virtual
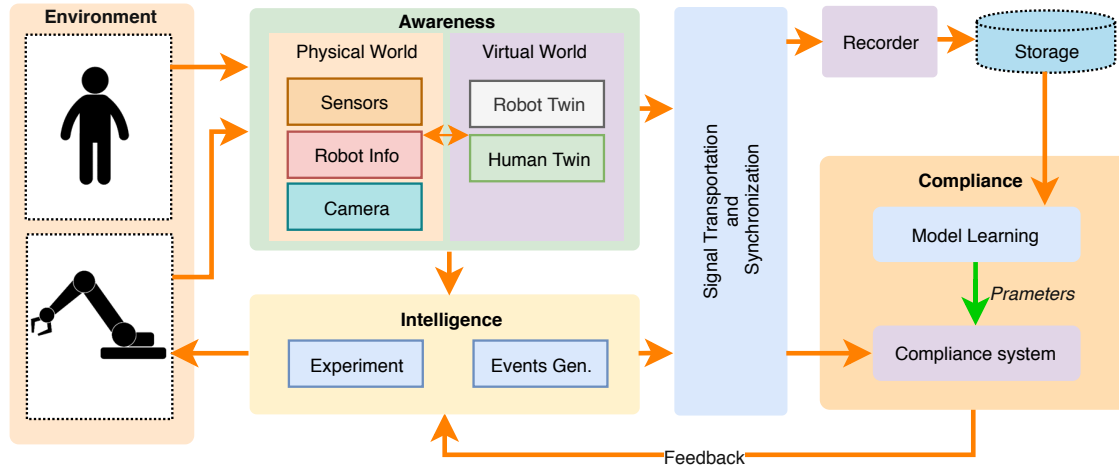
Figure 3.10: An overview block diagram of the proposed framework for monitoring Human Physiological Response during a Human-Robot Collaborative Task.

world updates its state constantly based on the sensory data received from the physical world to update itself and generate new data for the framework. Overall *Awareness* is responsible for sensing the physical and virtual world and providing this data to the rest of the system. Such a setup helps digitally represent a combined human-robot state, which can then be associated with the human physiological state.

The *Intelligence* represents the control of robot actions during an HRC experiment. Programming an experiment is part of the *Intelligence* since it controls the robot's speed, acceleration, and trajectory. The *Intelligence* module processes the data from the *Awareness* module to generate event markers as well as robot actions that can be used as stimuli to elicit a human response. In addition to *Awareness*, it also receives input from *Compliance* module, which is a form of interpretation of human expectation. The *Intelligence* module interprets this human command/feedback into actionable robot commands.

Using human physiological signals as feedback to the robot or a form of actionable control will help achieve a complete human-in-the-loop closed loop system. Here, the *Compliance* sub-module is responsible for inference from the physiological signals or any form of commands from the human that can be used to modify the robot's behavior. Thus, achieving a higher level of *Compliance* for the robot and managing the human expectation by interpreting the human physiological state can be a gateway to a more interactive human-robot collaboration.

Awareness, *Intelligence*, and *Compliance* are the main parts of the framework [94], however, to integrate these three modules a communication layer for data transformation and synchronization is required. This is critical as many sensor devices and other systems do not have the same frequency and timing clock. The communication layer is responsible for transferring data in real-time and synchronizing the data from different sources such as physiological signal collection devices, cameras, the representation of the human-robot state in the digital twin, and robot state information.

When designing human physiological signal-related experiments, the following aspects are critical.

- Experiment design

- Event markers generation

- Synchronization

The importance of these is elaborated in the following Sections.

3.6.1 Experiment Design

The experiment and its parameters need to be well-defined when designing an experiment. The task needs to be real or as realistic as possible to maintain the integrity of the robot's motion to act as stimuli to elicit the human physiological response. For example, an industrial task is a good option for the experiment. Hence, the industrial task may improve the involvement of the subject sharing human-robot collaboration workspace. In addition, the tasks need to be simple and controlled to increase the repeatability of a human-robot interaction scenario. A complex task may result in more uncertainty.

3.6.2 Event Marker Generation

The event marker generation is part of the experiment design. In the experiment, important events need to be investigated and generated by the experiment. Having markers during an experiment gives more intuition about the experiment, such as *"Experiment Start/End"*, *"Task Start/End"*, *"Robot Coming towards Human"*, etc. The event markers help to synchronize signals across different channels. For example, extracting Galvanic Skin Response

and Heart Rate signal between *"Experiment Start"* and *"Experiment End"* is trivial when the event markers are present during the signal recording. Thus, the markers can be used during post-processing for efficient data segmentation and epoching.

The auto-generation of event markers during an HRC experiment is critical. The choice of event markers depends on the experiment setup and the objective of the experiment. The biggest advantage of auto-generation of event markers is the experiment can be performed uninterrupted. These event markers can be used to effectively post-process and analyze the data as data segmentation, and epoching of the collected signals becomes easier. A list of events that are automatically generated during the HRC task for Case Studies I and II are listed in Table 3.2.

Table 3.2: The table shows potential event-markers that can be used during experiments

	Event Marker	**Definition**
Case Study	Experiment start	Experiment started
	Task [n] init	nth task initialized but subject has not complete loading yet
	Task [n] start	nth task started robot unloading all the parts
	Task [n] end	nth task unloading is done
	Robot approaching	Each time robot comes toward human will generate a event
	Pick up successful	Master pin is loaded
	Pick up failed	Master pin is not loaded
	Experiment end	Experiment is complete

3.6.3 Synchronization

Synchronization of signals from different sensors is crucial for the human physiological response. All the signals from the human and robot need to be synchronized with event markers. Thus, a central synchronization system is necessary. In the proposed framework for the physiological computing system, Lab Stream Layer (LSL) is used for interfacing subsystems, which integrates data from all different devices being used. The Lab Stream Layer is a system for collecting time series data over a local network with built-in time synchronization [98]. LSL stream is nearly real-time, and it is commonly used in biological signal collection systems such as OpenBCI, Pupil Lab, etc. Therefore, the LSL layer is selected as the central core of the data acquisition system in the proposed framework. In

the framework, each device has an application node that is responsible for acquiring signals from the device in real-time and pushing them to the LSL stream. A node is responsible for recording all-time series data from the LSL stream into a local file for post-processing and analysis. Along with LSL, Robot Operating System (ROS) and ZeroMQ are used to monitor data in real-time during the experiment [94].

3.7 Subjective Data Collection
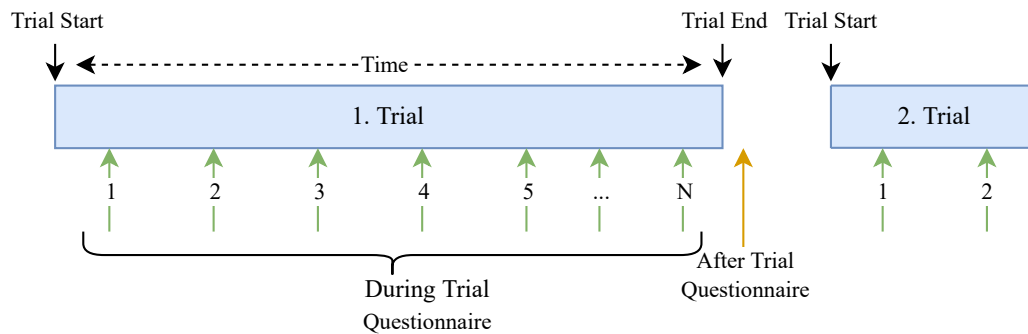
Figure 3.11: Two data collection methods: During and After trial [4].

Fig. 3.11 demonstrates the two data collection methods that were evaluated in this study. The first method is *after-trial* wherein participants reported their subjective responses only after the conclusion of the trial. This method is very common in the field. The second method is *during-trial* where the participants reported their subjective responses multiple times throughout the whole trial [4].

In order to minimize the time that the participants spent reporting the subjective responses, a custom Android app was developed, as shown in Fig. 3.12. The app allows the participant to enter their subjective responses immediately after each part (iteration) assembly on the tablet screen. This minimizes the duration of the data collection and maintains the integrity of the experiment. The during-trial approach produces more data and will provide a better idea of how the subjective responses change during the trials.
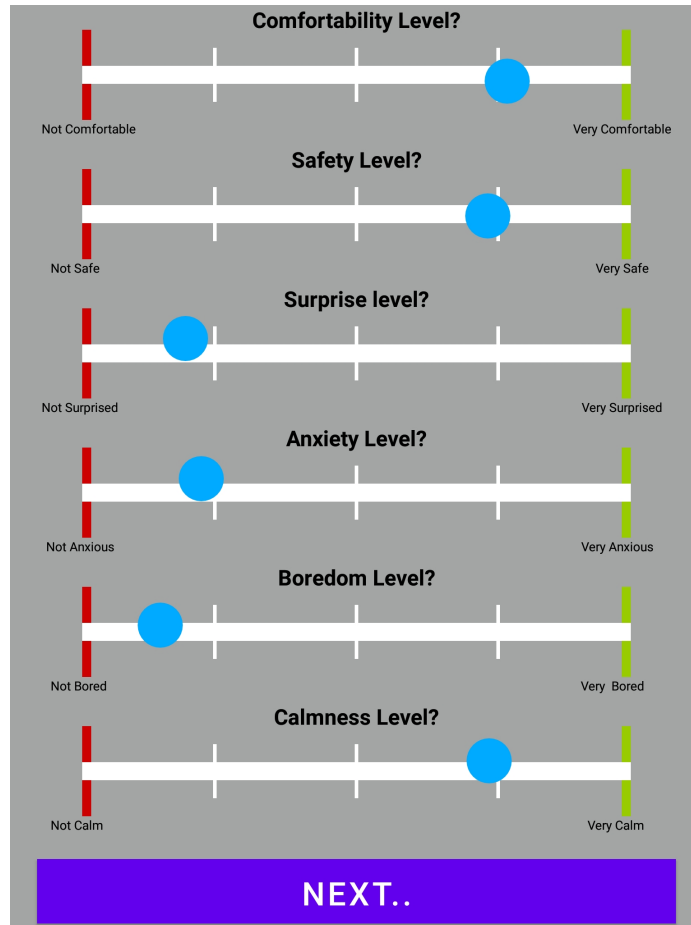
Figure 3.12: The custom android app that participant enters the subjective metric by touching the bars on the screen. The blue circle on the screen is randomly selected; it does not represent actual reporting.

3.8 Data Labelling

Supervised ML algorithms require labels for data during the training phase. In this research, we used *during trial* subjective responses as the label for the corresponding signals that were collected during the cycle, as shown in Fig. 3.11. Whenever the subject entered their response, the signal was labeled for 6 seconds both before and after the response was entered. Hence, 12 secs. of signals were labeled according to the subjective response. This approach was applied to each trial, which resulted in the dataset for the ML algorithms.

3.9 Emotion Estimation from Physiological Signals

Here, we estimate human emotion from physiological signals consisting of ECG, GSR, and Pupillometery signals. The physiological signals were collected during the HRC task and labeled based on subjective responses. As shown in Fig. 3.13, the signals were pre-processed and windowed, after which features were extracted. The extracted features are listed in Table 2.1.
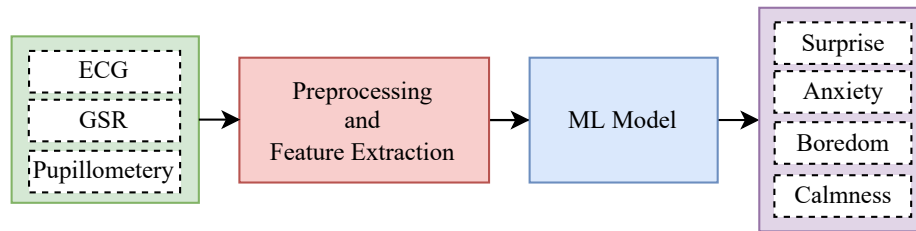
Figure 3.13: Emotion estimation from physiological signals.

3.9.1 Random Forest Regressor

The Random Forest (RF) regressor is a tree-based learner that uses multiple sub-samples of the dataset to train weak learners and uses majority voting among weak learners for prediction [99]. An RF regressor was used for the estimation of *surprise*, *anxiety*, *boredom*, and *calmness*. To find the best parameters, grid-search with cross-validation was applied to the training dataset. The best parameters were selected as: the number of estimators (*n_estimator = 100*), criterion (*mean squared error*), and the max depth (*max_depth = 6*). The rest of the configuration in scikit-learn was kept as default.

3.9.2 Functional Neural Network

A functional Neural Network (NN) regression model that estimates *surprise*, *anxiety*, *boredom*, and *calmness* from physiological signal was developed. All extracted features in table 2.1 were normalized before use in training. Each branch of the NN model consists of three layers, which have 16, 8, and 1 neuron; respectively. To prevent over-fitting, two dropout layers (0.3) were added between the hidden layers in Fig. 3.14. The number of layers was

selected by trial and error with the least complex model being selected. The *Adam* optimizer with *huber* loss was used for training. Five-fold cross-validation was used during training. The data were randomly divided into training (17 subjects' data) and testing (3 subjects' data). The test set was never used during training.
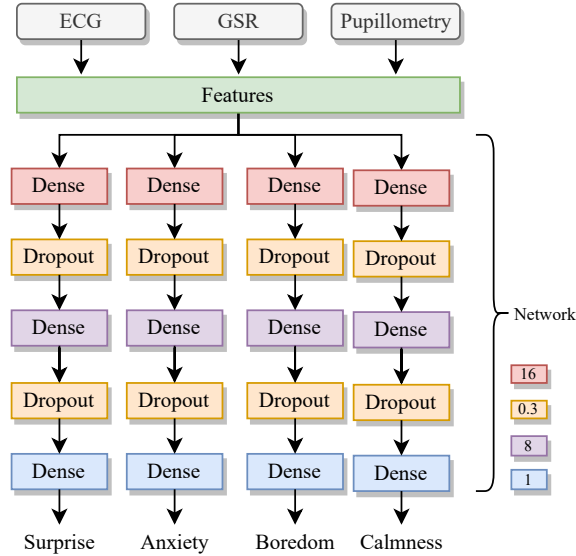
Figure 3.14: Functional Neural Network diagram that estimates emotions

In summary, we estimated emotions from physiological signals using ML models. Next we used equation 3.3 to transform the emotion domain into the AV domain. Finally, equations 3.4 and 3.7 were used to estimate the CI/UnCI level.

3.10   Real-time Estimation

To begin, Fig. 3.15 shows the real-time estimation of the CIES systems. Here you will see the raw signal streams are acquired by an independent thread for each signal. Each thread has a ring buffer with a size of 30 secs signal. The threads are responsible for pushing new data to the ring and removing the old data from the buffer by overwriting it.

ML pipeline retrieves data from all threads and applies preprocessing, feature extraction, and a transformation to the extracted feature. Then the input is provided to the pretrained ML models that estimate emotions which are required for calculating an AV point (see section 3.3). After that, the AV point is used to estimate CI/UnCI based on the formula
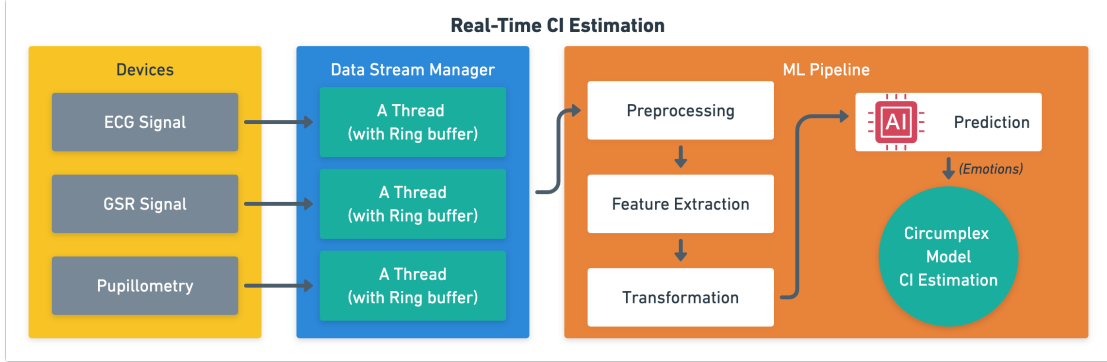
Figure 3.15: Real-time estimation of CIES system

in [3.4](). Since we are estimating CI/UnCI every second, it would be beneficial to apply an exponential filter to smooth the output. The exponential filter with $\alpha = 0.75$ was used for smoothing the output. The exact formula is shown in equation [3.8]():

$$r = \alpha * r_{prev} + (1 - \alpha) * r_{current} \tag{3.8}$$

where r is either CI or UnCI, α is a smoothing constant.

Fig. [3.16]() shows both estimated and filtered UnCI. The filtered UnCI was calculated by equation [3.8]().
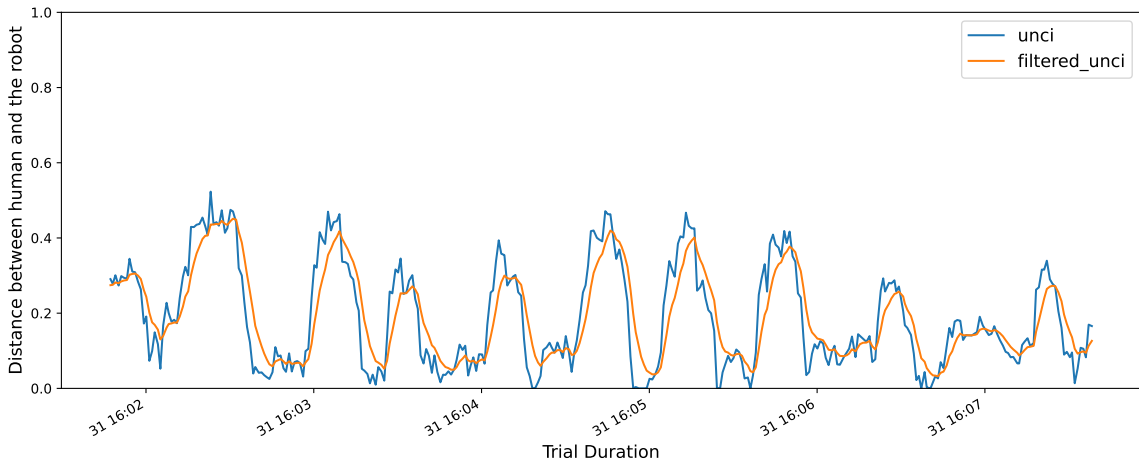
Figure 3.16: Estimated and filtered UnCI

Chapter 4

Stationary Experiment with Sawyer Robot:

Results and Discussion

In this chapter, we provide information about the experiment setup, the data collection protocol, and the statistics of people enrolled in the experiment. Then, the results of the experiment and its limitation are discussed.

4.1   Sequential Collaboration Task

This experiment is a sequential collaboration since the participant waits for the robot to provide a part for the assembly [100]. The data was collected from healthy college students (N=20). The participants consisted of 17 male and 3 female subjects (*Mean Age*= 24.70, *SD*= 2.99). The experiment setup consisted of a joint task between the robot and the human. The joint task is an assembly task, where the robot provides a part from Table-1 and the human picks up a part from Table-2 as shown in Fig. 4.1. The human is responsible for picking up and screwing the two parts together while the robot is holding one of the parts.

In the experiment, the robot changes its behavior by changing its *velocity*, *trajectory*, and *sensitivity* (three independent variables). The velocity was set to be in two levels: *normal* and *fast*. These two modes were achieved by setting the global speed ratio to 0.7 for normal mode and to 1.0 for fast mode. The trajectory was defined as two modes: *normal* and *extreme*. The normal trajectory was defined as the robot moving from Table-1 to Table-2 with minimum joint movements. The normal trajectory was smooth and predictable. On the other hand, the extreme behavior passes multiple waypoints between pick and place locations, which makes the robot movements jerkier and less predictable.

The sensitivity was defined in two modes as well: *normal* and *sensitive*. The sensitivity was measured by the amount of force applied to the robot's end-effector during the time the robot waits for the human to assemble the part. The sensitivity threshold was set to be $\pm 11 N/s^2$ and $\pm 8 N/s^2$ for *normal* and *sensitive*, respectively.
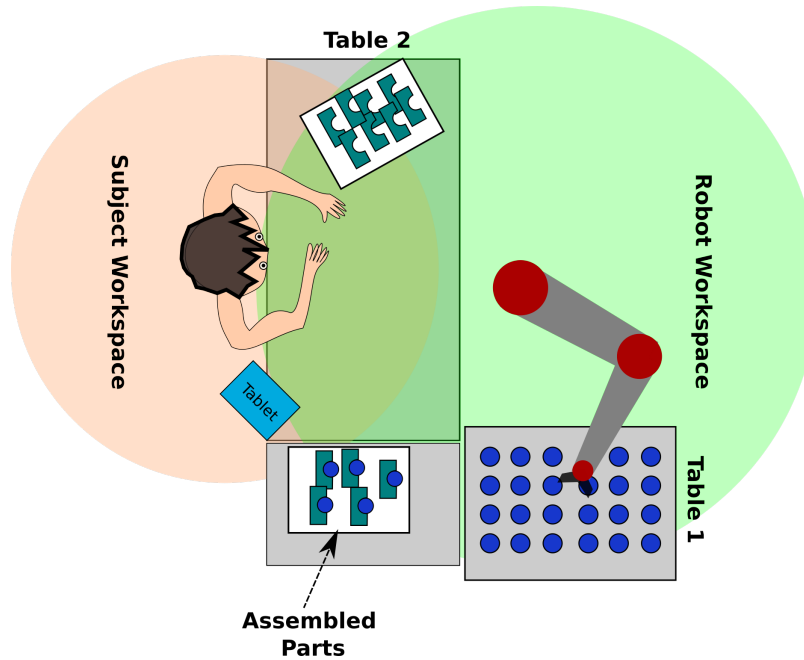
Figure 4.1: The experiment diagram is shown in a bird view.

This experiment consisted of 11 trials and one baseline. The robot can provide up to 24 parts (iterations) in a trial. The robot picks a part from *Table 1* and moves it in front of the participant for assembling. The participant needs to pick an item from Table 2 and assemble the part in five seconds, then the robot moves the assembled part to a drop-off location and moves to Table 1 for the next part. While the robot is dropping the part, the participant enters the subjective responses on the tablet where we asked the subject to report *surprise*, *anxiety*, *boredom*, *calmness*, and *comfortability* level in a continuous value range [0, 1]. The trial is completed when five minutes have elapsed, or there are no items left on Table 1 for the robot to pick up. The participant had an option to tap the robot's end effector to notify the robot not to wait anymore during assembling time. Hence, the participant can minimize the trial duration. Before the experiment started, all the participants were trained for one trial so that they get used to the task and the tablet.

In this experiment, to have better signal labeling, a custom Android app was developed. The app allows participants to enter their subjective responses immediately after each part assembly (iteration). Hence, this approach produced more subjective data and a better idea of how the dependent variables changed during the trials.

The experiment was approved by the Human Subject Research office at the Rochester

Institute of Technology. Informed consent from each participant was collected before the experiment.

Next, we will discuss devices that were used during data collection.

4.2 Data Collection Devices

The devices used for experiments were the Shimmer 3 GSR, Pupil Lab headset, and the Biopac BioHarness.

- *BioHarness* is a wireless chest strap that allows recording of an ECG signal. In addition to the ECG, the device provides respiration rate, heart-rate, RR intervals, acceleration (3-axes), and device information.

- *Shimmer3 GSR+* is a widely used device in research due to its Bluetooth connectivity. The device provides one GSR channel that measures the conductance of the skin and one PPG channel that measures the amount of reflected light (volumetric variations of blood circulation) from the vein [101]. The sensor sampling rate was set to 128 Hertz (Hz). During the experiment, we asked participants to minimize their motion when using the hand on which the sensor was placed. This was critical since the GSR signal is sensitive to motion artifacts and cannot be removed from the signal.

- *Pupil labs* headset is open-source hardware (eyeglasses) that has three cameras, two of which look at the eyes and one point to the subject's perspective [102]. The eye cameras operate at 120 frames per second (fps), and the world camera records at 30 fps. This device is widely used in research and provides various signals such as pupil diameters, gaze location, and a real-time camera stream. This research used the headset to collect pupil dilation and gaze location.

Each sensor was calibrated for each subject independently. The subjects started with a baseline recording where the subject sat in front of the robot and was asked to relax for five minutes. The subjects who had glasses were asked to remove them for better pupil signal quality. The pupil lab headset was connected to a Samsung S8 smartphone with a Pupil Mobile app that transmits data to a local machine. The Shimmer3 GSR+ connects via Bluetooth to a computer, and a custom application [94] developed to acquire these signals

were used to send data using the Lab Stream Layer (LSL) protocol. The BioHarness was connected over Bluetooth to a computer as well. All devices were synchronized using the LSL stream library [98]. A modified version of the custom data collection app that generates automated event markers (*trial start/stop*) and manual event markers during data collection was used [103].
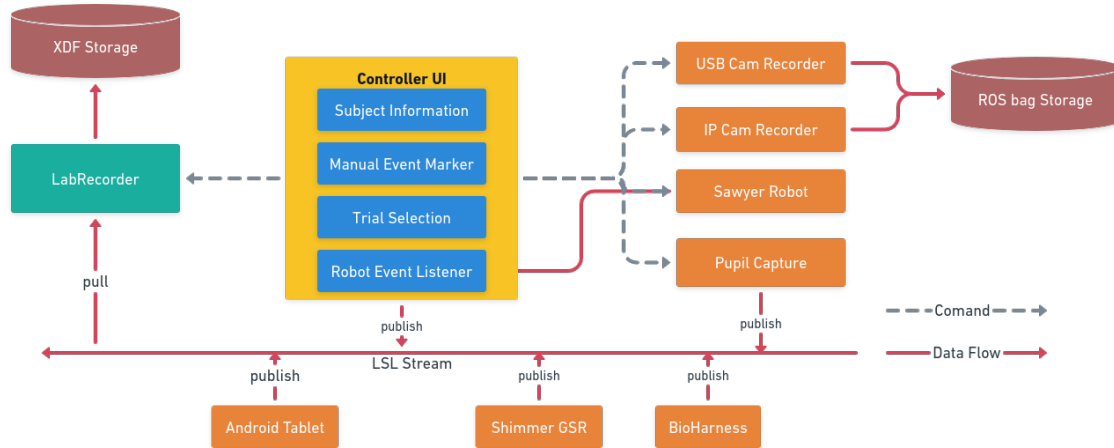
Figure 4.2: Overview of the Data collection system. The Controller UI is the main part of the system where the experiment is controlled. It helps to gather subject information, generate event markers, and launch all the other apps that record data, such as a USB Cam recorder, Pupil Capture, etc. In addition, it toggles start/stop in LabRecorder to record LSL streams. All the data in the experiment is recorded in XDF files except camera streams and Pupil Camera streams.

4.3 During and After Trial Subjective Measurement Collection

Before this experiment, we conducted two other experiments where we collected subjective data only after trials. However, we learned from those experiments that the data collection after-trial was insufficient; hence we developed a custom Android application that allows a participant to enter the subjective response during the trial.

In addition to the traditional approach, we collected subjective data during trials as well. The subjective responses were collected as continuous numbers between [0, 1], where 0 is

strongly disagreed and 1 is strongly agreed. The subject touched the bar shown on the tablet screen, which was converted to a continuous value. The questions were related to *surprise*, *anxiety*, *boredom*, *calmness*, and *comfortability* level after each iteration during the trial. In the traditional approach, the subjects filled out a questionnaire after each trial. The questions were in the format (I felt calm while working with the robot) and the answer was in a Likert scale consisting of *strongly disagree*, *disagree*, *neutral*, *agree*, and *strongly agree* as a discrete value. These metrics were related to the overall interaction with the robot. For example, if the robot's motion/action surprised the subject, we expected the participant to report agreeing or strongly agreeing with the corresponding question.

4.4   Data Collection System

Collecting signals from multiple devices and controlling the robot's behavior is challenging. Multiple problems, such as synchronization, bandwidth, storage, and control, can arise. We extend our previous approach for data collection [94] that can send commands to the robot to change its behavior on the fly. Fig. 4.2 shows the command and data flow between multiple systems. The controller UI program is the main program that controls the entire data collection system. The controller is responsible for acquiring the participant's information, trial selection, automated event generation, manual event generation, and sending commands to the rest of the system. Lab Stream Layer (LSL) is a widely used protocol in time series data collection that has built-in time synchronization and distributed ability in the local network [98]. The LSL was used for signal transferring into the local network. The custom Android app also sends the subjective response to the LSL stream. The recording program (LabRecorder) can record all selected streams that are used for signal recording. For the video recording, we took another approach. The video from two streams (USB cam and IP cam) are recorded as a rosbag. The video data was used only to inspect a trial if necessary.

4.5   Results and Discussion

This section discusses the results of our experiments. We evaluate the estimation of CI_{axis} from the subjective response, calculating *CI* using the circumplex model, and estimating emotions from physiological signals by using multiple ML models. Further, the performance of the proposed approach models is tabulated and discussed.

4.5.1   Comparison of Comfortability During and After Trial Response

Fig. 4.3 shows the reported comfortability for during and after trial. The x-axis shows the number of times the subject reported comfortability, and the y-axis shows the perceived comfortability level range 0 to 1. The after-trial response is a value that the participant reported at the end of the trial (dashed line). The during-trial responses were reported multiple times during the ongoing trial (dotted line). As it can be seen from the figure, the after-trial reporting cannot capture the variation in comfortability that happened during the trial. Collecting subjective responses enabled us to capture these changes and label the physiological signals more accurately.

The subject's response shown in Fig. 4.3 indicates that at the beginning of the trial, the subject was not comfortable as the time passed in the trial, the subject's perceived comfortability changed.

4.5.2   Correlation Between Comfortability and Emotions

In this section, we analyze the correlation between comfortability and emotions of *anxiety*, *surprise*, *boredom* and *calmness*. Spearman correlation was used, and its results were presented with scatter plots as shown in Fig. 4.4. The test result shows that comfortability has a high negative correlation with surprise (-0.64), a medium negative correlation with anxiety (-0.48), a low negative correlation with boredom (-0.29), and a high positive correlation with calmness (0.67). This supports the proposed idea from Redondo et al. where they indicated comfortability is not an emotion, but a combination of multiple emotions [89].
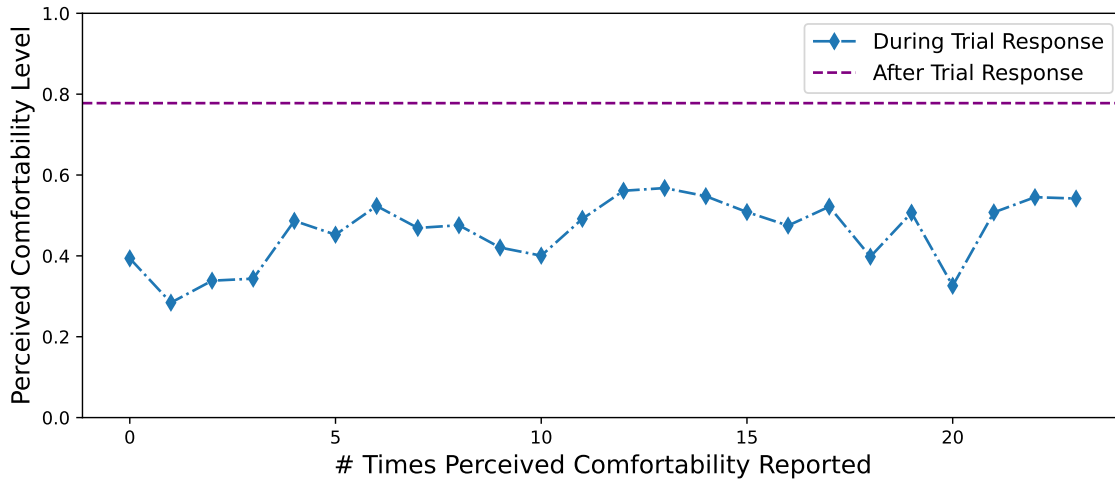
Figure 4.3: Comparison of after and during trial response and how comfortability changes during trial. The data shown here is taken from one trial from participant S13 from trial 7.

4.5.3 Analysis of Comfortability Axis

The subjective emotions (*surprise*, *boredom*, and *calmness*) are shown in Fig. 4.5a and were transformed into points (*arousal* and *valence* values) using equation 3.3. They were plotted in the AV domain as shown in Fig. 4.5b. Then, equation 3.5 was applied to the AV_{loc}s to estimate the best axis that represents the CI axis. As shown in Fig. 4.5b, the CI axis is close to *calmness*, which is expected since the subject reported high comfortability when they are calm.

4.5.4 Analysis of Uncomfortability Axis

Similar to the CI axis estimation, we used *surprise*, *anxiety*, and *boredom* (see Fig. 4.6a) emotions in 3.3 to transform the emotions into the AV domain as shown in Fig. 4.6b. Then, equation 3.5 was applied to the AV_{loc}s to estimate the best axis that represents the UnCI axis. As shown in Fig. 4.6b, the CI axis is close to *anxiety*, which is expected since the subject reported high uncomfortability when they reported high anxiety.
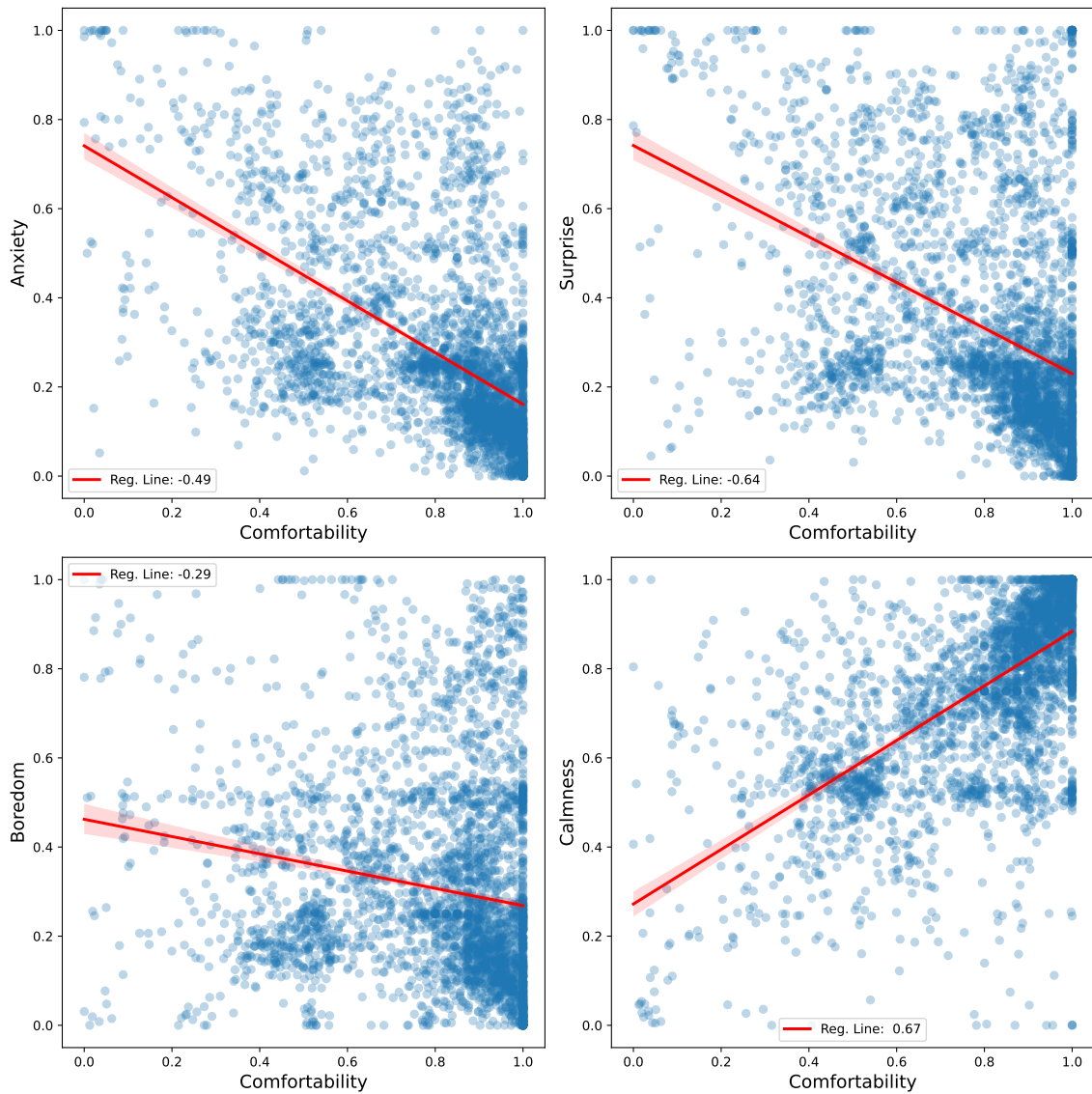
Figure 4.4: Correlation between reported comfortability and emotions

4.5.5 KDE Estimation

Two distributions obtained by KDE were fitted to the perceived comfortability and uncomfortability AV data points. Fig. 4.7 shows the fitted distribution in a heatmap. Unlike the circumplex model, the distribution for uncomfortability spreads into both quadrants I and II. The comfortability distribution stays in quadrant IV mostly.
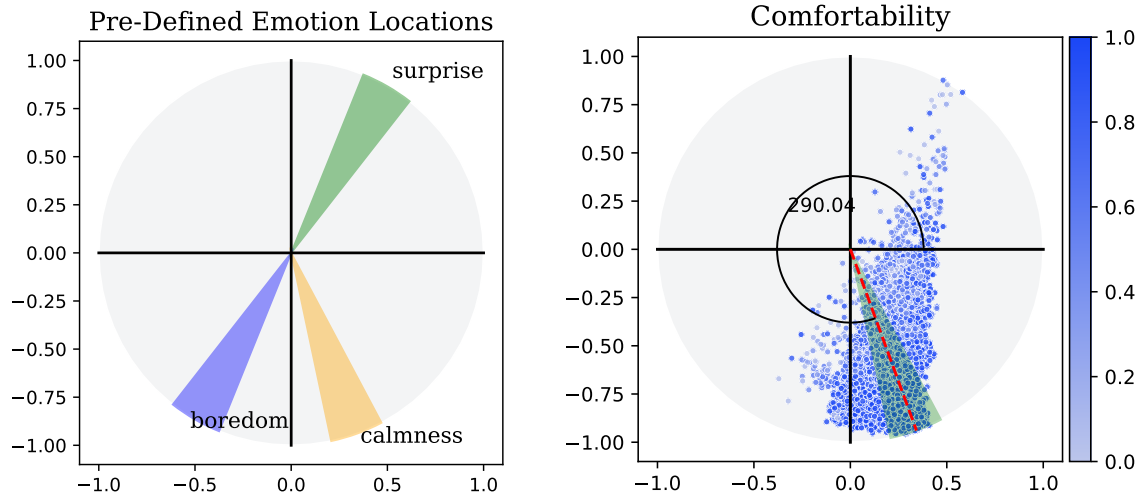
Figure 4.5: (a) shows the predefined emotion location for estimation of comfort axis, (b) indicates estimated CI axis location from subjective responses.
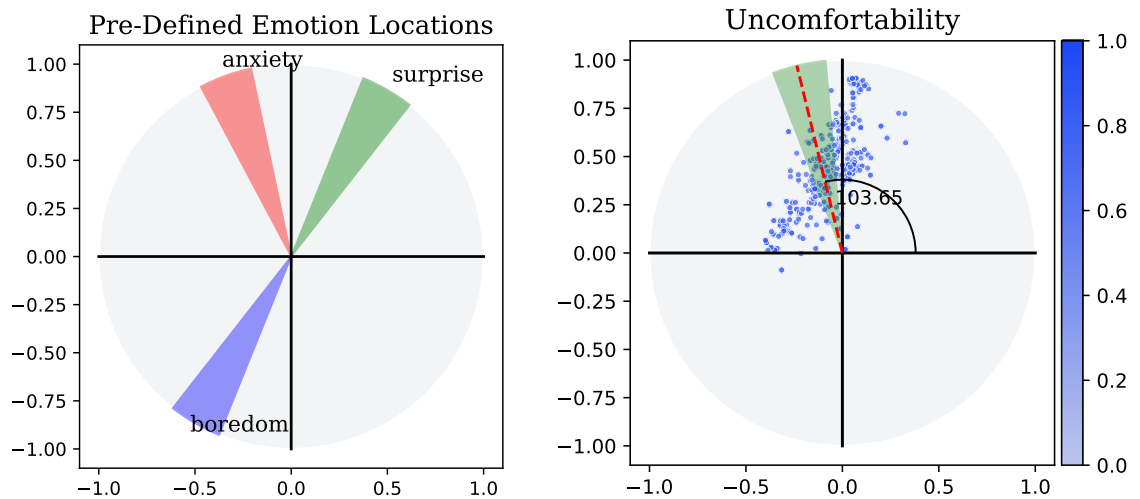
Figure 4.6: (a) shows the predefined emotion location for estimation of UnCI axis, (b) indicates estimated UnCI axis location from subjective responses.

4.5.6 Emotion and CI/UnCI Estimation from Physiological Signals

The dataset consisting of 2388 samples was used to develop the ML models. The dataset was divided into 70% and 30% for training and testing, respectively. Five-fold cross-validation was used during training. Table 4.1 shows the root-mean-square error (RMSE)
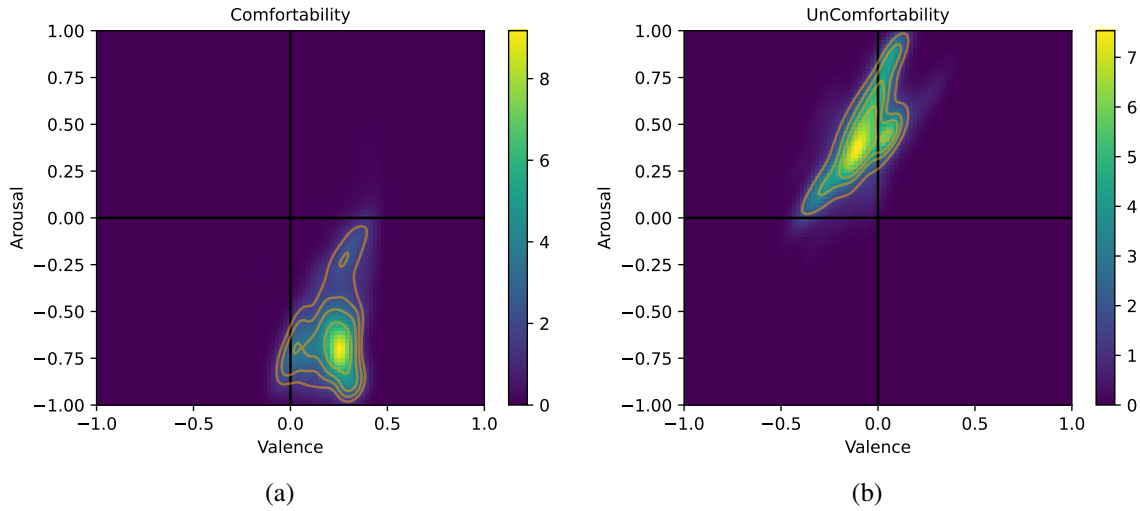
Figure 4.7: (a) shows the KDE fitted to comfortability AV data points (b) shows the KDE fitted to uncomfortability AV data points.

and mean absolute error (MAE) of emotions, comfortability, and uncomfortability for the Random Forest (RF) regressor, NN regressor, and Circumplex models.

Table 4.1: Performance of RF Regressor, NN, and Circumplex and KDE approaches.

Models	Surprise		Anxiety		Boredom		Calmness		Comfortability		Uncomfortability	
	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
RF	0.20	0.15	0.16	0.12	0.14	0.11	0.13	0.10	0.16	0.11	0.15	0.11
NN	0.22	0.15	0.18	0.12	0.18	0.11	0.15	0.10	0.19	0.11	0.19	0.11
Circumplex (RF)	-	-	-	-	-	-	-	-	0.35	0.30	0.28	0.24
Circumplex (NN)	-	-	-	-	-	-	-	-	0.34	0.30	0.26	0.24
KDE(RF)	-	-	-	-	-	-	-	-	0.47	0.40	0.51	0.43
KDE(NN)	-	-	-	-	-	-	-	-	0.50	0.46	0.60	0.54

Table 4.1 shows the RMSE and MAE of emotions, comfortability, and uncomfortability that estimates from the RF, NN, and Circumplex model approaches by using physiological signals. The circumplex model is only used to estimate comfortability and uncomfortability. The circumplex model approach uses estimated emotion from both RF and NN. From the table, we can see that the RMSE and MAE are higher for the circumplex model than for the RF and NN models. This is due to the estimation error that occurs when estimating emotion from two models because we used the estimated emotions to calculate arousal and valence using equation 3.2. Any error in emotion estimation can lead to a larger error in the

circumplex model approach. Although the error is high for the circumplex model approach, it is important to look at how it performs on trial data.

Next, we will analyze the performance of these four models on trial, which was randomly selected from the dataset that was never used in training.

4.5.7 Uncomfortability Estimation for a Single Trial

In this test, we used the leave one out approach. We kept a trial and did not use it in the training of the models. The x-axis shows the time elapsed in a trial, and the y-axis shows the UnCI level. During the trial, the robot's sensitivity was set to *normal*, the trajectory was *extreme*, and the velocity was *fast*. The ground truth (reported by subjects) is shown in blue. The blue line shows that the subject's uncomfort level changed during the trial. The data from a single trial fed into the ML models to estimate emotions, comfortability, and uncomfortability. Fig. 4.8 shows the uncomfortability estimation from the two models (RF and NN). From Fig. 4.8, the simple approach that estimates UnCI from physiological was unable to estimate UnCI when it is high in three locations. The RF regressor is slightly better than the NN model.

On the other hand, the circumplex and KDE approaches shown in Fig. 4.9 are better at estimating UnCI. Even though the circumplex model has lower values, it captures the overall shape of the prediction. The KDE approach made a high UnCI estimation where it was supposed to be low. This occurred due to limited data, where the emotion estimation is done only using approximately 20 seconds of data. The model improved estimation once there was enough signal estimation. Although these two approaches had higher RMSE and MAE values, their estimation is better when estimating the high uncomfortability values.

The proposed circumplex and KDE models use estimated emotion from the ML models to estimate *arousal* and *valence* and then estimate CI and UnCI from them. However, it would be better to estimate arousal and valence from the physiological signals. Since there is no standard way of calculating these metrics, and we did not ask participants their arousal and valence levels, we cannot estimate directly.

Although we estimated arousal and valence from emotion, this presents a huge opportunity as *arousal* and *valence* can be measured by physiological signals sensors and efficiently mapped to comfortability used in the circumplex model we created through these
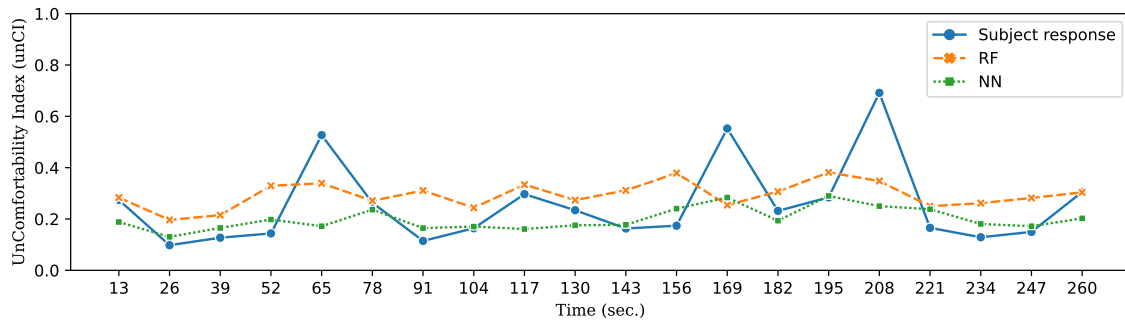
Figure 4.8: Performance of simple RF and NN that estimate Uncomfortability directly from physiological signals
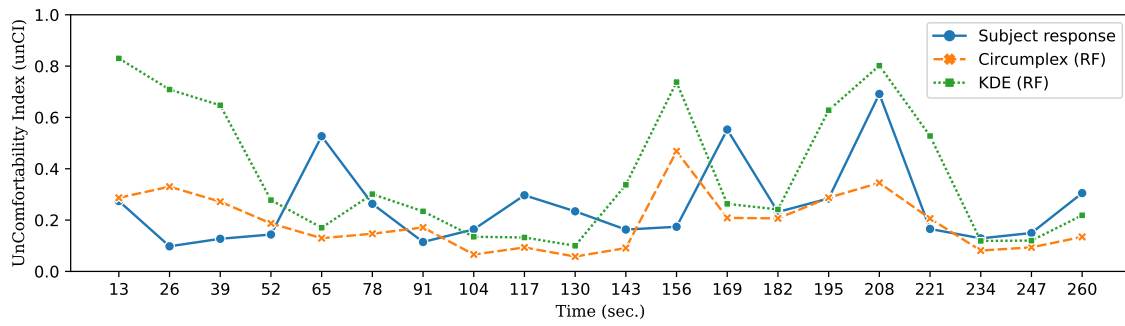
Figure 4.9: Performance of only Circumplex and KDE models that uses estimated emotions from RF regressors.

experiments.

4.5.8 Real-Time Estimation of UnCI

Controlling a robot during an experiment requires real-time estimation of UnCI. Hence, a real-time UnCI estimation system (CIES) that estimates UnCI from ECG, GSR, and pupil dilation signals was developed. The system uses a ring buffer that adds new incoming signals to the buffer and removes older ones as soon as the buffer is full. The estimation of the UnCI does not start until the buffer is full. Once the buffer is full, the CIES makes a copy of the buffer every second, cleans, preprocesses, and extracts features from each signal separately. The extracted features are concatenated, then fed to the RF regressor model for emotion estimation. To have a better estimation, the last ten predictions were averaged. Thus, this allows the CIES to have smoother CI estimation. Fig. 4.10 shows
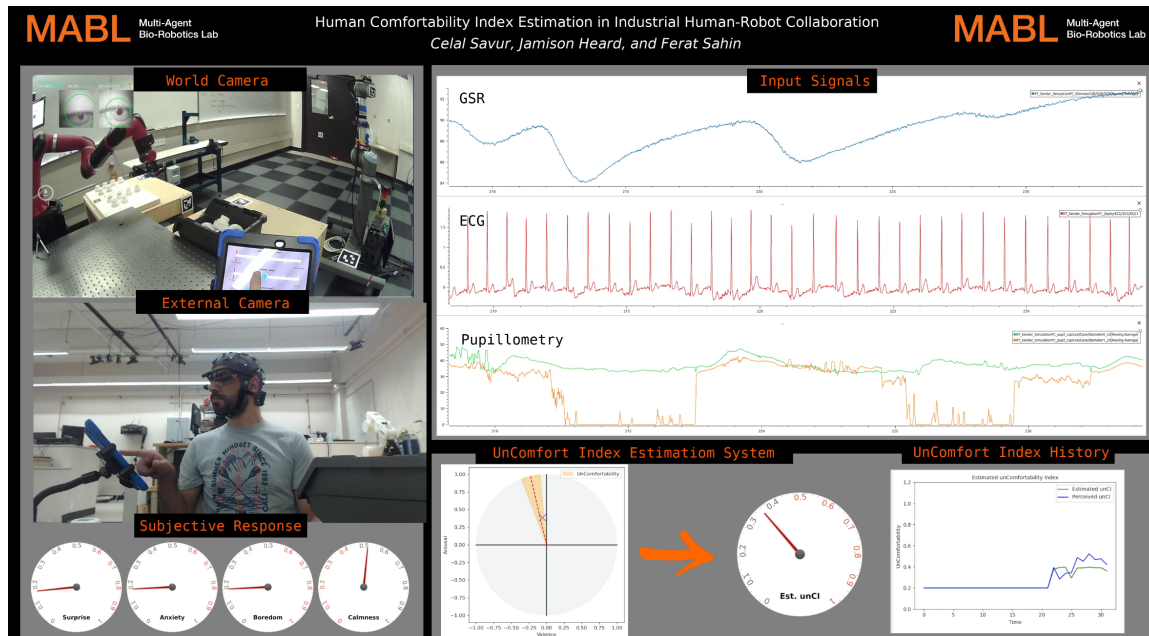
Figure 4.10: The world camera in the top-left corner shows the participant's view. The external camera shows the participant. The gauges in the bottom corner show the subjective responses reported during the trial. Input signals are the signal used in the CIES system for real-time UnCI estimation from GSR, ECG, and Pupil dilation signals by using the circumplex model with the RF regressor. The bottom-right corner shows the history of the estimated UnCI with the perceived UnCI reported by the participant. The video for the trial can be accessed from "https://youtu.be/qiTmN1ICVJo".

the data streams of GSR, ECG, Pupil dilation, and the real-time estimation of UnCI and perceived UnCI that was reported from the subject. As shown in Fig. 4.10 history plot, the circumplex model with RF regressor estimates the perceived UnCI with a small error. However, to validate the usefulness of the UnCI estimation system, a new experiment is required where the participant will evaluate the robot with CIES systems.

4.6 Limitations

While the results provided show great promise, it should be noted that the study was conducted in a lab room, which is a rather controlled environment. Subjects were essentially sedentary and only used one hand for the most part. Because it is generally recognized that motion artifacts affect physiological signals, data gathering in non-stationary conditions

can be difficult. Future studies should consider motion artifacts and create algorithms to eliminate them.

The proposed physiological computing system used multiple sensors for physiological signals. This is necessary for the system to work; however, it may be difficult to use such sensors sometimes. For example, the proposed system using pupillometry signals requires the participant to wear special glasses. However, if the participant already uses glasses, collecting the signals becomes challenging. However, as technology advances in the field of wearable sensors, this kind of challenge should be easier to deal with.

In the proposed research, we estimated arousal and valence from emotions, and the emotions' location on a circumplex model was approximated based on literature. However, it would be better to ask the participant to report arousal and valence during the trial. Then, arousal and valence can be used to estimate CI/UnCI.

4.7   Summary

In this chapter, we proposed a novel approach for human comfortability and uncomfortability index estimation by using subjective responses and physiological signals. One of the proposed approaches was inspired by Russel's circumplex model, which allows emotion to be represented in terms of arousal and valence [2]. The second approach was estimating CI/UnCI from arousal and valence by fitting a distribution obtained by KDE. To estimate arousal and valence, emotions were estimated from physiological signals.

Twenty subjects participated in the experiment. Physiological signals and subjective metrics were collected. The experiment was designed such that we collected the subjective metrics during the trial to have better signal labeling. Our results showed that estimating CI/UnCI from arousal and valence is a promising approach.

Our next experiment will focus on mimicking industrial settings. Hence, a conveyor and multiple robots will be added to the experiment to work alongside the human subject. This will help subjects to think he/she is in the factory floor working with robots.

In this research, arousal and valence were estimated from emotions, which is prone to error since the emotion estimation has noise. A better approach would be to ask the subject about their arousal and valence using a self-assessment manikin (SAM) approach [26]. In future experiments, we will ask the user this information as well as shown in 4.11.
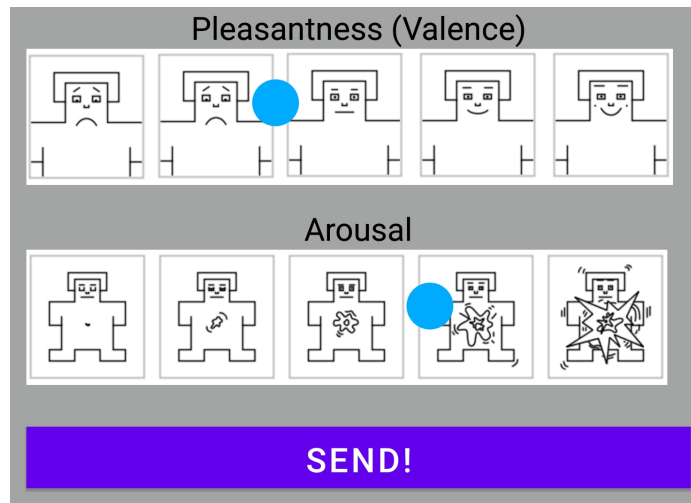
Figure 4.11: A screenshot of the tablet where the participant will report arousal and valence with a finger tap. The blue circle indicates reported valence and arousal by the subject.

In the next experiment, we will use the ML model trained in this research and validate it. The estimated uncomfortability will be used to modify the robot's velocity and cushioning distance between the human and the robot. As the human becomes less comfortable, the robot will reduce its velocity and increase the distance, and the robot will go to its default settings as the human becomes comfortable.

In conclusion, physiological metrics can be used to estimate the human comfortability index. The proposed method allows the CI and UnCI to be estimated only by two parameters: *arousal* and *valence*. In the current work, we calculated arousal and valence based on emotions; however, a better approach would be to ask subjects about their *arousal* and *valence* levels as it was done in [104]. In future experiments, we will ask participants to report their *arousal* and *valence* level during trials. We will develop a biocybernetics adaptation system [5] that estimates human UnCI during an HRC task in real-time and provide this information to the robot to adjust its behavior accordingly.

Chapter 5

Human Subject Experiment for CIES Model: Results and Discussion

The CIES model with dynamic SSM safety configurations is evaluated for an HRC setup where the human agent performs an assembly task with the robot. The experiment is performed with forty subjects and all the SSM safety configurations: TriSSM-Vo and TriSSM-Vr. This human subject research was approved by the IRB from Rochester Institute of Technology. Subjective responses are collected using questionnaires and interviews following each experiment. The human subjects are asked to rate the robot's behavior for a given SSM safety configuration in terms of comfortability, safety, emotions, and arousal-valence level. The emotions tracked include surprise, anxiety, calmness, and boredom.

The objective of this chapter is twofold. First, can the CIES model estimate human uncomfortability from the physiological signals? The second one is by using estimated uncomfortability, can CIES model makes participants less uncomfortable by changing the parameters of dynamic speed and separation monitoring. Our main hypotheses are to see whether there is a difference between the robot's behavior, the productivity of the participant and robot, and the effect of the distance vs. human estimated uncomfortability. The list below presents our main hypotheses in a general form.

 (i) CIES model modifies robot behavior by reducing directed velocity and increasing cushioning distance. Therefore, the participants would feel more comfortable when the dynamic speed and separation algorithms with the CIES model (VO_UNCI and VR_UNCI) are enabled than when they are not (VO and VR).

 (ii) The robot toggles between three modes (NORMAL, REDUCE, and STOP) based on the velocity and the distance between the robot and the human, as a result, the time spent in each mode would differ between robot behaviors. The time spent in REDUCE mode would be higher when the CIES model is active.

(iii) When a robot moves toward a human with a certain velocity, it would make the human uncomfortable. Therefore, the minimum distance between the human and the robot would affect perceived comfortability.

(iv) Similarly to the previous hypothesis if the robot increases the cushioning distance, that would make the person feel more comfortable. As a result, the participants' performance would be higher when the CIES model is active than when the CIES model is not.

In this chapter, the CIES model that trained on the Sawyer robot was evaluated in a non-stationary experiment. The robot's behavior adaptation based on the CIES model output is presented. The performance of the CIES model and its results are presented. In addition, the learning rate and effect of comfort on gender are discussed.

5.1 Experiment Setup
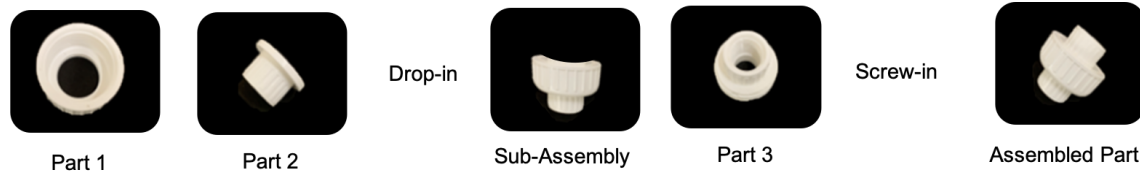
Figure 5.1: Parts used in the experiment. Three parts need to be assembled to complete a part.

The experimental setup consists of individual tasks for the robot and the human. The human task during the experiment was to assemble a product consisting of three parts, as shown in Fig. 5.1. One of the parts (Part-2) was provided by the robot and the other two parts (Part 1 and Part 3) were available in bins. The Sawyer robot picked parts from pallets (24 parts) and placed them on the conveyor, as shown in Fig. 5.2. Then the UR10 robot picked parts from the conveyor and placed the part in the bin on *Table-B* one by one, as shown in Fig. 5.2. The participant picked the components, in order from *Table-A*, *Table-B*, and *Table-C*, and assembled the part, then place it in the bin on *Table-D*, as shown in Fig. 5.3. During this process, the participant works in the shared workspace as shown in Fig.

5.4. After placing the part on Table-D, the subject was asked to record responses on the form provided in the tablet located next to Table-D.
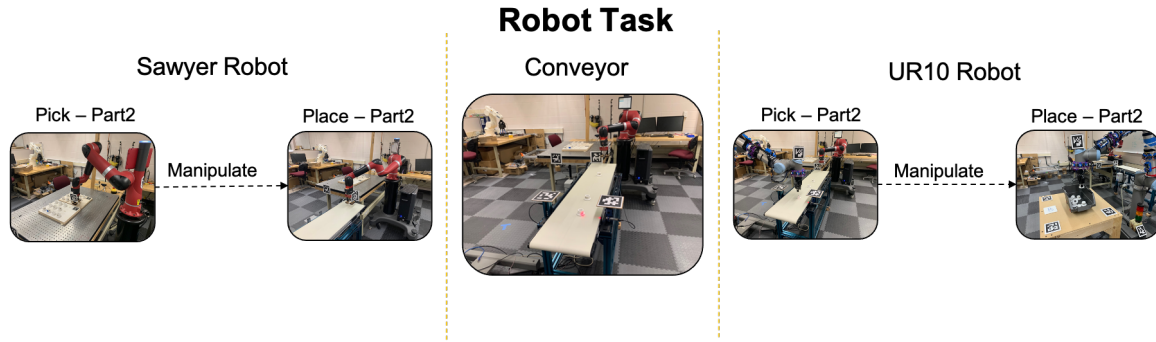
Figure 5.2: Sawyer and UR10 robots task, pick, and place location. There is a conveyor between Sawyer and UR10 robot.
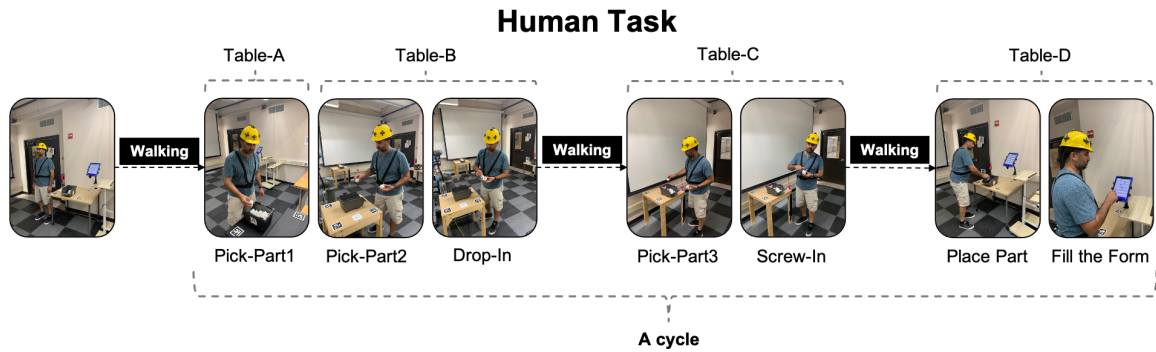
Figure 5.3: Human task is collecting parts from Table A, B, C and dropping the assembled part at Table-D.

This experiment consisted of 8 trials and each trial consisted of multiple cycles. A cycle is defined as a sequence of small tasks where a participant collects parts from Table-A, Table-B, and Table-C, assembles them, and then drops it onto Table-D, as shown in Fig. 5.3. The trials were randomized to reduce the order effect. Each trial took 6 minutes at max. In some cases, the experiment was taking less than 6 mins because the UR10 robot was completing all the parts. Hence, the trial is terminated either at 6 minutes limit or when there is no part left for the robot to pick up. Subjective responses of *comfortability*, *safety*, *surprise*, *anxiety*, *calmness*, *boredom*, *valence*, and *arousal* level were collected after each part was delivered to *Table-D* during each trial and after each trial.
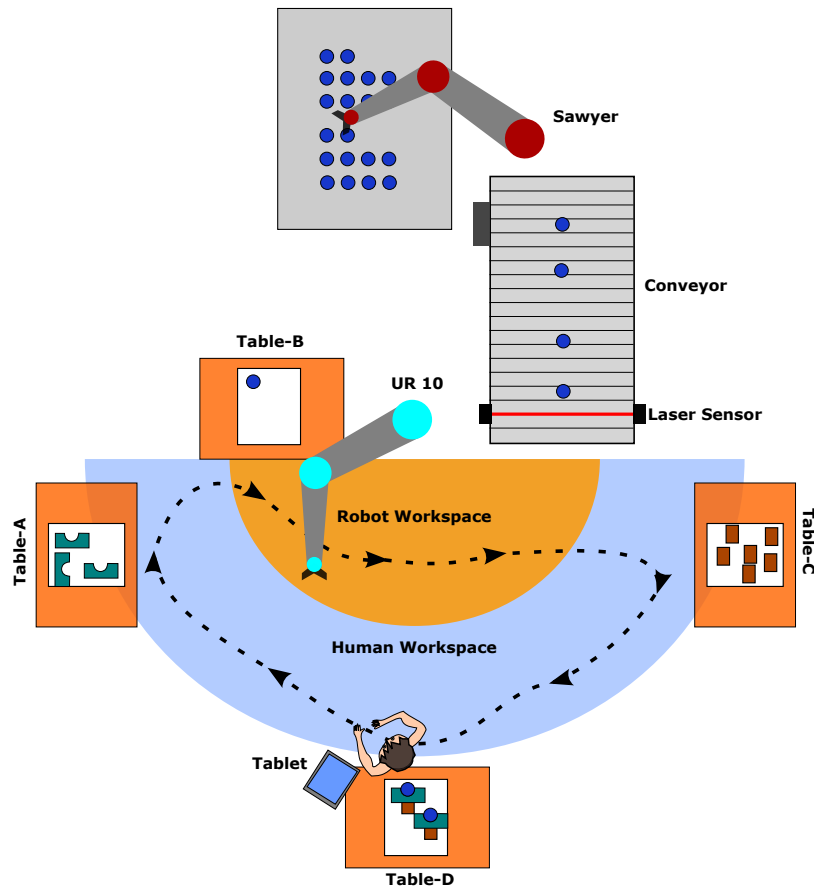
Figure 5.4: Overview of HRC experiment. The subject starts from Table-D and follows the path indicated, picks a component from Table-A then moves to Table-B, picking up the second component, and then crosses the robot workspace to go pick up the last component from Table-C. The subject then assembles the three components and places the part on Table-D. Finally, the subject answers the questionnaire on the Tablet. The subject repeats this cycle for approximately six minutes.

This experiment was designed to test the dynamic speed and separation monitoring (SSM) algorithm and CIES model with a time-of-flight sensor modality [12]. Table 5.1 lists the UR-10 robot behaviors. The TriSSM-Vo and TriSSM-Vr algorithms are explained in detail in [82].

Table 5.1: List of Robot's behaviors

Robot Algorithm	Description
TriSSM-Vo	A DSS algorithm taking into account both the robot's velocity and the human's velocity.
TriSSM-Vo-with-UnCI	UnCI metric provided to the TriSSM-Vo algorithm to adjust cushioning distance and directed speed.
TriSSM-Vr	A DSS algorithm taking into account only the robot's velocity, not the human's velocity.
TriSSM-Vr-with-UnCI	UnCI metric provided to the TriSSM-Vr algorithm to adjust cushioning distance and directed speed

5.1.1   Generalized Model (Subject Independent Model)

The CIES model was trained on the experiment conducted with the Sawyer robot (see chapter 4). The dataset was collected from twenty people. In order to have a generalized model (subject independent), the distribution of the output of the CIES model was used to select lower and upper thresholds. The lower threshold was selected as the mean of the distribution and the upper threshold was selected to be three standard deviations greater than the mean as shown in Fig. 5.5. These thresholds were selected to have a general model that works across multiple participants.
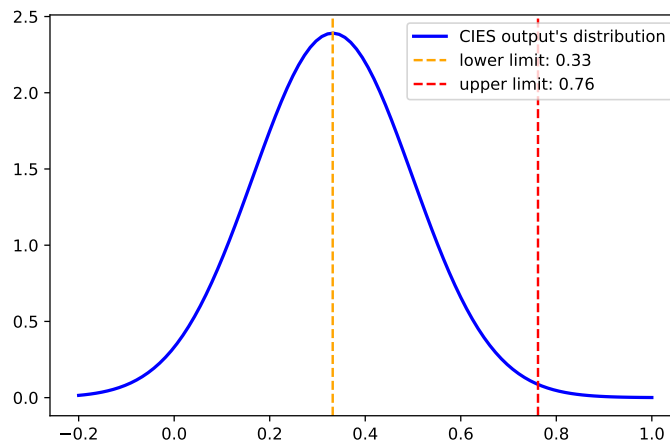
Figure 5.5: Distribution of output from uncomfortability estimation from CIES model

The mapped UnCI is calculated using equation 5.1

$$mapped_{unci} = interp(est_{unci}, [lower_{limit}, upper_{limit}], [0, 1])$$ (5.1)

where $mapped_{unci}$ is the general output, est_{unci} is estimated uncomfortability, $lower_{limit}$ is the mean of the output distribution of the CIES model, and $upper_{limiut}$ is three standard deviations from the mean. The $interp()$ is a function from NumPy [105] which applies interpolation between two ranges, and it selects the lower or upper bound if the point is not from the input range. For example, if we want to map 0.6 by using the $lower_{limit}$ (0.5), $upper_{limit}$ (0.8) and target range [0, 1] in $interp()$ function. It will be evaluated as $interp(0.6, [0.5, 0.8], [0, 1])$ and the result will be 0.33. However, if any input value is greater than the $upper_{limit}$, the $interp()$ function would evaluate as 1 which is the maximum of the target range. Similarly, if any input value is less than $lower_{limit}$ would map to 0 (minimum of target range).

5.1.2 Robot Behavior Adaptation

TriSSM-Vo and TriSSM-Vr algorithms were kept the same as Kumar et al. [82]. The other two algorithms (TriSSM-Vo-with-UncI and TriSS-Vr-with-UncI) were modified based on estimated $p_{unci} = mapped_{unci}$. Both algorithms were designed to reduce the directed velocity and increase the cushioning distance when a human is not feeling comfortable. Then as the human feels comfortable the velocity and cushioning distance return to their original value. Hence, the robot never increases its pace. The detail of TriSSM-Vo and TriSSM-Vr can be found in Kumar's thesis in [82].

Below is the formula for TriSMM-Vo-with-UnCI:

$$d_C(t_0) \geq \underbrace{(1 + p_{unci}k_{vo})v_o}_{directed\ speed}(T_R + T_{stop}) + v_R T_R + B + \underbrace{(1 + p_{unci}k_c)C_{dC}}_{cushioning\ constant}$$ (5.2)

where $d_C(t_0)$ is critical distance with respect to time, v_o is relative velocity between the robot and the human where the human's velocity is taken into account, T_R and T_{stop} are time

constants for robot to reduce and time to stop respectively, v_R is velocity of the robot, C_{dC} is cushioning constant, p_{unci} is estimated uncomfortability, k_{vo} and k_c are constants set to $k_{vo} = 0.5, k_c = 0.5$. This constant limits the effect of the CIES model on the DSS algorithm. In other words, the CIES model max 50% affects directed velocity and cushioning distance.

As it can be seen in equation 5.3 the p_{unci} has a scalar effect on both directed speed and cushioning distance. If the p_{unci} is low (close to 0.0), then there will not be any effect on the calculation of TriSSM-Vo. However, if p_{unci} gets larger (close to 1.0), then it will amplify both metrics. As a result, the robot will increase the cushioning distance and reduce the directed velocity. Hence, this will make the robot's cushioning distance more conservative when a co-worker is uncomfortable.

For the TriSSM-Vr-with-UnCI, we kept the effect of the uncomfortability the same. Below is the formula for TriSMM-Vr-with-UnCI:

$$d_C(t_0) \geq \underbrace{(1 + p_{unci}k_{vr})v_R}_{directed\ speed}(T_R + T_{stop}) + v_R T_R + B + \underbrace{(1 + p_{unci}k_c)C_{dC}}_{cushioning\ constant} \quad (5.3)$$

where $d_C(t_0)$ is critical distance with respect to time, v_R is robot's velocity, T_R and T_{stop} are time constants for robot to reduce speed and time to stop respectively, v_R is velocity of the robot, C_{dC} is cushioning constant, p_{unci} is estimated uncomfortability, and k_{vr} and k_c are constant set to $k_{vr} = 0.5, k_c = 0.5$.

It is important to mention that the TriSSM-Vr algorithm assumes that the human is stationary; thus, it only takes the robot's velocity into account. This is the main difference between TriSSM-Vr and TriSSM-Vo.

5.2 Subjective Data Collection

In order to evaluate robot behaviors, a list of questions was presented to the human subject after each cycle of a trial and after the trial. The questions are shown in Table 5.2. To make data entry easier an Android tablet explained in section 3.7 was used.

Table 5.2: A list of questions asked to subjects during and after the trials. During the trial, the subject response is recorded as a continuous number between [0, 1] based on a visual scale shown in Fig. 4.11 and when reporting after the trial the subject is presented with a Likert scale of [1, 9].

	Question	Response Type	Description
Q1	Comfortability level?	Range [0,1] Likert [1, 9]	How comfortable the subject felt while working with the robot. Subject informed that by comfortability, we do not mean physical comfort.
Q2	Safety level?	Range [0,1] Likert [1, 9]	How safe the subject felt while working with the robot.
Q3	Surprise level?	Range [0,1] Likert [1, 9]	How surprised the subject felt while working with the robot.
Q4	Anxiety level?	Range [0,1] Likert [1, 9]	What is the subject's anxiety level working with the robot?
Q5	Calmness level?	Range [0,1] Likert [1, 9]	How calm the subject felt while working with the robot.
Q6	Boredom level?	Range [0,1] Likert [1, 9]	How bored the subject felt while working with the robot. The subject was explicitly told to reset the boredom after each trial if they can.
Q7	Pleasantness level?	Range [0,1] Likert [1, 9]	How pleasant was the experience with the robot. This is indicating valence and it ranges from 0 to 1 where 0 indicates negative valence and 1 indicates positive valence.
Q8	Emotional intensity level?	Range [0,1] Likert [1, 9]	How much emotional intensity the subject felt while working with the robot. This is indicating the arousal and it ranges from LOW to HIGH.

Instead of directly asking *arousal* and *valence* levels, we rephrased the questions to make it easier to understand for the participants [106]. *Valence* was rephrased as 'pleasantness' and *arousal* was rephrased as 'emotional intensity'. Before the experiment, all

the questions were explained to the participants one by one until they understood all of the questions. The subject answered these questions at the end of each cycle during a trial via the Android tabled. At the end of the trial, the same questions were given in a Google form for the participants to evaluate their overall experience for the trial they had just completed [4].

5.3   Human Subjects Summary

For this experiment, the data were collected from healthy participants (N=40). The participants consist of 27 male and 13 female subjects. The demographics of the subjects for the experiment, in terms of sex, age, and familiarity with the robot (machinery), are shown in Table 5.3. The sample space consisted of people who had an average rating of 4.3 (SD=2.86) and self-reported in terms of experience and familiarity working with machinery between 1 and 9.

Table 5.3: The subjects' demographics regarding sex, age, and familiarity with machines/robots.

Subject Sex	Count	Avg. Age	Avg. rating in terms of experience and familiarity working with machinery such as CNC, lathe or robots, on a Likert Scale (1-Less Familiar to 9-very familiar)
Males	27	28.48 (±7.61)	5.07 (±2.85)
Females	13	26.61 (±7.83)	2.69 (±2.21)
Summary	**Total**: 40	Avg: 27.87 (±7.63)	Avg: 4.3 (±2.86)

5.4   Subjective Evaluation

This section will evaluate important hypotheses based on subjective responses and ML algorithm results. The data is collected from forty participants, and each participant repeats

eight trials, excluding baseline and training sessions. Table 5.4 shows number of trials per robot algorithms for both original and post-processed datasets. Few experiments and trials were removed due to sensor failure or technical problems during data collection. Thus, three subjects' data was removed due to sensor failure. For brevity, the robot behavior (algorithm) is referred to from here on in this form: TriSSM-Vo as VO, TriSSM-Vo-with-UnCi as VO_UNCI, TriSSM-Vr and VR, and TriSSM-Vr-with-UnCI as VR_UNCI.

Table 5.4 shows the number of trials per robot's algorithm for original and post-processed datasets. We will analyze our results with both datasets. In the original dataset, we kept the trial algorithm (VO, VR, VU_UNCI, and VR_UNCI) the same as what was used during data collection. Here we do not check whether the CIES model changes the robot's behavior or not. However, in the post-processing dataset, we can detect whether or not the CIES model triggers the robot's behavior; thus, we can correct the trial's algorithm. In the original dataset, in a few trials, either the CIES model did not trigger at all, or when it triggered, it did not have any change in the robot behavior. For this reason, we will analyze our results with the original and post-processed datasets separately.

Table 5.4: Number of trials per robot's algorithm for both original and post-processed datasets.

Robot Algorithm	Number of Trials (Original Dataset)	Number of Trials (Post-Processed Dataset)
VO	66	78
VO_UNCI	67	55
VR	64	72
VR_UNCI	69	61
Total	**266**	**266**

When creating the post-processed dataset, two steps were used. The first step was to fix the trial algorithm if the CIES output was always 0. For example, if a trial was set to be VO_UNCI and the CIES model output was 0, then the trial algorithm was set to VO. If the CIES did not change the robot's behavior, then it was VO, not VO_UNCI. Second, we detected all the locations where the CIES model provided an output value that was greater

than 0 in the trial. Then we detected the range and added a ±1 second margin. Once we had all the ranges, then we looked at the robot's behavior for those ranges. If the robot's behavior changed between NORMAL, REDUCE, and STOP during any of the ranges, we did not change the trial's algorithm (kept the original); otherwise, the trial's algorithm was set to either VO or VR, depending on the algorithm. For example, if the CIES model estimated uncomfortability only once during the trial (VO_UNCI) and the robot did not change its behavior, then the robot will behave as if its mode was VO. Therefore, setting the trial's mode to the VO would be logical.

In summary, Table 5.4 shows the number of trials for for original and post-processing datasets. It is important to mention that the number of trials for VO_UNCI and VR_UNCI is less for the post-processed dataset than for the original dataset. This happened because the CIES model could not change robot behavior during the trial for VO_UNCI and VR_UNCI. Thus, those trials were re-labeled as either VO or VR accordingly.

5.4.1   After Trial Rating for Robot's Behaviors

The participants rate the robot's behavior after each trial in terms of the question listed in table 5.2 on a Likert scale between 1 and 9 (1 is lowest and 9 is highest) by using a Google form. Each subject performs a series of trials that are randomized from Table 5.1.

The summary of the responses (37 participants) rating all the robot's behaviors is shown in Table 5.5 for the original and post-processed datasets. The ratings are averaged for a comparison of the robot's behavior. On average, both *VO_UNCI* and *VO* were rated with similar scores. However, *VO_UNCI* (7.612) is slightly lower than *VO* (7.773). FIn the second table, the difference between VO_UNCI and VO get smaller.

On the other hand, the CIES improves the average reported comfortability since the VR_UNCI (7.638) is greater than VO (7.188) for original labels. We observe that average comfortability is still higher for VR_UNCI for the post-processed dataset.

It is observed that for all robot behavior the overall perceived safety can be ordered as *VO_UNCI > VO > VR_UNCI > VR*. Regarding average surprise, the most surprising algorithm is *VR*. The algorithm that creates the least anxiety is *VO_UNCI*, and the *VO_UNCI* was rated as the calmest algorithm for both the original and the post-processed dataset.

In summary, *VO_UNCI* is the highest-rated algorithm in terms of comfortability and safety. The highest rated algorithm for surprise and anxiety is *VR*. Although all the average calmness levels are similar, the calmest algorithm was *VO_UNCI*.

Table 5.5: The summary of after-trial responses robot's behavior in terms of subjective metrics for Original and Post-Processed datasets

Original Dataset							
Algorithm	**Statistic**	**Comfort level?**	**Safety level?**	**Surprise level?**	**Anxiety level?**	**Boredom level?**	**Calmness level?**
VO	*AVG*	7.773	7.727	3.136	2.515	5.000	7.258
	STD	1.681	1.732	2.089	1.858	2.678	1.900
	Median	8.000	8.000	3.000	2.000	5.000	8.000
	STD-Error	0.207	0.213	0.257	0.229	0.330	0.234
VO_UNCI	*AVG*	7.612	7.761	3.269	3.030	4.582	7.090
	STD	1.907	1.587	2.384	2.342	2.595	2.179
	Median	8.000	8.000	2.000	2.000	5.000	8.000
	STD-Error	0.233	0.194	0.291	0.286	0.317	0.266
VR	*AVG*	7.188	7.312	3.859	3.047	4.531	7.344
	STD	2.152	2.210	2.754	2.535	2.576	2.191
	Median	8.000	8.000	3.000	2.000	4.500	8.000
	STD-Error	0.269	0.276	0.344	0.317	0.322	0.274
VR_UNCI	*AVG*	7.638	7.551	3.217	2.957	4.522	7.435
	STD	1.706	1.937	2.242	2.464	2.518	1.851
	Median	8.000	8.000	3.000	2.000	5.000	8.000
	STD-Error	0.205	0.233	0.270	0.297	0.303	0.223
Post-Processed Dataset							
Algorithm	**Statistic**	**Comfort level?**	**Safety level?**	**Surprise level?**	**Anxiety level?**	**Boredom level?**	**Calmness level?**
VO	*AVG*	7.705	7.603	3.205	2.615	4.923	7.205
	STD	1.676	1.804	2.158	1.922	2.662	1.909
	Median	8.000	8.000	2.500	2.000	5.000	8.000
	STD-Error	0.190	0.204	0.244	0.218	0.301	0.216
VO_UNCI	*AVG*	7.673	7.945	3.200	3.000	4.600	7.127
	STD	1.963	1.407	2.360	2.380	2.608	2.228
	Median	8.000	8.000	3.000	2.000	5.000	8.000
	STD-Error	0.265	0.190	0.318	0.321	0.352	0.300
VR	*AVG*	7.278	7.347	3.750	2.958	4.528	7.472
	STD	2.078	2.176	2.674	2.469	2.589	2.109
	Median	8.000	8.000	3.000	2.000	4.500	8.000
	STD-Error	0.245	0.256	0.315	0.291	0.305	0.249
VR_UNCI	*AVG*	7.590	7.541	3.262	3.049	4.525	7.295
	STD	1.764	1.946	2.301	2.533	2.494	1.909
	Median	8.000	8.000	3.000	2.000	5.000	8.000
	STD-Error	0.226	0.249	0.295	0.324	0.319	0.244

So far, we have looked at the average rating of reported metrics. Here we will define our hypotheses and see if our claims are statistically significant. For all hypotheses, we

are going to assume the mean difference is equal and we will only define expected outputs as a hypothesis. The desired hypotheses are bolded in the following equations. We will evaluate each hypothesis for both datasets.

H_a^1 : The median perceived comfortability rating for VO_UNCI (μ_{VO_UNCI}) is greater than VO (μ_{Vo}).

$$H1 = \left\{ H_0^1 : \mu_{VO\_UNCI} = \mu_{VO}, \quad \mathbf{H_a^1} : \mu_{VO\_UNCI} > \mu_{VO} \right\} \tag{5.4}$$

H_a^2 : The median perceived comfortability rating for VR_UNCI (μ_{VO_UNCI}) is greater than VR (μ_{VO}).

$$H2 = \left\{ H_0^2 : \mu_{VR\_UNCI} = \mu_{VR}, \quad \mathbf{H_a^2} : \mu_{VR\_UNCI} > \mu_{VR} \right\} \tag{5.5}$$

H_0^3 : The median perceived comfortability rating for VR_UNCI (μ_{VR_UNCI}) is as good as VO (μ_{VO}).

$$H3 = \left\{ \mathbf{H_0^3} : \mu_{VO} = \mu_{VR\_UNCI}, \quad H_a^3 : \mu_{VO} > \mu_{VR\_UNCI} \right\} \tag{5.6}$$

In the H_0^3, we aim to see if VR_UNCI is as good as VO in terms of comfortability. As we defined the VO and the VR algorithm in Section 2.5, the VO algorithm uses the relative velocity and distance between the human and the robot. And the VR algorithm uses the robot-directed speed and the distance between the human and the robot. In the VO algorithm, a human is a dynamic object in the workspace; however, in the VR algorithm, the human is a static object in the workspace. For example, if a human is coming toward the robot with a velocity, the VO algorithm will slow down and stop earlier than the VR since the relative velocity will be high toward the robot. Conversely, if the human moves away from the robot, the VO algorithm will not stop or slow down since the relative velocity will be away from the robot; however, the VR algorithm will work in the same way as the human coming toward the robot. By introducing the CIES model into the equation, would it be possible to make the VR algorithm as comfortable as the VO?

In order to measure the statistical significance of the differences of the *comfort*, *safety*,

surprise, *anxiety*, *boredom*, and *calmness* for the *VO*, *VO_UNCI*, *VR*, and *VR_UNCI* robot's behaviors, a one-tailed non-parametric (Wilcoxon) pairwise t-test for the two-sample populations was performed. The statistics and the resulting p-value from the t-test are shown in Table 5.6.

In this analysis, we used statistical significant value ($\alpha = 0.05$). Our first hypothesis (H_a^1) is that the CIES model would improve the *VO* algorithm in terms of perceived comfortability. However, as the p-value is 0.305, which is greater than the significance value, we failed to reject the null hypothesis (H_0^1). In the second hypothesis (H_0^2), we want to see if the CIES model improves VR in terms of comfortability, the test result shows a p-value of 0.037. Thus, we accepted the alternative hypothesis (H_0^2). The CIES model improved the VR algorithm and made it more comfortable based on after-trial data collection metrics for the original dataset.

Kumar et al. [82] claimed that the VO algorithm is more comfortable than the VR algorithm. Thus, we want to see whether the CIES model makes VR as comfortable as the VO algorithm (VO without the CIES model), which is our null hypothesis (H_a^3). From the statistic, the p-value is 0.541; thus, we reject the alternative and fail to reject the null hypothesis. This is interesting because VO is not more comfortable than VR_UNCI statistically. However, it is important to look at whether VO is more comfortable than VR as well. The result shows a p-value of 0.060 for $\mu_{VO} > \mu_{VR}$. Although we cannot say that VO is more comfortable than VR, what we can see is the p-value for VR_UNCI is greater than VR. This indicates that the CIES model improves the VR algorithm in terms of perceived comfortability. It is important to mention these results are from the after-trial responses for the original dataset.

In addition to perceived comfortability, we look into other metrics as well. The perceived safety for VO_UNCI is statistically significant, VR ($\mu_{VO\_UNCI} > \mu_{VR}$), with a $p-val = 0.006$. Although *VO* is not safer than *VR*, *VO_UNCI* is safer than *VR*. Hence, we can see

that the CIES model helps the *VO* algorithm to be perceived as safer by the participants.

When looking at the post-processed dataset, we observed a similar pattern. However, there are a couple of differences in terms of statistical results. In the original dataset, VR_UNCI was more comfortable than VR. However, it was not since the p-value is greater than the α value here. Similarly, VO was more boring than VR_UNCI, but we cannot say the same thing with the post-processed dataset. Finally, in the post-processed dataset, the VO_UNCI is safer than VR_UNC. The rest of the analyses are the same as the original dataset.

Table 5.6: The result of a non-parametric (Wilcoxon) pairwise test to measure the statistical significance of the differences in medians of the after-trial ratings for robot behaviors for original and post-processed datasets.

A	Cond.	B	Comfort	Safety	Surprise	Anxiety	Boredom	Calmness
Original Dataset								
VO	<	VO_UNCI	0.305	0.146	0.677	0.204	0.965	0.387
VO	>	VR	0.060	0.164	0.839	0.790	0.015	0.892
VO	>	VR_UNCI	0.541	0.293	0.390	0.961	0.014	0.766
VO_UNCI	>	VR	0.017	0.006	0.956	0.712	0.262	0.827
VO_UNCI	>	VR_UNCI	0.217	0.119	0.586	0.683	0.336	0.668
VR	<	VR_UNCI	0.037	0.133	0.962	0.402	0.433	0.385
Post-Processed Dataset								
A	Cond.	B	Comfort	Safety	Surprise	Anxiety	Boredom	Calmness
VO	<	VO_UNCI	0.167	0.094	0.686	0.621	0.956	0.267
VO	>	VR	0.089	0.229	0.750	0.678	0.019	0.942
VO	>	VR_UNCI	0.410	0.286	0.324	0.886	0.051	0.455
VO_UNCI	>	VR	0.013	0.005	0.852	0.927	0.300	0.733
VO_UNCI	>	VR_UNCI	0.134	0.015	0.470	0.825	0.526	0.403
VR	<	VR_UNCI	0.092	0.253	0.949	0.514	0.318	0.669
Significance ($\alpha = 0.05$)								

In terms of surprise, *VO_UNCI* is less surprising than *VR*, and *VR* is more surprising than *VR_UNCI*. For anxiety, *VO* triggers less anxiety in comparison to both *VR_UNCI*. For boredom, the *VO* algorithm is more boring than *VR* and *VR_UNCI* and less boring than *VO_UNCI*. Finally, there is no statistical significance for calmness between all the robot behaviors. It is important to mention these results are based on the after-trial responses (Original dataset) for which participants were evaluated six minutes (a full trial) with the

original trial's algorithm. In our previous research [4], we have seen that there might be differences between after-trial responses vs. during-trial responses.

The during-trial responses are likely to yield a better result because the participant is required to report multiple times while undergoing the trial. The responses capture more information about the participant's state throughout the trial. This is critical for capturing perceived comfortability as it may change multiple times during the trial.

5.4.2 During Trial Rating for Robot's Behaviors

During-trial responses were collected while undergoing the trial. The participant was asked to report after each cycle (a cycle is defined as completing an assembled part shown in 5.3). Each cycle took 15 seconds on average. As the responses were collected during the trial, it provides additional information about the progress of the human's emotional and physiological state [4].

Like after-trial responses, during-trial responses are summarized based on the robot's behavior and shown in Table 5.7 for both original and post-processed datasets. The rating for during-trial is a range between 0 and 1. It can be observed that *VO* and *VO_UNCI* have similar average ratings in terms of *comfort*, *safety*, *surprise*, *anxiety*, *boredom*, and *calmness*. Similarly, *VR* and *VR_UNCI* have the same ratings. When we compare *VOs* (VO and VO_UNCI) with the *VRs* (VR and VR_UNCI) algorithm, it is clear that *VOs* are better in all reported metrics for both datasets. When we look at the median comfort level for all four algorithms, the mean/median values for the CIES algorithm are higher. Hence, this shows that participants felt more comfortable when CIES was active.

As we have seen there is a difference in average rating from the participants from Table 5.7. Table 5.8 shows the non-parametric (Wilcoxon) pairwise test to measure statistical difference in medians. We are analyzing the same hypothesis (see hypothesis 5.4, 5.5, 5.6) with the alternative data collection method of during trial. In this analysis, we used

Table 5.7: The summary of during-trial response robot's behavior for Original and Post-Processed datasets

Original Dataset							
Algorithm	**Statistic**	**Comfort level?**	**Safety level?**	**Surprise level?**	**Anxiety level?**	**Boredom level?**	**Calmness level?**
VO	*AVG*	0.877	0.869	0.208	0.176	0.476	0.783
	STD	0.149	0.162	0.185	0.175	0.313	0.192
	Median	0.923	0.926	0.148	0.134	0.486	0.842
	STD-Error	0.018	0.020	0.023	0.022	0.039	0.024
VO_UNCI	*AVG*	0.863	0.866	0.229	0.193	0.436	0.774
	STD	0.178	0.157	0.214	0.203	0.289	0.210
	Median	0.922	0.908	0.169	0.146	0.443	0.826
	STD-Error	0.022	0.019	0.026	0.025	0.035	0.026
VR	*AVG*	0.836	0.824	0.252	0.209	0.409	0.759
	STD	0.201	0.207	0.222	0.222	0.308	0.233
	Median	0.900	0.902	0.178	0.113	0.317	0.812
	STD-Error	0.025	0.026	0.028	0.028	0.038	0.029
VR_UNCI	*AVG*	0.856	0.850	0.230	0.191	0.429	0.775
	STD	0.176	0.172	0.215	0.195	0.289	0.209
	Median	0.895	0.898	0.157	0.136	0.420	0.810
	STD-Error	0.021	0.021	0.026	0.023	0.035	0.025
Post-Processed Dataset							
Algorithm	**Statistic**	**Comfort level?**	**Safety level?**	**Surprise level?**	**Anxiety level?**	**Boredom level?**	**Calmness level?**
VO	*AVG*	0.867	0.856	0.219	0.183	0.463	0.774
	STD	0.150	0.171	0.188	0.182	0.312	0.194
	Median	0.897	0.907	0.158	0.146	0.460	0.807
	STD-Error	0.017	0.019	0.021	0.021	0.035	0.022
VO_UNCI	*AVG*	0.875	0.883	0.219	0.186	0.446	0.785
	STD	0.183	0.140	0.217	0.201	0.286	0.213
	Median	0.940	0.918	0.165	0.121	0.456	0.844
	STD-Error	0.025	0.019	0.029	0.027	0.039	0.029
VR	*AVG*	0.840	0.828	0.244	0.203	0.409	0.764
	STD	0.194	0.202	0.216	0.215	0.309	0.225
	Median	0.900	0.902	0.178	0.113	0.322	0.812
	STD-Error	0.023	0.024	0.025	0.025	0.036	0.027
VR_UNCI	*AVG*	0.855	0.849	0.236	0.196	0.432	0.770
	STD	0.182	0.175	0.221	0.200	0.286	0.216
	Median	0.895	0.898	0.157	0.136	0.420	0.810
	STD-Error	0.023	0.022	0.028	0.026	0.037	0.028

a significant value ($\alpha = 0.05$) as well. Our first hypothesis H_a^1 is that the CIES model would improve the *VO* algorithm in terms of perceived comfortability. However, as the

p-value is 0.541 which is greater than the significance value, we failed to reject the null hypothesis (H_0^1). Similarly, in the second hypothesis (H_0^2)) we want to see if the CIES model improves VR, the test result shows a p-value of 0.105, as a result, we failed to reject the null hypothesis.

Before analyzing H^3, it is important to consider whether the VO is more comfortable than VR. From the table 5.8, the p-value is 0.011 for the original and 0.042 for the post-processed datasets. Thus we rejected the null hypothesis. Now, we know that VO is more comfortable than VR, but is this correct that the VO is more comfortable than VR_UNCI? In the H_a^3, we check whether VO is better than VR_UNCI algorithms in terms of comfortability. The p-value is 0.124 for the original and 0.078 for the post-processed datasets. Hence we cannot say VO is better than VR_UNCI. As a result, we can conclude that the CIES model improved the VR algorithm in terms of comfortability.

Table 5.8: The result of a non-parametric (Wilcoxon) pairwise test to measure the statistical significance of the differences in medians of the during trial ratings for robot's behavior

A	Cond.	B	Comfort	Safety	Surprise	Anxiety	Boredom	Calmness
Original Dataset								
VO	<	VO_UNCI	0.549	0.264	0.516	0.827	0.805	0.314
VO	>	VR	0.011	0.002	0.919	0.889	0.003	0.204
VO	>	VR_UNCI	0.124	0.086	0.782	0.697	0.074	0.400
VO_UNCI	>	VR	0.003	0.001	0.929	0.958	0.091	0.042
VO_UNCI	>	VR_UNCI	0.041	0.022	0.935	0.959	0.153	0.124
VR	<	VR_UNCI	0.105	0.124	0.805	0.736	0.071	0.464
Post-Processed Dataset								
A	Cond.	B	Comfort	Safety	Surprise	Anxiety	Boredom	Calmness
VO	<	VO_UNCI	0.129	0.048	0.884	0.993	0.581	0.109
VO	>	VR	0.042	0.003	0.812	0.783	0.004	0.138
VO	>	VR_UNCI	0.078	0.011	0.652	0.867	0.236	0.363
VO_UNCI	>	VR	0.005	0.001	0.871	0.993	0.042	0.007
VO_UNCI	>	VR_UNCI	0.002	0.002	0.995	0.999	0.340	0.038
VR	<	VR_UNCI	0.289	0.371	0.629	0.541	0.040	0.484
Significance (α = 0.05)								

In addition, we can see that both *VO* and *VO_UNCI* algorithms are better than *VR* and *VR_INCI* in terms of comfort and safety. The p-value is less than the significant

value of $\alpha = 0.05$. In terms of surprise, VR_UNCI is more surprising than VO_UNCI for post-processed datasets, and there is no difference between other robots' algorithms. Only VO is more boring than VR for the original dataset, and there is no difference between algorithms. However, for the post-processed dataset, $VO > VR$, $VO\_UNCI > VR$, and $VR\_UNCI > VR$.

In conclusion, VO and VO_UNCI algorithms are more comfortable and safer but trigger more anxiety in comparison to VRs algorithms.

5.4.3 Analysis of Robot's Modes

Tri-Modal DSS has three levels: *NORMAL*, *REDUCE*, and *STOP* modes. Here we are going to analyze the duration of each mode during trials. The MAX trial time is set to six minutes however, in some cases (when a subject is fast), the participant may finish the task before the given time. Therefore, we will analyze the trial percentage time instead of the average time. Thus, we used the formula in equation 5.7 to calculate the percentage of the duration

$$p = \frac{mode_{time}}{total_{time}} \tag{5.7}$$

where p is a percentage, $mode_{time}$ is the total amount of time in seconds that the robot stays in one of the modes: *NORMAL*, *REDUCE*, and *STOP*, and $total_{time}$ is trial duration in seconds.

Table 5.9 shows the summary of the average percent time spent in each mode for the robot algorithms (DSS and DSS_CIES) during trials for both original and post-processed datasets. It can be observed that the REDUCE and STOP percent times are higher for DSS_CIES than DSS. The NORMAL average percent time is slightly higher for DSS for both datasets. In order to see if these changes are statistically significant between robot algorithms, we will need to test the hypothesis.

Our hypothesis for the robot's percent time spent in each mode is listed below:

Table 5.9: NORMAL, REDUCE, and STOP values as the percentage of the whole trial for both datasets of original and post-processed.

Dataset	Algorithm	Statistic	NORMAL	REDUCE	STOP
Original Dataset	DSS	*AVG*	0.839	0.070	0.090
		STD	0.041	0.020	0.032
		Median	0.842	0.066	0.088
		STD-Error	0.004	0.002	0.003
	DSS_CIES	*AVG*	0.829	0.073	0.098
		STD	0.046	0.024	0.035
		Median	0.833	0.071	0.093
		STD-Error	0.004	0.002	0.003
Post-Processed Dataset	DSS	*AVG*	0.837	0.070	0.091
		STD	0.041	0.021	0.031
		Median	0.842	0.065	0.090
		STD-Error	0.003	0.002	0.003
	DSS_CIES	*AVG*	0.830	0.073	0.097
		STD	0.047	0.023	0.037
		Median	0.834	0.072	0.090
		STD-Error	0.004	0.002	0.003

H_a^4 : The median percent time in NORMAL mode for DSS (μ_{DSS}) is greater than DSS_CIES (μ_{DSS_CIES}).

$$H4 = \left\{ H_0^4 : \mu_{DSS} = \mu_{DSS\_CIES}, \quad \mathbf{H_a^4} : \mu_{DSS} > \mu_{DSS\_CIES} \right\} \tag{5.8}$$

H_a^5 : The median percent time in REDUCE mode for DSS_CIES (μ_{DSS_CIES}) is greater than DSS (μ_{DSS}).

$$H5 = \left\{ H_0^5 : \mu_{DSS\_CIES} = \mu_{DSS}, \quad \mathbf{H_a^5} : \mu_{DSS\_CIES} > \mu_{DSS} \right\} \tag{5.9}$$

H_a^6 : The median percent time in STOP mode for DSS (μ_{DSS}) is greater than DSS_CIES (μ_{DSS_CIES}).

$$H6 = \left\{ H_0^6 : \mu_{DSS} = \mu_{DSS\_CIES}, \quad \mathbf{H_a^6} : \mu_{DSS} > \mu_{DSS\_CIES} \right\} \tag{5.10}$$

Table 5.10 shows the resulted p-values of the one way non-parametric (Wilcoxon) pairwise test for robot modes. In the first hypothesis H_a^4, the claim is that the robot would stay in NORMAL mode for DSS more than DSS_CIES. The p-value for this hypothesis is 0.028 and 0.078 for the original dataset and post-processed datasets, respectively. Thus, we

accepted the alternative for the original dataset and failed to reject the null hypothesis for the post-processed dataset.

In the H^5 hypothesis, we expect the percent REDUCE time to be higher for DSS_CIES. The p-value is lower than the $\alpha = 0.05$ value for both datasets. As a result, we reject the null hypothesis and accept the alternative hypothesis. This is expected because the CIES model makes the robot switch to REDUCE mode by increasing the cushioning distance and reducing the velocity; hence, this would have increased the time spent in REDUCE mode. Finally, hypothesis H^6 claims that the percent STOP time is higher for DSS_CIES than DSS. The p-values for STOP modes are 0.007 (original) and 0.042 (post-processed) for datasets. Similar to the REDUCE mode, as the CIES model changes cushioning distance and robot's velocity, it stops more than DSS algorithms.

Table 5.10: The result of a non-parametric (Wilcoxon) pairwise test to measure the statistical significance of the differences in medians of the robot's modes for the robot's behavior

Dataset	A	Cond.	B	Mode	p-value
Original Dataset	DSS	>	DSS_CIES	*NORMAL*	0.028
	DSS_CIES	>	DSS	*REDUCE*	0.050
	DSS_CIES	>	DSS	*STOP*	0.007
Post-Processed Dataset	DSS	>	DSS_CIES	*NORMAL*	0.078
	DSS_CIES	>	DSS	*REDUCE*	0.016
	DSS_CIES	>	DSS	*STOP*	0.042
Significance ($\alpha = 0.05$)					

Box plots shown in Fig. 5.6 are the robot's modes with respect to robot behaviors. The top left subplot shows the robot's NORMAL mode, and it can be seen that DSS and DSS_CIES algorithms have similar values. However, DSS_CIES has less variance, and there is no outlier. On the other hand, DSS algorithms have a large variance with multiple outliers. The REDUCE subplot shows a variety of algorithms where the DSS has a low median with few outliers and DSS_CIES has a high median with no outliers. For the STOP mode, both modes DSS and DSS_CIES have a similar average median value.

There are commonly used robot metrics used to measure collaboration metrics, such as
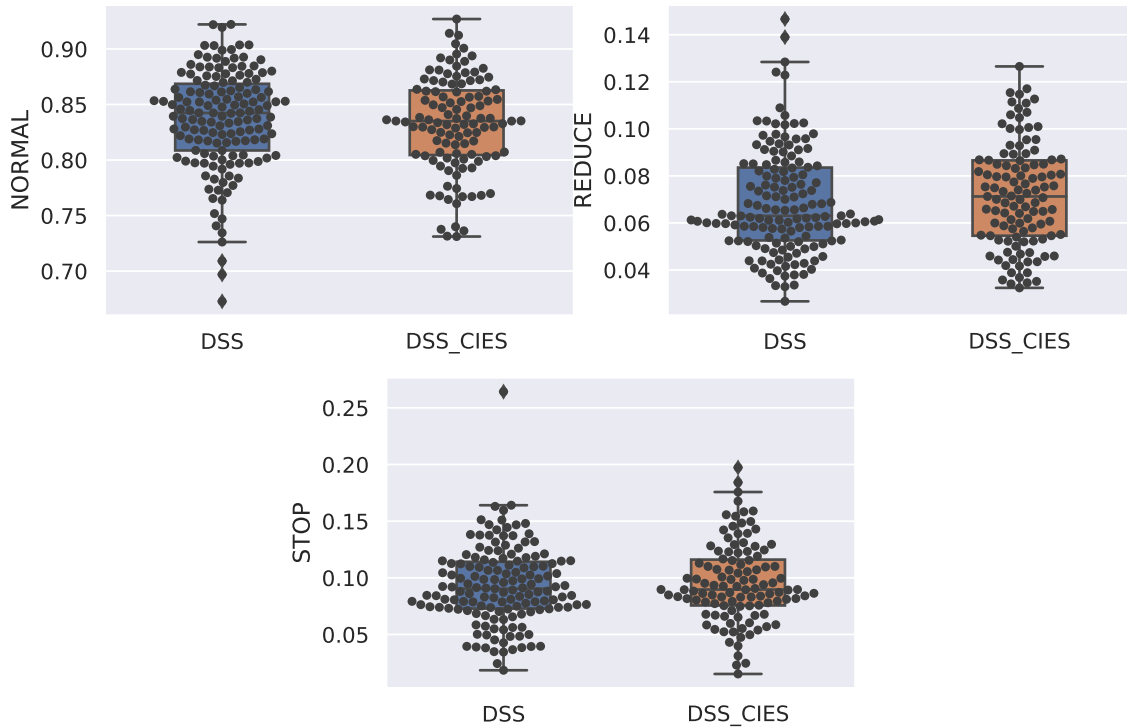
Figure 5.6: Box plots for percentage time of robot modes (NORMAL, REDUCE, and STOP) for original dataset

robot idle time and human idle time, as discussed in [107]. Table 5.11 shows the average time spent in robot modes of NORMAL, REDUCE, and STOP for four robot algorithms. In this experiment, the robot idle time is designated as robot STOP time. The robot stop time is $VO\_UNCI > VO > VR\_UNCI > VR$ for both datasets. This order shows the effect of the CIES model on the robot algorithm.

Next, we will discuss the effect of distance on estimated uncomfortability, which is critical as we expect humans to feel less comfortable when facing the robot.

5.4.4 Estimated Uncomfortability vs. Robot's Behavior

Presented here is analysis of estimated uncomfortability (see section 3.1.4) for robot behaviors. Table 5.12 shows non-parametric (Wilcoxon) pairwise test for filtered_unci and mapped_unci during different algorithms.

Table 5.11: Average time (seconds) spent in each robot's modes

Label Type	Algorithm	NORMAL		REDUCE		STOP	
		AVG	*STD*	*AVG*	*STD*	*AVG*	*STD*
Original Label	VO	301.311	12.135	21.045	5.047	35.061	10.659
	VO_UNCI	300.036	15.327	20.901	5.667	37.689	12.137
	VR	298.372	16.029	28.844	6.780	29.029	11.610
	VR_UNCI	294.121	15.980	31.300	7.928	32.589	12.850
Post-Processed Label	VO	301.222	11.722	20.848	4.921	35.435	10.259
	VO_UNCI	299.883	16.382	21.149	5.945	37.732	12.955
	VR	297.399	16.382	29.404	7.553	29.749	11.377
	VR_UNCI	294.712	15.737	30.960	7.348	32.207	13.386

H_a^7 : The *VR* algorithm would be less comfortable than *VO* algorithm.

$$H7 = \left\{ H_0^7 : \mu_{VO} = \mu_{VR}, \quad \mathbf{H_a^7} : \mu_{VR} > \mu_{VO} \right\} \tag{5.11}$$

H_a^8 : The *VR_UNCI* algorithm would be as good as *VO* in terms of estimated comfortability.

$$H8 = \left\{ \mathbf{H_0^8} : \mu_{VR\_UNCI} = \mu_{VO}, \quad H_a^8 : \mu_{VR\_UNCI} > \mu_{VO} \right\} \tag{5.12}$$

Table 5.12 shows the pairwise non-parametric Wilcoxon test results. In H^7, we claimed that the VR is less comfortable than the VO algorithm or that VR is more uncomfortable since we are evaluating estimated uncomfortability. The result from the test showed that the VR algorithm is not more uncomfortable than VO. Hence we cannot say that VR is more uncomfortable as the p-value is 0.369. As a result, we failed to reject the null hypothesis for H^7.

In H^8, we claimed that VR_UNCI is as comfortable as VO. As we are focusing on estimated uncomfortability, we can look at if VR_UNCI is more uncomfortable than V0 (H_a^8). The p-value for this analysis is 0.373, which is greater than the α value. Hence, we cannot say that VR_UNCI is more uncomfortable than VO. Before concluding this hypothesis, looking at the *VR* algorithm is critical. The *VR* algorithm is less comfortable (more uncomfortable) [82]. As shown in Table 5.12, there is no difference between the *VO* and *VR* algorithm in terms of uncomfortability. Although we failed to reject the null hypothesis

for H_0^8, it is difficult to say this happened due to the CIES model contribution to the *VR* algorithm. Thus, we cannot conclude whether the CIES model improved the comfortability of the *VR* algorithm. It is important to mention that these results were drawn from estimated uncomfortability. There are two reasons that we cannot see any difference regarding estimating uncomfortability. The first reason is that the subject did not feel the change in the robot's behavior, which did not trigger a physiological state change. The second reason is there were limited instances where the subjects felt uncomfortable. Both VO and VR algorithms are safe; we were trying to make them safer and more comfortable via the CIES model. To evaluate the real performance of the CIES model, a custom experiment that focuses on making humans uncomfortable is required.

Table 5.12: The result of a non-parametric (Wilcoxon) pairwise test to measure the statistical significance of the differences in medians of the estimated uncomfortability (mapped) for the robot's behavior

A	Cond.	B	Original Dataset	Post-Processed Dataset
			UNCI	UNCI
VO	<	VO_UNCI	0.690	0.541
VO	<	VR	0.369	0.549
VO	<	VR_UNCI	0.373	0.318
VO_UNCI	<	VR	0.297	0.541
VO_UNCI	<	VR_UNCI	0.020	0.024
VR	<	VR_UNCI	0.131	0.042
Significance ($\alpha = 0.05$)				

Table 5.12 shows that VR_UNCI is more uncomfortable than VO_UNCI for both datasets. We also observed that for the post-processed dataset, VR_UNCI is more uncomfortable than VR. This is interesting because we expect the CIES model to make VR more comfortable.

5.4.5   Uncomfortability vs. Distance between Human and Robot

In non-stationary, both the human and the robot move during the trial, and the distance changes. In addition, when the robot and the human are going toward each other, this

creates a dangerous scene since the relative velocity is high. Therefore, we look at the distance between the human and the robot versus estimated uncomfortability.

Fig. 5.7 shows distance versus uncomfortability during a trial. The y-axis shows the distance between human and robot as measured by the ToF ring on the robot. The x-axis shows the time elapsed during the trial in seconds. From the subjects' responses, we can see that whenever the robot comes close to the person, the person's uncomfortability increases, and while the robot is away from the person, their uncomfortability decreases.
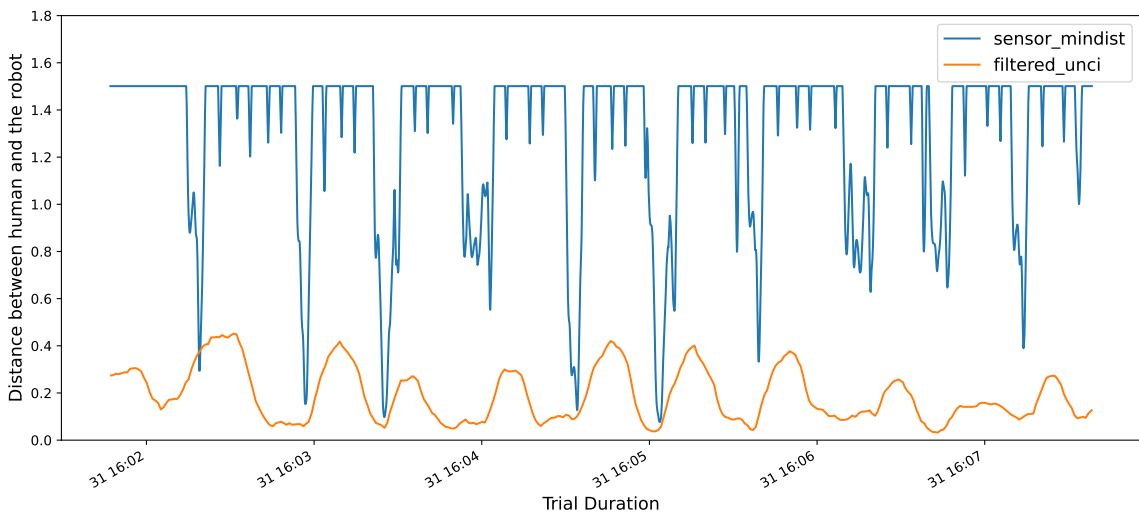
Figure 5.7: The estimated uncomfortability and distance between human and the robot during one of the trials (the participant (P31) at trial 2)

In Fig. 5.7 we observed how a person's uncomfortability changes throughout a trial. Thus, we looked at all the trials where the distance between the human and the robot was less than 0.5 meters, and we marked those events. Then we look at the estimated uncomfortability 2 secs before and 10 secs after the events and check whether the uncomfortability is increased, not changed, or decreased as shown in Fig. 5.8. Once the 12-seconds window is selected, the estimated uncomfortability at the beginning and end of the window are compared. These events are categorized into three groups: uncomfortability increased more than 5%, decreased less than -5%, and the no change in uncomfortability and the

events are labeled as 'increased', 'decreased', and 'no changes' respectively.
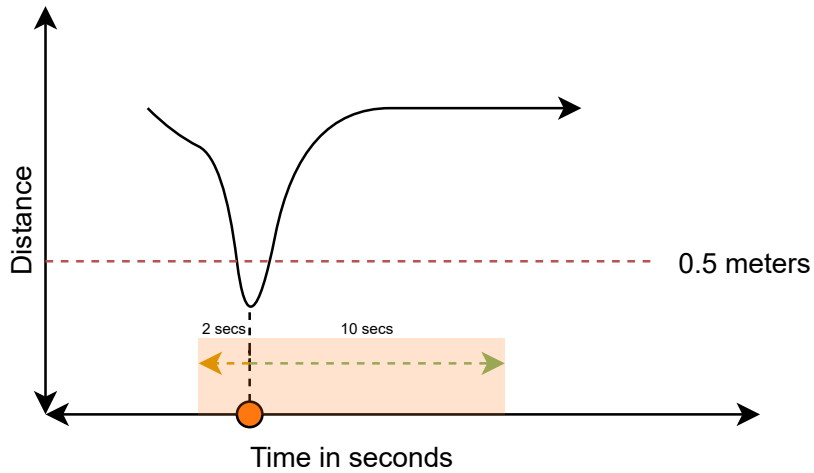
Figure 5.8: An example of event processing based on the distance between the human and the robot.

Table 5.13 shows there were 2543 total events for which the distance between human and robot was less than 0.5 meters. Among all the events, 44.5% of these events, human uncomfortability increased, 31.2% of times it decreased, and 24.3% times the change was minimal (5% or less).

Table 5.13: Tally of uncomfortability events for which the distance between human and robot was less than 0.5 meters.

Events	Number of Events	%
UnComfortability Increased (>5%)	1130	**44.5%**
UnComfortability Decreased (<5%)	795	31.2%
UnComfortability within (5%)	618	24.3%
Total	*2543*	*100%*

Fig. 5.7 shows that there is a correlation between estimated uncomfortability and distance between the human and the robot. This is expected since getting closer to the robot will increase the chance of collision; thus, participants feel uncomfortable. Thus changing robot trajectory to maximize the distance between human and robot would be one way of making the human more comfortable. It is important to mention that in some cases, we

may not be able to change the robot's trajectory due to limited space or other limitations.

5.4.6 AV Domain and Emotions

Unlike the stationary (Sawyer) experiment, arousal and valence data was collected from the participant during the experiment. Here, we are going to discuss arousal and valence regarding comfortability and emotions. First, we will evaluate the AV domain with comfortability, and then we will look at the emotions reported by the participants.
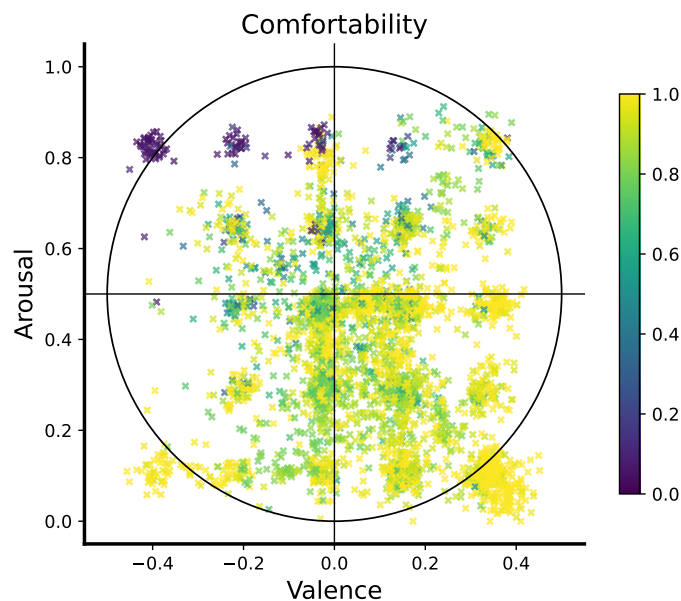
Figure 5.9: Reported arousal and valence rating highlighted based on perceived comfortability

Fig. 5.9 highlights AV points with respect to perceived comfortability. The high comfortability (yellow) points are mainly located in quadrant IV, and the low comfortability points are in quadrant II close to the edge. In section 3.1.4, the arousal and valence-based model was introduced. The result from the stationary (Sawyer) experiment showed that comfortability is in quadrant IV.

Fig. 5.10 shows the emotions (*anxiety*, *surprise*, *boredom*, and *calmness*) with respect to arousal and valence. The circle on the figures shows Russel's circumplex model. As we
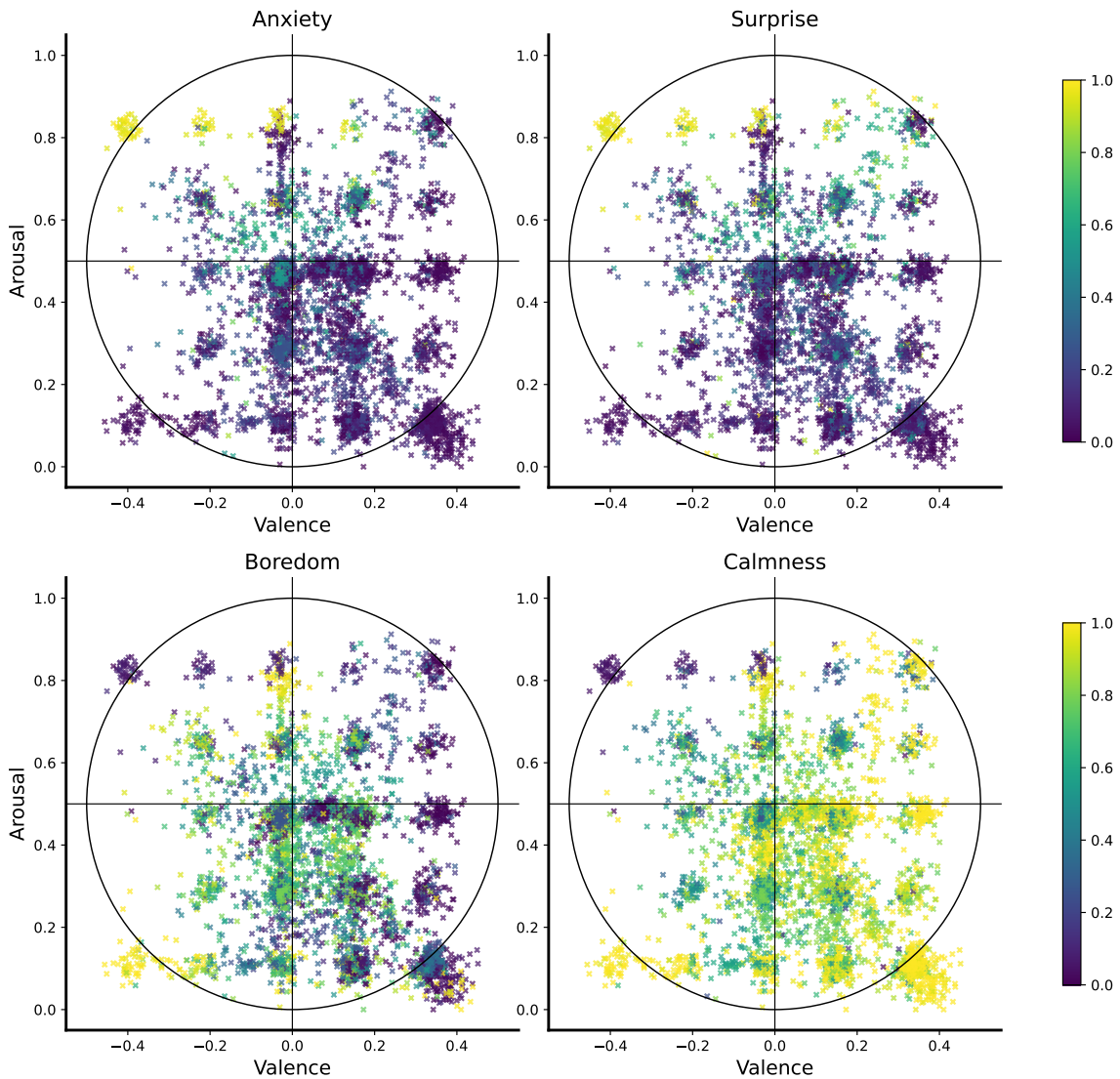
Figure 5.10: Reported arousal and valence rating highlighted with respect to perceived *surprise*, *anxiety*, *boredom*, and *calmness*

disused basic emotion location in section 1.2, the surprise was in quadrant I and anxiety in quadrant II close to the high arousal axis, boredom in quadrant III, and calmness in quadrant IV close to the low arousal axis. It is clear to see the reported emotions are dominant in those quadrants.

Although there are a few surprises in the second quadrant, many reports are greater than 0.8 in the first quadrant. It is important to note that these are subjective responses. Thus,

the evaluation of emotions may vary between people.

In summary, we can see the pattern that these emotions are in different quadrants. Thus, Russel's circumplex model provides a framework to estimate emotion from only arousal and valence.

5.4.7 Analysis of Emotions and Comfortability

In the previous section, we look at dominant emotions in the AV domain. Here we are going to look at the correlation between comfortability and emotions.

Table 5.14 shows the Pearson correlation matrix. In the first row, comfortability has a negative medium correlation with both surprise and anxiety, and this is expected since a person cannot be comfortable and at the same time feel anxiety. On the other hand, it has a high positive correlation with calmness and a weak correlation with boredom. Similarly, if a person is comfortable, he is expected they feel calm.

Table 5.14: Correlation Matrix between emotion and comfort values reported after trial from the participants

	comfortability	surprise	anxiety	boredom	calmness
comfortability	1.00	-0.56	-0.56	0.12	0.61
surprise	-0.56	1.00	0.55	-0.13	-0.41
anxiety	-0.56	0.55	1.00	-0.12	-0.53
boredom	0.12	-0.13	-0.12	1.00	0.21
calmness	0.61	-0.41	-0.53	0.21	1.00

This subject reporting supports our assumption that we use calmness, boredom, and surprise for comfortability estimation and anxiety, surprise, and boredom for uncomfortability estimation.

5.4.8 Familiarity with Robot vs Comfortability

As experience and familiarity are important in human-robot collaboration, we want to see if there is a correlation between familiarity with robots and comfortability. Before applying

any test, we need to check if these two metrics follow a normal distribution. Hence, we plot KDE plot in Fig. 5.11 for both metrics. As can be seen from the figure, both metrics do not follow a normal distribution. Therefore, instead of applying Pearson correlation, we applied Spearman, which does not make any assumption on the distribution.
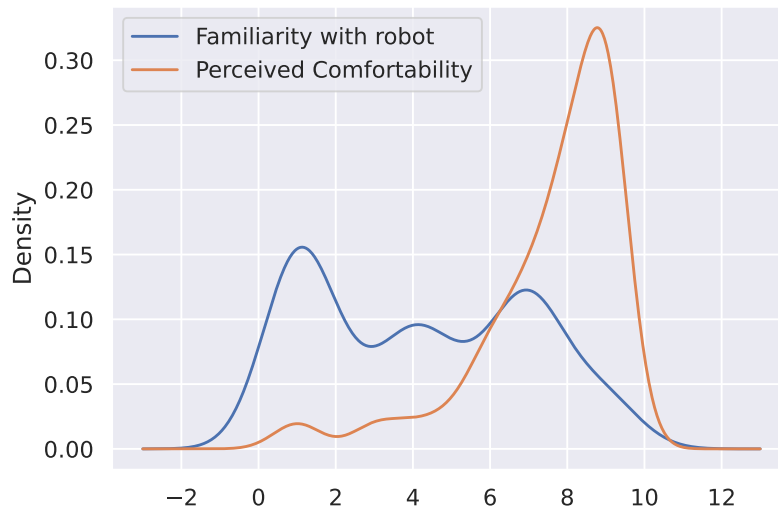
Figure 5.11: Distribution of both comfort and familiarity with the robot.

A Spearman's correlation was run to determine the relationship between comfortability and familiarity with robot values. There was a weak, positive monotonic correlation between comfortability and familiarity with robot (r_s= .22, n = 270, p-value < 0.00). This result indicates that people who have experience with robots tend to have a higher comfortability with the robot. This is expected as we get to experience with a system we become more comfortable.

Next, we will analyze how the performance of the participant change in learning through the experiment.

5.5 Objective Evaluation

The evaluation of experiments for all the configurations is done in terms of productivity and learning rate. The productivity is calculated based on how many parts were assembled divided by trial duration (in minutes) as shown in equation 5.13. For each trial, six minutes is given to the subject to complete a trial, and as soon as the time is reached, the robot stops, and the trial stops as well. The learning rate is calculated by the percentage change in the number of assembled parts between the first and last trial.

The productivity formula is defined as

$$productivity = \frac{num\_assembled\_parts}{trial_{duration}} \tag{5.13}$$

where *num_assembled_parts* is a total number of parts assembled by the subject during trials, and *trial_{duration}* is the trial duration in minutes.

5.5.1 Analysis of Learning Rate

The human brain has a unique way of learning, changing from person to person. As we practice a new concept or physical activity, we get better at it [13]. Researchers are working on the human brain to understand the learning rate and integrate this with a system that can adapt accordingly [108].

The experiment explained in 5.1 is a repetitive task where humans do the same task multiple times. Hence, we measure the productivity of the human by the number of parts assembled during the trial. The objective here is not to find a person learning rate; rather, we want to see how much performance increased after eight trials.

Fig. 5.12 shows the cumulative percentage change with respect to the first trial for all participants. The colored line shows the individual rate of change, and the dashed black line is the average change. It is clear that the number of assembled parts increased in general. However, there are some cases where the participant assembled fewer parts during
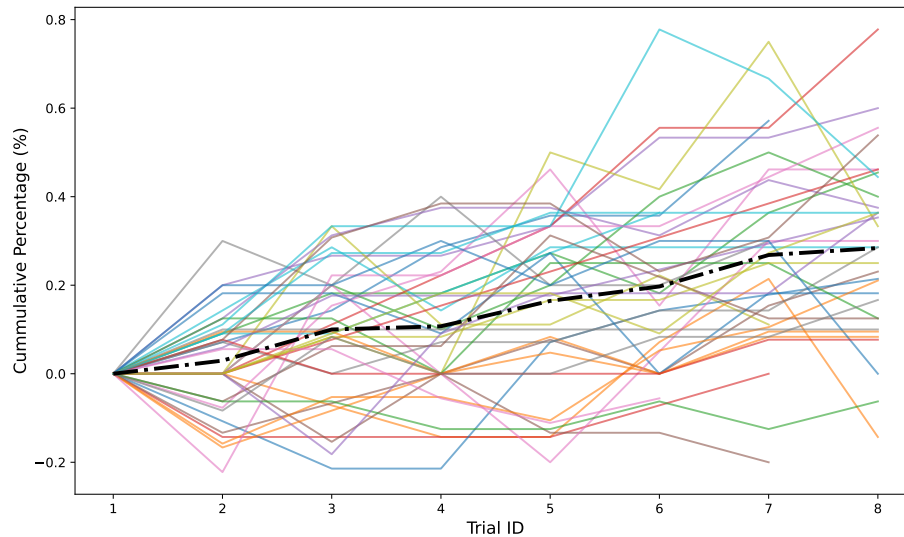
Figure 5.12: Cumulative percentage with respect to the first trial across all the participants.

the second trial. This decrease can be due to the random robot behavior selection between trials. For example, a subject may experience one of the VO algorithms in the first trial and then experience VR in the second trial. As we know, the VR algorithm is more aggressive than VO. In conclusion, the participant productivity (the number of assembled parts during the trial) increased by 26%.

5.5.2 Analysis of Productivity

In this section, we analyze the productivity of the participants based on the robot's algorithm of VO, VO_CIES, VR, and VR_CIES for the post-processed dataset. Although the trials during the experiment were randomized to reduce the order effect, we only selected the last two trials since we know there is a learning rate effect. On average, a trial is approximately 6 minutes. The average number of assembled parts for VO is 14.1 in comparison to 14.6 for VO with CIES, as shown in 5.15. Although this difference is small, it is large if extrapolated to a longer working period. On average, if a subject works with the robot for an hour, the subject will complete 141 parts working with VO and 146 when working with

VO with CIES active. Hence, productivity increased by 3.5%.

Similarly, the average number of assembled parts is 13.7 for the VR algorithm and 14.3 for VR with CIES from trials. If we run the experiment for an hour, the average completed parts would be 137 and 143 for VR and VR_CIES, respectively. We can conclude that the CIES improves VR slightly with a 4.3% improvement in productivity and an increase in comfortability.

Table 5.15: Summary of Robot's algorithm regarding the trial duration, number of assembled parts, and productivity for the last two trials for the Post-Processed dataset.

Algo	# Trials	Trial Duration (Sec.)		# Assembled Part		Productivity (Min.)	
		Mean	*STD*	*Mean*	*STD*	*Mean*	*STD*
VO	18	357.505	2.949	14.141	3.444	2.373	0.577
VO_UNCI	14	358.764	2.589	14.655	4.070	2.450	0.677
VR	15	357.880	3.246	13.736	3.369	2.302	0.561
VR_UNCI	14	357.879	3.538	14.311	3.645	2.398	0.605

The mean of the algorithms differ from each other, and it looks like CIES increases productivity; however, to conclude, we need to conduct a statistical test. Hence, we conducted a pairwise Tukey HSD test.

H_a^9 : The productivity of VO is high in comparison to the VR (μ_{VR}).

$$H9 = \left\{ H_0^9 : \mu_{VO}^{num\_parts} = \mu_{VR}^{num\_parts}, \quad \mathbf{H_a^9} : \mu_{VO}^{num\_parts} > \mu_{VR}^{num\_parts} \right\} \tag{5.14}$$

H_a^{10} : The productivity of VR_UNCI is as good as the VO (μ_{VO}).

$$H10 = \left\{ \mathbf{H_0^{10}} : \mu_{VO}^{num\_parts} = \mu_{VR\_UNCI}^{num\_parts}, \quad H_a^{10} : \mu_{VO}^{num\_parts} > \mu_{VR\_UNCI}^{num\_parts} \right\} \tag{5.15}$$

The statistics result in Table 5.16 shows that there is no statistical difference between VO and VR algorithms. The p-values are greater than the $\alpha = 0.05$ value; hence, we failed to reject the null hypothesis for $H9$.

Similarly to the $H9$, we reject the alternative hypothesis as the p-value is larger than the α value for H_a^{10}. However, in the $H10$, we want to see whether the CIES model improves the productivity of the VR algorithm. The result of the H9 hypothesis showed that we

could not say that VO is better than VR interns in productivity. Hence, we cannot say that VR_UNCI improves the VR algorithm since both algorithms' productivity may be as good as VO.

In conclusion, although there is a mean difference between VR and VO with CIES and without, the improvement of CIES is statistically not significant. Hence, with the existing setup, CIES does not increase productivity.

Table 5.16: Pairwise Tukey test for last two trials from the participants.

A	B	mean(A)	mean(B)	diff	se	p-value
VO	VO_UNCI	2.548	2.869	-0.321	0.231	0.509
VO	VR	2.548	2.484	0.064	0.226	0.992
VO	VR_UNCI	2.548	2.622	-0.073	0.231	0.989
VO_UNCI	VR	2.869	2.484	0.385	0.240	0.386
VO_UNCI	VR_UNCI	2.869	2.622	0.248	0.245	0.743
VR	VR_UNCI	2.484	2.622	-0.137	0.240	0.940

Fig. 5.13 shows the scatter plot between the number of assembled parts and the mean comfortability during trial responses. The red line shows the regression line that shows there is a positive correlation between these two metrics. The correlation is calculated using the "Spearman" correlation method, which is a non-parametric version of the "Person" correlation for repeated measurements. The test showed that there is a 0.29 positive correlation between these two metrics. This result indicates that the subject's productivity increases as they get comfortable. In other words, making the subject comfortable will increase the productivity of humans.

5.5.3 Comfortability vs Gender

In this section, we investigate if there is a statistical difference between male and female participants in terms of comfortability.

H_a^{10} : There is a difference between male and female subjects regarding comfortability.
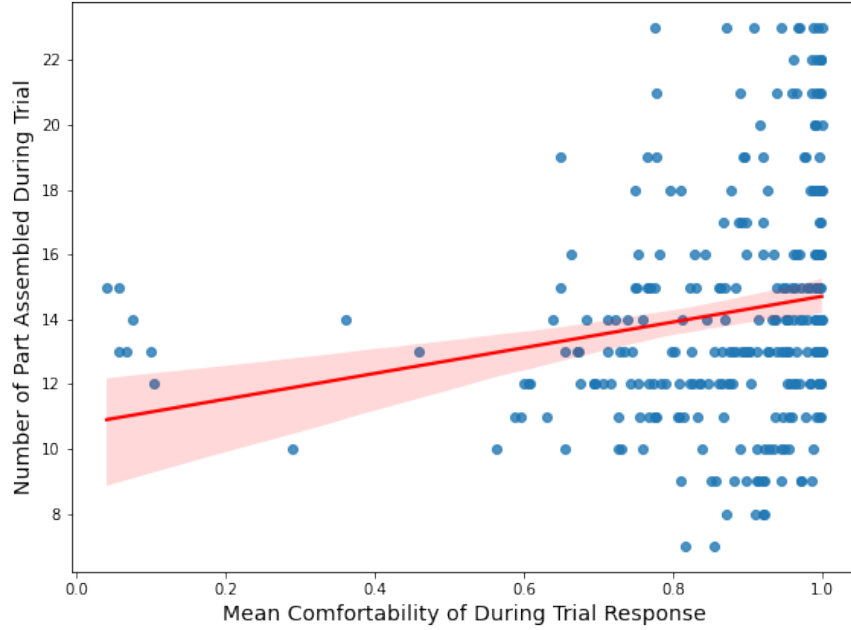
Figure 5.13: Productivity vs. Reported Comfortability

$$H11 = \left\{ H_0^{11} : \mu_{MALE}^{comfort} = \mu_{FEMALE}^{comfort}, \quad \mathbf{H_a^{11}} : \mu_{MALE}^{comfort} > \mu_{FEMALE}^{comfort} \right\} \quad (5.16)$$

Hence, we conducted an ANOVA test between two groups with the assumption of equal variance. The test result shows there is a difference between the two groups with $\alpha = 0.008$. Then a Tukey-HSD post-hoc test was applied to see how these two groups differ from each other. The $p - value = 0.009$ in Table 5.17 shows that there is statistical significance between two groups where male subjects feel more comfortable than female subjects. Thus, we reject the null hypothesis (H_0^{11}) for this analysis. This result is expected since the average familiarity robot for female participants is 2.69 and 5.07 for males. As we discussed in 5.4.8, familiarity positively affects perceived comfortability.

Table 5.17: Statistical significance between female and male participants

A	B	mean(Female)	mean(Male)	p-value
Female	**Male**	7.12	7.765	0.009

5.6 Stationary vs. Non-Stationary Experiments

In this section, we are comparing both stationary (Sawyer) and non-stationary (UR10) experiments. We will compare comfortability and uncomfortability for both experiments in terms of the axis and KDE approaches.

5.6.1 Analysis of Comfortability

For the comfortability axis estimation, surprise, boredom, and calmness emotions were used to estimate the comfortability axis. In order to find the comfortability axis, we used the responses that comfortability is greater than 0.8. Fig. 5.14 shows the estimated axis from both experiments. For the stationary experiment the comfortability axis is estimated to be $290.03°$ and for non-stationary is $278.65°$. Although there is a $12.62°$ slight difference, both axes overlap and are located in the same quadrant. This shift can be explained by the reported boredom level. As shown in Fig. 5.14, there are many responses that are reported as comfortable and high boredom levels at the same time. Hence, as we used boredom in the estimation of the axis, the boredom emotion pulls the comfortability axis towards the third quadrant.

Fig. 5.15 shows estimated a distribution obtained by KDE for both experiments. In the Fig. 5.15-b there is one main peak in the distribution, however, in the Fig. 5.15-c, there are two peaks. Similar to the comfortability axis estimation, this can be explained by the boredom level. It looks like some of the participants feel comfortable at the same time they felt bored. Hence, the distribution differs between the two experiments. Although there is a difference between KDE, both KDE locations are close to each other.
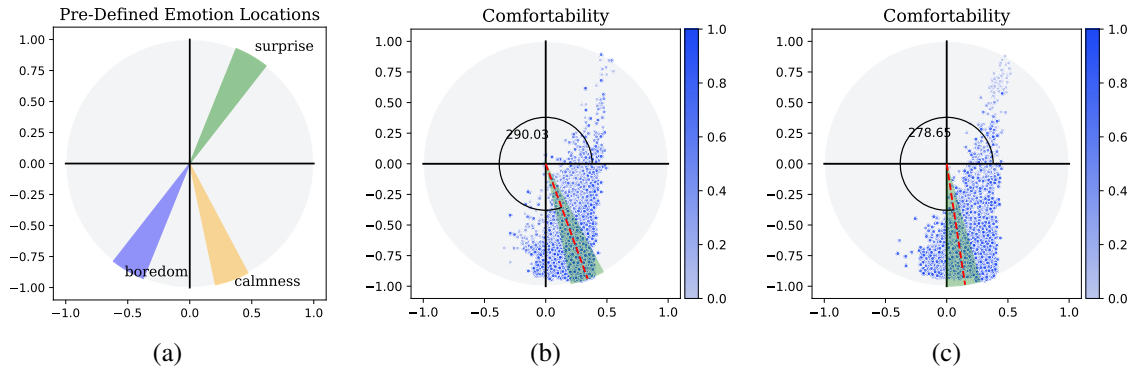
Figure 5.14: Estimated comfortability axis location for stationary and non-stationary experiments. (a) shows the emotion axis locations, (b) shows comfortability axis estimation from the stationary experiment, and (c) shows the comfortability axis from the non-stationary experiment.

Figure 5.15: Estimated comfortability distribution obtained by KDE for stationary and non-stationary experiments. (a) shows the emotion axis locations, (b) shows comfortability KDE estimation from the stationary experiment, and (c) shows the comfortability KDE from the non-stationary experiment.

5.6.2 Analysis of Uncomfortability

For the uncomfortability axis estimation, surprise, anxiety, and boredom emotions were used. In order to find the uncomfortability axis, we used the responses that comfortability is less than 0.4. Fig. 5.16 shows the estimated axis from both experiments. For the stationary experiment the uncomfortability axis is estimated to be 104.28° and for non-stationary

is 88.32°. There is a 15.96° difference between both axes. This difference can be explained by a number of uncomfortability responses. As shown in Fig. 5.16, there are many uncomfortable responses for the stationary experiment and limited AV points for the non-stationary experiment. Thus, we will need more AV points to estimate the uncomfortability axis on the circumplex model for the non-stationary experiment.
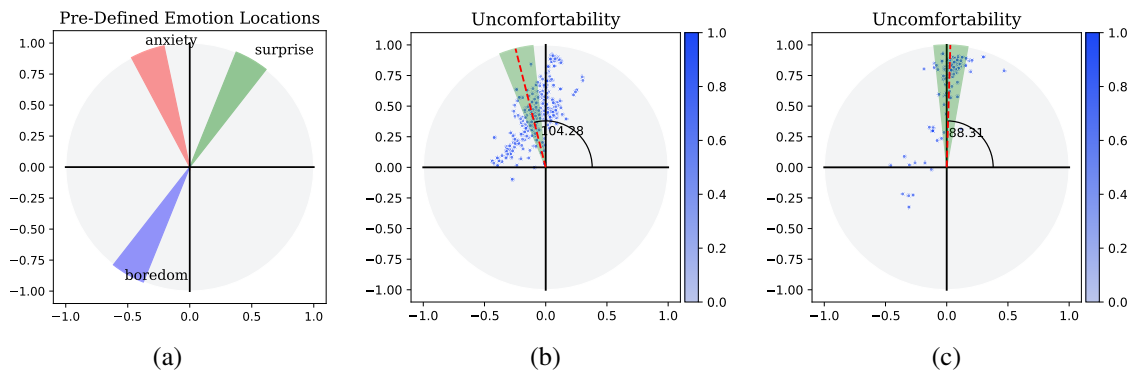
Figure 5.16: Estimated uncomfortability axis location for stationary and non-stationary experiments. (a) shows the emotion axis locations, (b) shows uncomfortability axis estimation from the stationary experiment, and (c) shows the uncomfortability axis from the non-stationary experiment.

Fig. 5.17 shows estimated distribution obtained by KDE for both experiments. In Fig. 5.17-b the distribution is large as there are many AV points for uncomfortability, however, in Fig. 5.15-c, the KDE is small due to a lack of AV points. Similar to the uncomfortability axis estimation, the distribution needs more AV points to estimate better distribution for uncomfortability.

5.7 Models Performance Evaluation

In this section, the performance of two CIES models (Random Forest) was evaluated. The first model was trained on a stationary experiment dataset (see chapter 4). The second
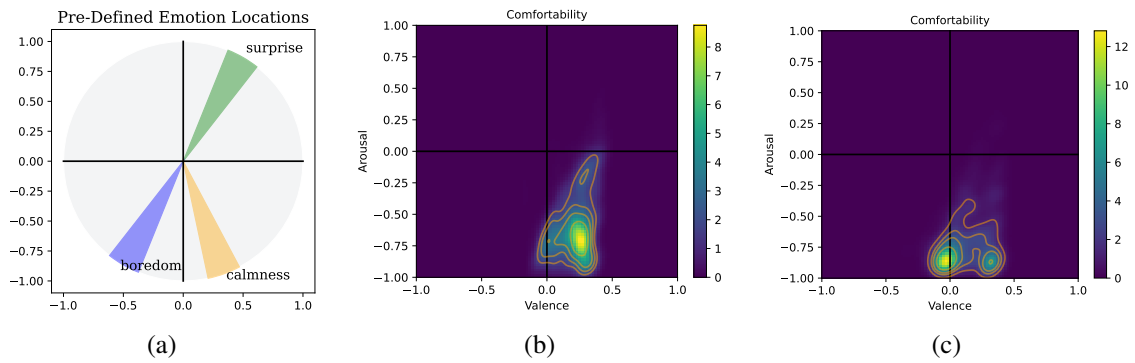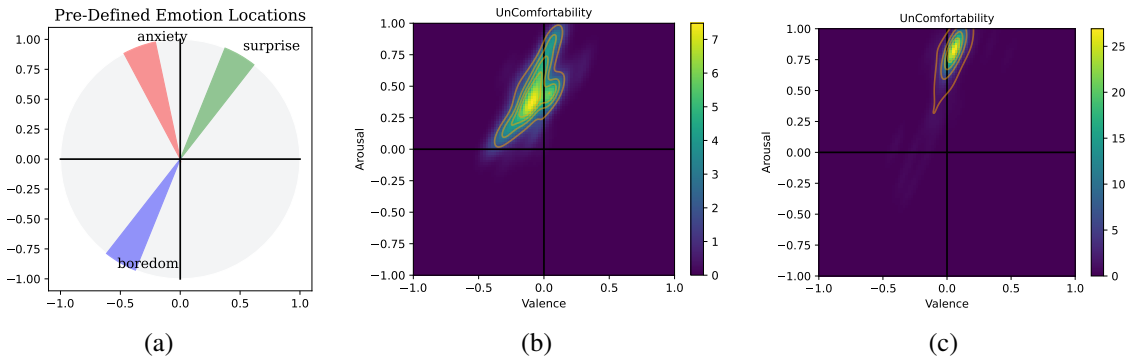
Figure 5.17: Estimated uncomfortability distribution obtained by KDE for stationary and non-stationary experiments. (a) shows the emotion axis locations, (b) shows uncomfortability KDE estimation from the stationary experiment, and (c) shows the uncomfortability KDE from the non-stationary experiment.

model was trained on randomly selected twenty subjects' data from a non-stationary experiment (see chapter 5). The remaining dataset from seventeen subjects from non-stationary was used as a test dataset to evaluate both models. The test dataset consists of 1961 samples.

Table 5.18 presents the performance of both the models for comfortability and uncomfortability. Since we used a non-stationary experiment dataset for the test. The CIES model that trained on the same experiment our perform for both comfortability and uncomfortability. However, the RMSE and MAE score of uncomfortability for the stationary experiment is really close to the other model.

Table 5.18: Performance of two CIES models that trained on different experiment

Experiment Type	Comfortability		Uncomfortability	
	RMSE	MAE	RMSE	MAE
Stationary Model (RF)	0.46	0.43	0.29	0.24
Non-Stationary Model (RF)	0.33	0.30	0.28	0.20

In conclusion, these results are expected since the comfortability and uncomfortability axes are close to each other for both experiments.

5.8 Circumplex Model using Arousal and Valence

During non-stationary experiment, *arousal* and *valence* rating were collected during the ongoing trial from participants as subjective responses. Here, instead of estimating *arousal* and *valence* from emotions, we used reported *arousal* and *valence* in circumplex model to estimate the location of comfortability and uncomfortability axes. Fig. 5.18 shows both axes locations on the circumplex model. The comfortability axis location is 298.50° and uncomfortability is 120.38°. When emotions were used to calculate these axes, they were 290.03°, and 104.28° for comfortability and uncomfortability respectively. Although the axes are slightly shifted, they are near to each other. This indicates that reported emotions and arousal and valence are correlated.
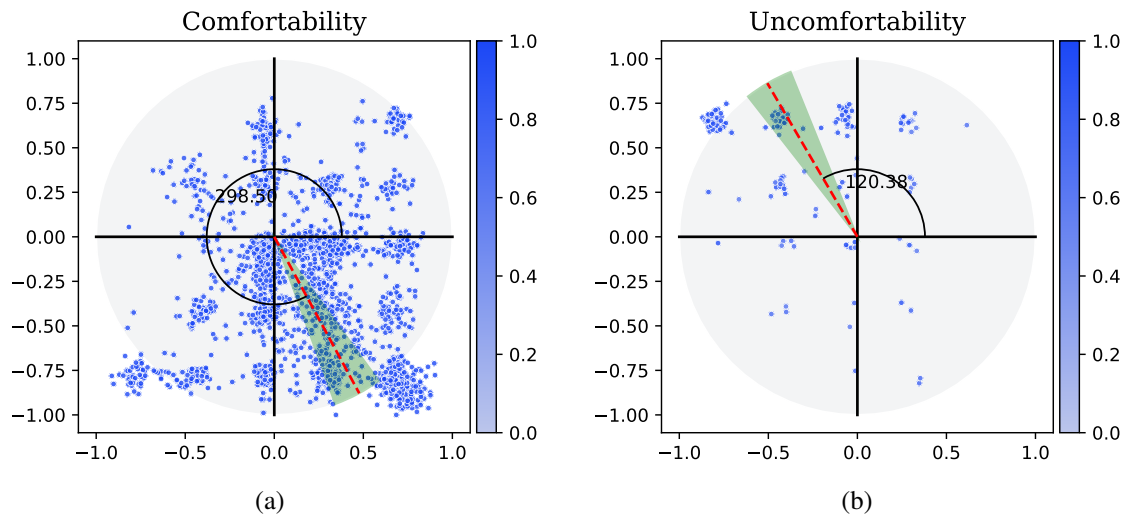
Figure 5.18: (a) shows circumplex model trained on reported arousal and valence rating similarly (b) shows circumplex model for uncomfortability trained on AV points

5.9 Limitations

While the results provided show great promise, it should be noted that the study was conducted in a lab setting, which is a rather controlled environment. Subjects were essentially sedentary and only used one hand for the most part. Because it is generally recognized that motion artifacts affect physiological signals, data gathering in non-stationary conditions can be difficult. Future studies should consider motion artifacts and create algorithms to eliminate them.

The proposed physiological computing system used multiple sensors for physiological signals. This is necessary for the system to work; however, it may be difficult to use such sensors sometimes. For example, the proposed system using pupillometry signals requires the participant to wear special glasses. However, if the participant already uses glasses, then it becomes challenging to collect the signals. However, as technology advances in the field of wearable sensors, this kind of challenge should be easier to deal with.

The proposed model was tested in a human subject experiment. The CIES models tested here were trained on an entirely different experiment and data collected from different people than this experiment. Hence, this is a major limitation for CIES models. However, the model trained on another dataset the results are still promising. In addition, the previous experiment was a stationary experiment where participants were sitting which reduces the motion artifacts. However, the current experiment was non-stationary, which reduce the quality of data in some cases.

The CIES model used in this experiment builds upon emotions that were used to estimate arousal and valence, then the circumplex model is used to calculate CI/UnCI. Estimation of arousal and valence from emotion has its own limitations since we are trying to come from the emotion domain to the AV domain. However, it would be better if we asked the participant to report arousal and valence during the trial. Then, arousal and valence can

be used to estimate CI/UnCI.

For machine learning algorithms, the data and its quality are the most important part of learning. The CIES model used in this experiment was trained on only twenty people's data. For this type of model, this data size is a limitation. Hence, having more data that covers a variety of uncomfortability occurrences would make a great improvement for the CIES model.

One of the challenges in ML algorithms is to generalize a model for multiple people. Although we generalize the model based on twenty people from previous experiments, it would be better to have an adaptive model that learns from human responses as it changes from person to person.

Another limitation of this experiment is that we change only the robot's velocity and cushioning distance between human and robot. However, changing robot trajectory based on human's uncomfortability would further this work.

5.10 Summary

In this chapter, we evaluate the CIES system with the DSS system of VO and VR. Our objective was to see if CIES has the potential to estimate uncomfortability by using this estimated comfortability index and whether the robot behavior can be modified to make the person more comfortable. Therefore, we designed an experiment that estimates human uncomfortability in real-time. During the DSS algorithm, the estimated value was not used; it was just recorded. However, during the DSS_CIES configuration, the estimated value was affecting the DSS algorithms where the directed velocity and cushioning distance were modified. This adaptation is explained in detail in section 5.1.2. Table 5.19 shows all the hypotheses from this chapter. The analyses were done on both original and post-processed datasets.

In the H1 and H2, we observed how the CIES model improves comfortability and if this

Table 5.19: Summary of all hypothesis proposed in chapter 5

Hypo.	Statement	Original Dataset (p-value)	Post-Processed Dataset (p-value)
H^1	The median perceived comfortability rating for VO_UNCI (μ_{VO_UNCI}) is greater than VO (μ_{VO}).	0.549	0.129
H^2	The median perceived comfortability rating for VR_UNCI (μ_{VR_UNCI}) is greater than VR (μ_{VR}).	0.105	0.289
H^{3*}	The median perceived comfortability rating for VR_UNCI (μ_{VR_UNCI}) is as good as VO (μ_{VO}).	0.124	0.078
H^4	The median percent time in NORMAL mode for DSS (μ_{DSS}) is greater than DSS_CIES (μ_{DSS_CIES}).	0.028	0.078
H^5	The median percent time in REDUCE mode for DSS_CIES (μ_{DSS_CIES}) is greater than DSS (μ_{DSS}).	0.050	0.016
H^6	The median percent time in STOP mode for DSS (μ_{DSS}) is greater than DSS_CIES (μ_{DSS_CIES}).	0.007	0.042
H^7	The VR algorithm would be less comfortable than VO algorithm.	0.369	0.549
H^{8*}	The VR_UNCI algorithm would be as good as VO.	0.373	0.318
H^9	The productivity of VO is high in comparison to the VR (μ_{VR})	-	0.992
H^{10*}	he productivity of VR_UNCI is as good as the VO (μ_{VO})	-	0.989
H^{11}	There is a difference between males and female subjects in terms of comfortability.	0.009	

* Unlike the rest of the hypotheses, this hypothesis supports the null hypothesis.
The value reported for H1, H2, and H3 from the trial responses.

comfortability is statistically significant. From the post-processed dataset, we can see the p-value of 0.129 and 0.289 for VO_UNCI and VR_UNCI, respectively. This is important since the p-value is not very large for these two analysis. Although we cannot say the CIES improves VO and VR algorithms in terms of comfortability statistically, it is clear that the CIES model is slightly improving both algorithms. In future research, it would be beneficial to tune the effect of the CIES model on robot behavior.

From hypothesis H3, we observed that the CIES model made the VR algorithm as good as VO in terms of comfortability. We also observed that the productivity of VO is not better than both VR and VR_UNCI. In addition, the results from the experiment show that CIES has the potential to estimate uncomfortability and the changes regarding the minimum distance between human and robot (see section 5.4.5). We can see that as the distance decreases, uncomfortability increases and vice versa.

In this experiment, the statistical analysis showed that there is no difference between DSS and DSS with the CIES model regarding estimated uncomfortability. This happened due to the adaptation of the CIES model to the DSS. The mean of mapped UnCI for all the trials was 0.05 (0.09). The contribution of this mapped UnCI value on DSS (see section 5.1.2) is divided by half since we multiply with 0.5. Hence, the overall contribution of the mapped UnCI on DSS was 0.025, which modified the directed speed and distance only by 2.5%. This change may not be noticeable to the participant as the change is minimal. However, in a few cases, the mapped UnCI was 0.64. as a result, the effect of UnCI goes to 0.32, which means that the directed velocity will be reduced by 32% and cushioning distance will increase by 32% at MAX. It should be noted that the number of this type of trial was limited to see the statistical differences between modes.

The CIES model used in this experiment was from a previous experiment in which the data was collected during an entirely different task. The upper and lower limit were extracted from the previous experiment. Hence, the generalized models we used in this experiment could not be generalized enough for all participants. An alternative solution would be to use the subject-dependent model on UnCI and have control as to how much the CIES would have an effect on DSS algorithms.

In conclusion, the CIES has the potential to estimate uncomfortability even though the dataset was collected from a different experiment. The subject dependence on generalization was insufficient to see the effect of the CIES with DSS. Having a personalized and

adaptive model that modifies DSS would be a future direction of this research. In addition, instead of modifying the velocity and distance between human and robot, changing the robot's trajectory may be more effective in making the human more comfortable.

Chapter 6

Conclusion and Future Work

In this chapter, we present the conclusions and summary of this research, discuss the impact of this research on other studies, and present the future direction of this research.

6.1   Conclusion

The research presented in this dissertation showed the implementation of human comfort index estimation during human-robot collaboration tasks. We proposed a novel approach for human comfortability and uncomfortability index estimation by using subjective responses and physiological signals. One of the proposed approaches was inspired by Russel's circumplex model, which allows emotion to be represented in terms of arousal and valence [2]. The second approach was the estimation of CI/UnCI from arousal and valence by fitting a distribution obtained by KDE. To estimate arousal and valence, emotions were estimated from physiological signals.

In the first experiment (see chapter 4), twenty subjects participated in the experiment, where physiological signals and subjective metrics were collected. The experiment was designed such that we collected the subjective metrics during the trial to have better signal labeling. Our results showed that estimating CI/UnCI from arousal and valence is a promising approach.

In the following experiment (see chapter 5) where uncomfortability estimation from the CIES model and robot behavior was controlled. The experiment focused on mimicking industrial settings. Hence, a conveyor and multiple robots were added to the experiment to work alongside the human subject. This helped the subject to think he/she is on the factory floor working with robots.

In the literature, it is known that the GSR signal is sensitive to emotional changes [33, 56]. However, it is suggested to use the GSR signal in conjunction with other signals such as ECG, EEG, or EMG [25]. In this thesis, three physiological signals: GSR, ECG, and Pupillometry were used. Our observation showed that the signals that provide information about comfortability or uncomfortability can be ordered (high to low) as GSR, ECG, and Pupillometry signals. We observed that removing the Pupillometry signal has little effect on the accuracy of the CIES model. We did not analyze the contribution of each signal independently since it is not the focus of this research.

There are two types of robots, industrial and collaborative. Industrial robots are in use in cages, and they work fast, precisely, and robustly. Hence, this type of robot is more productive and needs to be considered for mass production [82]. On the other hand, collaborative robots can work alongside humans. Thus, their productivity is less than industrial robots since they need to work at a lower speed. The objective of the collaborative robot is to maintain productivity while keeping humans safe. The CIES model aims to provide feedback to the collaborative robot to make humans more comfortable; as a result, productivity will increase. The experimental result shows that although the mean of the robot modes where CIES active had higher productivity, it is not statistically significant. This may happen due to multiple reasons; one of them is the generalized model used. The second reason is the effect of the CIES model on the robot behavior was small when it changed velocity and cushioning distance. The CIES model can improve if it has more power to modify robot behavior and use a personalized model instead of a generalized model.

We used the ML model trained on the Sawyer experiment and validated on a non-stationary experiment in which the UR-10 robot was used in collaboration. The estimated uncomfortability was used to modify the robot's velocity and cushioning distance between humans and robots. Hence, as the human becomes less comfortable, the robot will reduce its velocity and increase the distance. The robot will go to its default settings as the human

becomes comfortable. The subjective and objective analyses were completed. Although the CIES model estimated uncomfortability during the trial, the effect of the CIES model on robot behavior change is limited. Hence, the CIES model was unable to make participants comfortable in comparison to the DSS system.

In conclusion, physiological metrics can be used to estimate the human comfortability index. The proposed method allows the CI and UnCI to be estimated only by two parameters: *arousal* and *valence*. In the current work, we calculated arousal and valence based on emotions; however, a better approach would be to ask subjects about their *arousal* and *valence* levels as it was done in [104]. In future work, we will use reported *arousal* and *valence* level [26] to train a model to estimate these two metrics directly from physiological signals without using emotions.

6.2   Future Work

In the last experiment, we modified only the robot's velocity and cushioning distance. In addition to these two parameters, it would be useful to change the robot trajectory on the fly as the human becomes uncomfortable, as shown in figure 6.1. There are existing online trajectory planners available for the arm robot [109, 110]. The robot path planner can toggle between two modes, the first one is to minimize joint movement when a human is comfortable, and the second one is selecting a path that maximizes the distance between human and robot when a human is not comfortable.

In the final experiment (non-stationary) we used a generalized model (see 5.1.1). The generalized model helps ML models to work across multiple participants; however, to have a better estimation of comfortability or uncomfortability, a subject-dependent model is needed. The CIES model can start estimation from a generalized model; however, as the data is collected from a particular subject, the generalized model needs to be updated for the subject hence the performance of the CIES model can be improved. When we get
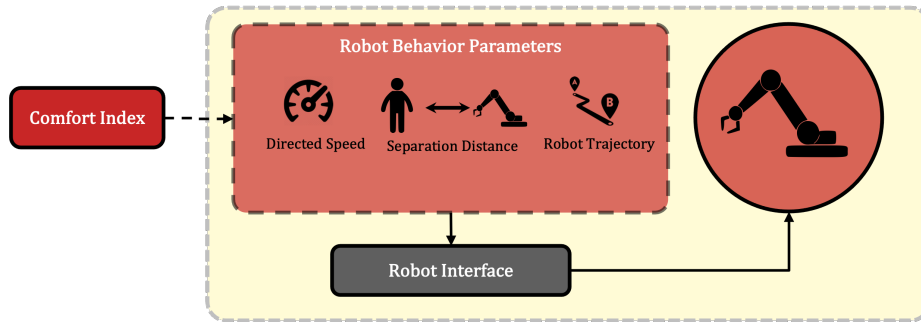
Figure 6.1: Robot's behavior that can be modified by comfort index

used to a task, it is human nature to get comfortable doing it. When we get comfortable, our physiological signals will go to normal since the sympathetic nervous system will be in control. Thus, the CIES model would estimate accordingly. The CIES model may fail only if the person has a different physiological response to the same stimulus, which is unlikely to happen.
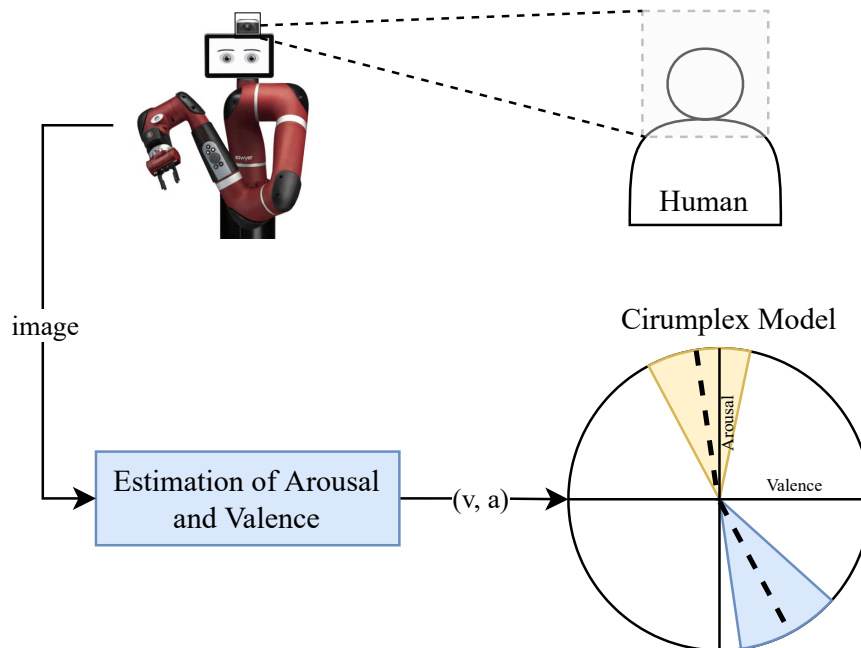
Figure 6.2: Arousal and valence estimation from image then the estimation of comfortability and uncomfortability.

In this research, we used emotions to estimate arousal and valence. Using reported arousal and valence would be a good direction for future research. In addition, we used three sensors (ECG, GSR, and Pupillometry) to estimate emotions. However, some new researchers use only a camera to estimate continuous arousal and valence, which looks promising [3]. Fig. 6.2 shows an overall diagram where a camera faces the human, and a system uses these images to estimate arousal and valence. Then this can be used in the circumplex model to estimate the comfortability and uncomfortability. Although this approach does not require any physiological signal, it requires the subject's face to be in the camera view. Thus, this type of system can be useful for a stationary experiment where the subject is sitting in front of the robot. A separate experiment is needed to test this system in the HRC settings. Thus, this approach for a human-robot collaboration task would need further research.

Robot
Mode: NORMAL
Velocity: 0.5 m/sec.

Comfort Level

Human Field of View

Figure 6.3: Augmented/Mixed reality sample view for CIES and robot's information.

The current CIES model estimates a human comfortability/uncomfortability level and provides this information to the robot. This information is valuable feedback for the robot to adjust its immediate behavior. However, the human does not know what information is being sent to the robot as feedback. This is a limitation of the CIES system. To tackle this, augmented/mixed reality can be used to display information that is sent to the robot. Thus, humans would know what is happening internally in the CIES system. In addition, displaying the robot's mode (NORMAL, REDUCE, and STOP), velocity, relative position,

and distance between human and robot would be helpful, as shown in Fig. 6.3. However, separate studies may need to be conducted to understand how the comfortability of a human agent may change when receiving information from the robot by the means of augmented reality devices in HRC.

Bibliography

[1] S. Kumar, C. Savur, and F. Sahin, "Survey of Human–Robot Collaboration in Industrial Settings: Awareness, Intelligence, and Compliance," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 51, no. 1, pp. 280–297, Jan. 2021.

[2] J. A. Russell, "A circumplex model of affect," *Journal of Personality and Social Psychology*, vol. 39, no. 6, pp. 1161–1178, 1980.

[3] A. Toisoul, J. Kossaifi, A. Bulat, G. Tzimiropoulos, and M. Pantic, "Estimation of continuous valence and arousal levels from faces in naturalistic conditions," *Nature Machine Intelligence*, vol. 3, no. 1, pp. 42–50, Jan. 2021.

[4] M. Sahin and C. Savur, "Evaluation of Human Perceived Safety during HRC Task using Multiple Data Collection Methods," in *2022 17th Annual Conference System of Systems Engineering, SoSE 2022*, 2022, p. 6.

[5] S. H. Fairclough, "Fundamentals of physiological computing," *Interacting with Computers*, vol. 21, no. 1-2, pp. 133–145, Jan. 2009.

[6] NSF, "Information and Intelligent Systems (IIS): Core Programs," https://www.nsf.gov/pubs/2018/nsf18570/nsf18570.htm, 2019.

[7] D. Kulic and E. A. Croft, "Anxiety detection during human-robot interaction," in *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2005, pp. 616–621.

[8] D. Kulić and E. Croft, "Physiological and subjective responses to articulated robot motion," *Robotica*, vol. 25, no. 1, pp. 13–27, Jan. 2007.

[9] L. Tiberio, A. Cesta, and M. Belardinelli, "Psychophysiological Methods to Evaluate User's Response in Human Robot Interaction: A Review and Feasibility Study," *Robotics*, vol. 2, no. 2, pp. 92–121, 2013.

[10] S. H. Fairclough, "Physiological Computing and Intelligent Adaptation," in *Emotions and Affect in Human Factors and Human-Computer Interaction*. Elsevier, 2017, no. 2017, pp. 539–556.

[11] C. Savur and F. Sahin, "Real-Time American Sign Language Recognition System Using Surface EMG Signal," in *2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA)*. Miami: IEEE, Dec. 2015, pp. 497–502.

[12] S. Kumar, C. Savur, and F. Sahin, "Dynamic Awareness of an Industrial Robotic Arm Using Time-of-Flight Laser-Ranging Sensors," *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 2850–2857, 2018.

[13] R. C. Wilson, A. Shenhav, M. Straccia, and J. D. Cohen, "The Eighty Five Percent Rule for optimal learning," *Nature Communications*, vol. 10, no. 1, pp. 1–9, 2019.

[14] E. Musk, "An integrated brain-machine interface platform with thousands of channels," *Journal of Medical Internet Research*, vol. 21, no. 10, pp. 1–14, 2019.

[15] J. R. Hughes, "Electroencephalography. Basic principles, clinical applications and related fields," *Electroencephalography and Clinical Neurophysiology*, vol. 54, no. 4, pp. 473–474, Oct. 1982.

[16] W. Klimesch, "Memory processes, brain oscillations and EEG synchronization," *International Journal of Psychophysiology*, vol. 24, no. 1-2, pp. 61–100, Nov. 1996.

[17] J. O'Keefe and N. Burgess, "Theta activity, virtual navigation and the human hippocampus," *Trends in Cognitive Sciences*, vol. 3, no. 11, pp. 403–406, Nov. 1999.

[18] B. Yılmaz, S. Korkmaz, D. B. Arslan, E. Güngör, and M. H. Asyalı, "Like/dislike analysis using EEG: Determination of most discriminative channels and frequencies," *Computer Methods and Programs in Biomedicine*, vol. 113, no. 2, pp. 705–713, Feb. 2014.

[19] Y. Zhang, Y. Chen, S. Bressler, and M. Ding, "Response preparation and inhibition: The role of the cortical sensorimotor beta rhythm," *Neuroscience*, vol. 156, no. 1, pp. 238–246, Sep. 2008.

[20] R. Alcaide, N. Agarwal, J. Candassamy, S. Cavanagh, M. Lim, B. Meschede-Krasa, J. McIntyre, M. V. Ruiz-Blondet, B. Siebert, D. Stanley, D. Valeriani, and A. Yousefi, "EEG-Based Focus Estimation Using Neurable's Enten Headphones and Analytics Platform," *bioRxiv*, p. 2021.06.21.448991, 2021.

[21] J. W. Hurst, "Naming of the Waves in the ECG, With a Brief Account of Their Genesis," *Circulation*, vol. 98, no. 18, pp. 1937–1942, Nov. 1998.

[22] M. Ali, F. Machot, A. Mosa, M. Jdeed, E. Machot, and K. Kyamakya, "A Globally Generalized Emotion Recognition System Involving Different Physiological Signals," *Sensors*, vol. 18, no. 6, p. 1905, Jun. 2018.

[23] K.-H. Choi, J. Kim, O. S. Kwon, M. J. Kim, Y. H. Ryu, and J.-E. Park, "Is heart rate variability (HRV) an adequate tool for evaluating human emotions? – A focus on the use of the International Affective Picture System (IAPS)," *Psychiatry Research*, vol. 251, no. February, pp. 192–196, May 2017.

[24] G. Lu, F. Yang, J. A. Taylor, and J. F. Stein, "A comparison of photoplethysmography and ECG recording to analyse heart rate variability in healthy subjects," *Journal of Medical Engineering and Technology*, vol. 33, no. 8, pp. 634–641, Nov. 2009.

[25] W. Boucsein, *Electrodermal Activity*. Boston, MA: Springer US, 1992.

[26] M. M. Bradley and P. J. Lang, "Measuring emotion: The self-assessment manikin and the semantic differential," *Journal of Behavior Therapy and Experimental Psychiatry*, vol. 25, no. 1, pp. 49–59, Mar. 1994.

[27] M. Yik, J. A. Russell, and J. H. Steiger, "A 12-Point Circumplex Structure of Core Affect," *Emotion*, vol. 11, no. 4, pp. 705–731, 2011.

[28] Tobii, "Tobii pro," https://www.tobiipro.com/.

[29] P. Bonifacci, L. Desideri, and C. Ottaviani, "Familiarity of Faces: Sense or Feeling?" *Journal of Psychophysiology*, vol. 29, no. 1, pp. 20–25, Jan. 2015.

[30] S. Mathôt, "Pupillometry: Psychology, Physiology, and Function," *Journal of Cognition*, vol. 1, no. 1, pp. 1–23, Feb. 2018.

[31] R. Paprocki and A. Lenskiy, "What Does Eye-Blink Rate Variability Dynamics Tell Us About Cognitive Performance?" *Frontiers in Human Neuroscience*, vol. 11, no. December, pp. 1–9, Dec. 2017.

[32] C. Savur and F. Sahin, "American Sign Language Recognition system by using surface EMG signal," in *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. Banff: IEEE, Oct. 2016, pp. 002 872–002 877.

[33] D. Kulic and E. A. Croft, "Affective State Estimation for Human–Robot Interaction," *IEEE Transactions on Robotics*, vol. 23, no. 5, pp. 991–1000, Oct. 2007.

[34] K. GOUIZI, F. BEREKSI REGUIG, and C. MAAOUI, "Emotion recognition from physiological signals," *Journal of Medical Engineering and Technology*, vol. 35, no. 6-7, pp. 300–307, Oct. 2011.

[35] A. al-Qerem, F. Kharbat, S. Nashwan, S. Ashraf, and k. blaou, "General model for best feature extraction of EEG using discrete wavelet transform wavelet family and differential evolution," *International Journal of Distributed Sensor Networks*, vol. 16, no. 3, 2020.

[36] P. Rani, N. Sarkar, and C. Liu, "Maintaining Optimal Challenge in Computer Games through Real-Time Physiological Feedback Mechanical Engineering," *Task-Specific Information Processing in Operational and Virtual Environments: Foundations of Augmented Cognition*, pp. 184–192, 2006.

[37] V. Villani, L. Sabattini, C. Secchi, and C. Fantuzzi, "A Framework for Affect-Based Natural Human-Robot Interaction," in *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, Aug. 2018, pp. 1038–1044.

[38] C. Dobbins, S. Fairclough, P. Lisboa, and F. F. G. Navarro, "A Lifelogging Platform Towards Detecting Negative Emotions in Everyday Life using Wearable Devices," in *2018 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*. IEEE, Mar. 2018, pp. 306–311.

[39] S. Kumar and F. Sahin, "A framework for a real time intelligent and interactive Brain Computer Interface," *Computers and Electrical Engineering*, vol. 43, pp. 193–214, Apr. 2015.

[40] J. S. Artal-Sevil, A. Acon, J. L. Montanes, and J. A. Dominguez, "Design of a Low-Cost Robotic Arm controlled by Surface EMG Sensors," in *2018 XIII Technologies Applied to Electronics Teaching Conference (TAEE)*. Canary Island: IEEE, Jun. 2018, pp. 1–8.

[41] Y. Mangukiya, B. Purohit, and K. George, "Electromyography(EMG) sensor controlled assistive orthotic robotic arm for forearm movement," in *2017 IEEE Sensors Applications Symposium (SAS)*. IEEE, 2017, pp. 1–4.

[42] L. Shu, J. Xie, M. Yang, Z. Li, Z. Li, D. Liao, X. Xu, and X. Yang, "A review of emotion recognition using physiological signals," *Sensors (Switzerland)*, vol. 18, no. 7, 2018.

[43] S. Zoghbi, D. Kulić, E. Croft, and M. Van Der Loos, "Evaluation of affective state estimations using an on-line reporting device during human-robot interactions," *2009*

IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS 2009, pp. 3742–3749, 2009.

[44] C. Bartneck, D. Kulić, E. Croft, and S. Zoghbi, "Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots," *International Journal of Social Robotics*, vol. 1, no. 1, pp. 71–81, 2009.

[45] NASA, "NASA Task Load Index," *NASA Website*, 2006.

[46] M. Joosse, A. Sardar, M. Lohse, and V. Evers, "BEHAVE-II: The Revised Set of Measures to Assess Users' Attitudinal and Behavioral Responses to a Social Robot," *International Journal of Social Robotics*, vol. 5, no. 3, pp. 379–388, Aug. 2013.

[47] T. Ninomiya, A. Fujita, D. Suzuki, and H. Umemuro, "Development of the Multi-dimensional Robot Attitude Scale: Constructs of People's Attitudes Towards Domestic Robots," in *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2015, vol. 9388 LNCS, pp. 482–491.

[48] T. Nomura, T. Suzuki, T. Kanda, and K. Kato, "Measurement of negative attitudes toward robots," *Interaction StudiesInteraction Studies. Social Behaviour and Communication in Biological and Artificial Systems*, vol. 7, no. 3, pp. 437–454, 2006.

[49] C. M. Carpinella, A. B. Wyman, M. A. Perez, and S. J. Stroessner, "The Robotic Social Attributes Scale (RoSAS)," in *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, vol. Part F1271. New York, NY, USA: ACM, Mar. 2017, pp. 254–262.

[50] C. D. Spielberger, "State-Trait Anger Expression Inventory–2."

[51] D. Kulic and E. A. Croft, "Real-time safety for human - robot interaction," in *ICAR '05. Proceedings., 12th International Conference on Advanced Robotics, 2005.*, vol. 2005. IEEE, 2005, pp. 719–724.

[52] C. T. Landi, V. Villani, F. Ferraguti, L. Sabattini, C. Secchi, and C. Fantuzzi, "Relieving operators' workload: Towards affective robotics in industrial scenarios," *Mechatronics*, vol. 54, no. April, pp. 144–154, 2018.

[53] P. Rani, N. Sarkar, and C. Smith, "Affect-sensitive human-robot cooperation - theory and experiments," in *2003 IEEE International Conference on Robotics and Automation (Cat. No.03CH37422)*. IEEE, 2003, pp. 2382–2387.

[54] C. Liu, P. Rani, and N. Sarkar, "Human-robot interaction using affective cues," *Proceedings - IEEE International Workshop on Robot and Human Interactive Communication*, pp. 285–290, 2006.

[55] W. L. Hu, K. Akash, N. Jain, and T. Reid, "Real-Time Sensing of Trust in Human-Machine Interactions," *IFAC-PapersOnLine*, vol. 49, no. 32, pp. 48–53, 2016.

[56] P. Rani, N. Sarkar, and J. Adams, "Anxiety-based affective communication for implicit human–machine interaction," *Advanced Engineering Informatics*, vol. 21, no. 3, pp. 323–334, Jul. 2007.

[57] S. Erebak and T. Turgut, "Caregivers' attitudes toward potential robot coworkers in elder care," *Cognition, Technology and Work*, vol. 21, no. 2, pp. 327–336, 2019.

[58] J. T. Butler and A. Agah, "Psychological effects of behavior patterns of a mobile personal robot," *Autonomous Robots*, vol. 10, no. 2, pp. 185–202, 2001.

[59] J. Abdur-Rahim, Y. Morales, P. Gupta, I. Umata, A. Watanabe, J. Even, T. Suyama, and S. Ishii, "Multi-sensor based state prediction for personal mobility vehicles," *PLoS ONE*, vol. 11, no. 10, pp. 1–30, 2016.

[60] C. Dobbins and S. Fairclough, "Detecting and Visualizing Context and Stress via a Fuzzy Rule-Based System during Commuter Driving," *2019 IEEE International Conference on Pervasive Computing and Communications Workshops, PerCom Workshops 2019*, pp. 499–504, 2019.

[61] P. W. Ferrez and J. D. R. Millán, "You are wrong! - Automatic detection of interaction errors from brain waves," in *IJCAI International Joint Conference on Artificial Intelligence*, 2005, pp. 1413–1418.

[62] S. K. Ehrlich and G. Cheng, "Human-agent co-adaptation using error-related potentials," *Journal of Neural Engineering*, vol. 15, no. 6, p. 066014, Dec. 2018.

[63] M. Val-Calvo, J. R. Álvarez-Sánchez, J. M. Ferrández-Vicente, A. Díaz-Morcillo, and E. Fernández-Jover, "Real-Time Multi-Modal Estimation of Dynamically Evoked Emotions Using EEG, Heart Rate and Galvanic Skin Response," *International Journal of Neural Systems*, vol. 30, no. 04, p. 2050013, Apr. 2020.

[64] E. Mower, D. J. Feil-Seifer, M. J. Matarić, and S. Narayanan, "Investigating implicit cues for user state estimation in human-robot interaction using physiological measurements," in *Proceedings - IEEE International Workshop on Robot and Human Interactive Communication*, 2007, pp. 1125–1130.

[65] D. Novak, B. Beyeler, X. Omlin, and R. Riener, "Workload estimation in physical human-robot interaction using physiological measurements," *Interacting with Computers*, vol. 27, no. 6, pp. 616–629, 2015.

[66] I. Iturrate, R. Chavarriaga, L. Montesano, J. Minguez, and J. D. R. Millán, "Teaching brain-machine interfaces as an alternative paradigm to neuroprosthetics control," *Scientific Reports*, vol. 5, pp. 1–11, 2015.

[67] S. K. Ehrlich and G. Cheng, "A Feasibility Study for Validating Robot Actions Using EEG-Based Error-Related Potentials," *International Journal of Social Robotics*, vol. 11, no. 2, pp. 271–283, Apr. 2019.

[68] A. F. Salazar-Gomez, J. Delpreto, S. Gil, F. H. Guenther, and D. Rus, "Correcting robot mistakes in real time using EEG signals," in *Proceedings - IEEE International Conference on Robotics and Automation*. IEEE, 2017, pp. 6570–6577.

[69] F. Dehais, E. A. Sisbot, R. Alami, and M. Causse, "Physiological and subjective evaluation of a human-robot object hand-over task," *Applied Ergonomics*, vol. 42, no. 6, pp. 785–791, 2011.

[70] V. Villani, F. Pini, F. Leali, and C. Secchi, "Survey on human–robot collaboration in industrial settings: Safety, intuitive interfaces and applications," *Mechatronics*, vol. 55, no. March, pp. 248–266, Nov. 2018.

[71] M. Nokata, K. Ikuta, and H. Ishii, "Safety-optimizing method of human-care robot design and control," in *Proceedings 2002 IEEE International Conference on Robotics and Automation (Cat. No.02CH37292)*, vol. 2. IEEE, 2002, pp. 1991–1996.

[72] D. Kulic and E. Croft, "Estimating Robot Induced Affective State using Hidden Markov Models," in *ROMAN 2006 - The 15th IEEE International Symposium on Robot and Human Interactive Communication*. IEEE, Sep. 2006, pp. 257–262.

[73] E. T. Hall, "Proxemics and Design," *Design and Environment*, 1971.

[74] P. Rani, C. Liu, N. Sarkar, and E. Vanman, "An empirical study of machine learning techniques for affect recognition in human–robot interaction," *Pattern Analysis and Applications*, vol. 9, no. 1, pp. 58–69, May 2006.

[75] J. Too, C. Tan, F. Duan, Y. Zhang, K. Watanabe, R. Kato, and T. Arai, "Human-Robot Collaboration in Cellular Manufacturing: Design and Development," *International Conference on Intelligent Robots and Systems*, pp. 29–34, 2009.

[76] T. Arai, R. Kato, and M. Fujita, "Assessment of operator stress induced by robot collaboration in assembly," *CIRP Annals - Manufacturing Technology*, vol. 59, no. 1, pp. 5–8, 2010.

[77] G. Schirner, D. Erdogmus, K. Chowdhury, and T. Padir, "The future of human-in-the-loop cyber-physical systems," *Computer*, vol. 46, no. 1, pp. 36–45, 2013.

[78] P. Rani and N. Sarkar, "Emotion-sensitive robots - a new paradigm for human-robot interaction," in *4th IEEE/RAS International Conference on Humanoid Robots, 2004.*, vol. 1.   IEEE, 2004, pp. 149–167.

[79] ISO 2016, "ISO/TS 15066:2016," https://tinyurl.com/ebfesr2j, 2016.

[80] P. Rybski, P. Anderson-Sprecher, D. Huber, C. Niessl, and R. Simmons, "Sensor fusion for human safety in industrial workcells," in *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*.   IEEE, Oct. 2012, pp. 3612–3619.

[81] P. A. Lasota, T. Fong, and J. A. Shah, "A Survey of Methods for Safe Human-Robot Interaction," *Foundations and Trends in Robotics*, vol. 5, no. 3, pp. 261–349, 2017.

[82] S. Kumar, "Dynamic Speed and Separation Monitoring with On-Robot Ranging Sensor Arrays for Human and Industrial Robot Collaboration," Ph.D. dissertation, Rochester Institute of Technology, 2020.

[83] R. M. Bowen, "Online Novelty Detection System: One-Class Classication of Systemic Operation," Ph.D. dissertation, Rochester Institute of Technology, 2015.

[84] P. Rani, J. Sims, R. Brackin, and N. Sarkar, "Online stress detection using psychophysiological signals for implicit human-robot cooperation," *Robotica*, vol. 20, no. 6, pp. 673–685, 2002.

[85] C. Karyotis, F. Doctor, R. Iqbal, A. James, and V. Chang, "A fuzzy modelling approach of emotion for affective computing systems," *IoTBD 2016 - Proceedings of the International Conference on Internet of Things and Big Data*, no. IoTBD, pp. 453–460, 2016.

[86] A. Mollahosseini, B. Hasani, and M. H. Mahoor, "AffectNet: A database for facial expression, valence, and arousal computing in the wild," *arXiv*, vol. 10, no. 1, pp. 18–31, 2017.

[87] N. Du, F. Zhou, E. M. Pulver, D. M. Tilbury, L. P. Robert, A. K. Pradhan, and X. J. Yang, "Examining the effects of emotional valence and arousal on takeover

performance in conditionally automated driving," *Transportation Research Part C: Emerging Technologies*, vol. 112, no. September 2019, pp. 78–87, 2020.

[88] F. Galvão, S. M. Alarcão, and M. J. Fonseca, "Predicting exact valence and arousal values from EEG," *Sensors*, vol. 21, no. 10, 2021.

[89] M. E. L. Redondo, A. Vignolo, R. Niewiadomski, F. Rea, and A. Sciutti, "Can Robots Elicit Different Comfortability Levels?" in *Social Robotics*, A. R. Wagner, D. Feil-Seifer, K. S. Haring, S. Rossi, T. Williams, H. He, and S. Sam Ge, Eds. Cham: Springer International Publishing, 2020, vol. 12483, pp. 664–675.

[90] Pedregosa, F. and Varoquaux, G. and Gramfort, A. and Michel, V., and Thirion, B. and Grisel, O. and Blondel, M. and Prettenhofer, P., and Weiss, R. and Dubourg, V. and Vanderplas, J. and Passos, A. and, and Cournapeau, D. and Brucher, M. and Perrot, M. and Duchesnay, E., "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[91] D. Makowski, T. Pham, Z. J. Lau, J. C. Brammer, F. Lespinasse, H. Pham, C. Schölzel, and S. H. A. Chen, "NeuroKit2: A Python toolbox for neurophysiological signal processing," *Behavior Research Methods*, vol. 53, no. 4, pp. 1689–1696, Aug. 2021.

[92] C. Carreiras, A. P. Alves, A. Louren\c{c}o, F. Canento, H. Silva, and A. Fred, "BioSPPy: Biosignal Processing in Python," https://github.com/PIA-Group/BioSPPy/.

[93] H. F. S. Gamboa, "Multi-modal behavioral biometrics based on HCI and electro-physiology," Ph.D. dissertation, UNIVERSIDADE TECNICA DE LISBOA, 2008.

[94] C. Savur, S. Kumar, S. Arora, T. Hazbar, and F. Sahin, "HRC-SoS: Human robot collaboration experimentation platform as system of systems," in *2019 14th Annual Conference System of Systems Engineering, SoSE 2019*. IEEE, 2019, pp. 206–211.

[95] T. Cichon and J. Rossmann, "Simulation-based user interfaces for digital twins: Pre-, in-, or post-operational analysis and exploration of virtual testbeds," *31st Annual European Simulation and Modelling Conference 2017, ESM 2017*, pp. 365–372, 2017.

[96] S. Kumar, S. Arora, and F. Sahin, "Speed and Separation Monitoring using On-Robot Time-of-Flight Laser-ranging Sensor Arrays," in *2019 IEEE 15th International Conference on Automation Science and Engineering (CASE)*. IEEE, Aug. 2019, pp. 1684–1691.

[97] M. Safeea and P. Neto, "Minimum distance calculation using laser scanner and IMUs for safe human-robot interaction," *Robotics and Computer-Integrated Manufacturing*, vol. 58, pp. 33–42, Aug. 2019.

[98] SCCN, "Lab Stream Layer (LSL)," 2018.

[99] A. Ethem, *Introduction to Machine Learning*. MIT, 2014.

[100] A. BAUER, D. WOLLHERR, and M. BUSS, "Human–Robot Collaboration: A Survey," *International Journal of Humanoid Robotics*, vol. 05, no. 01, pp. 47–66, 2008.

[101] E. Mejía-Mejía, J. M. May, R. Torres, and P. A. Kyriacou, "Pulse rate variability in cardiovascular health: A review on its applications and relationship with heart rate variability," *Physiological Measurement*, vol. 41, no. 7, p. 07TR01, Aug. 2020.

[102] M. Kassner, W. Patera, and A. Bulling, "Pupil: An Open Source Platform for Pervasive Eye Tracking and Mobile Gaze-based Interaction," in *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*. New York, NY, USA: ACM, Sep. 2014, pp. 1151–1160.

[103] C. Savur, S. Kumar, and F. Sahin, "A framework for monitoring human physiological response during human robot collaborative task," *Conference Proceedings - IEEE International Conference on Systems, Man and Cybernetics*, vol. 2019-Octob, pp. 385–390, 2019.

[104] S. Koelstra, C. Mühl, M. Soleymani, J. S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras, "DEAP: A database for emotion analysis; Using physiological signals," *IEEE Transactions on Affective Computing*, vol. 3, no. 1, pp. 18–31, 2012.

[105] C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. F. del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, and T. E. Oliphant, "Array programming with NumPy," *Nature*, vol. 585, no. 7825, pp. 357–362, Sep. 2020.

[106] V. Shuman, D. Sander, and K. R. Scherer, "Levels of Valence," *Frontiers in Psychology*, vol. 4, 2013.

[107] G. Hoffman, "Evaluating Fluency in Human–Robot Collaboration," *IEEE Transactions on Human-Machine Systems*, vol. 49, no. 3, pp. 209–218, Jun. 2019.

[108] X. Wu, T. Wang, C. Liu, T. Wu, J. Jiang, D. Zhou, and J. Zhou, "Functions of Learning Rate in Adaptive Reward Learning," *Frontiers in Human Neuroscience*, vol. 11, p. 592, Dec. 2017.

[109] H. Pham and Q.-C. Pham, "A New Approach to Time-Optimal Path Parameterization Based on Reachability Analysis," *IEEE Transactions on Robotics*, vol. 34, no. 3, pp. 645–659, Jun. 2018.

[110] I. A. Sucan, M. Moll, and L. E. Kavraki, "The Open Motion Planning Library," *IEEE Robotics & Automation Magazine*, vol. 19, no. 4, pp. 72–82, Dec. 2012.

Chapter A

Appendix

A.1 Devices

The devices used in this study are listed below:

- Pupil Capture/Player is the tool provided by pupil dilation. The pupil headset used to collect subject's pupil dilation, gaze location and world camera stream recorded.
  https://pupil-labs.com/products/core/

- Shimmer3 GSR is the device has capability to measure GSR and PPG signals. The reason this device selected that it has Bluetooth capability and access to raw data from device.
  https://www.shimmersensing.com/products/shimmer3-wireless-gsr-sensor

- BioHarness is a wireless chest strap that allows recording of an ECG signal. In addition to the ECG, the device provides respiration rate, heart-rate, RR intervals, acceleration (3-axes), and device information.
  https://tinyurl.com/yhx75k6c

- UR-10 is a cobot with 10 kg payload. It is a widely used robot that being used in collaboration tasks. UR-10 used in minimum speed and separation monitoring experiment.
  https://www.universal-robots.com/products/ur10-robot/

- Sawyer Robot is another co-bot that developed using ROS. the Sawyer was used in task planning and execution experiment.
  https://www.rethinkrobotics.com/sawyer

- OptiTrack is camera-based tracking system that gives millimetric precision in the workspace, it is suitable for robotic application, in this research, the optitrack used to track subject's location and helmet position.

  https://optitrack.com/

A.2 Software

The software used in this study are listed below, in addition, custom software developed as well.

- VREP Simulation was used to create the digital-twin of the environment.

  http://www.coppeliarobotics.com/

- LSL Stream was used to transfer data and save data.

  https://github.com/sccn/labstreaminglayer

- ZMQ is messaging protocol that used communication between nodes.

  https://zeromq.org/

- Qt Framework is used to develop tools.

  https://www.qt.io/

- Rosbag is used to record and play back the recorded bags.

  http://wiki.ros.org/rosbag

- PlotJuggler is used to plot the time-series data.

  https://github.com/facontidavide/PlotJuggler

- Neurokit for physiological signal processing.

  https://neurokit.readthedocs.io/