



HAL
open science

Electrophysiological Measures for Human–Robot Collaboration Quality Assessment

Mathias Rihet, Guillaume Sarthou, Aurélie Clodic, Raphaëlle Roy

► **To cite this version:**

Mathias Rihet, Guillaume Sarthou, Aurélie Clodic, Raphaëlle Roy. Electrophysiological Measures for Human–Robot Collaboration Quality Assessment. *Discovering the Frontiers of Human-Robot Interaction*, Springer Nature Switzerland, pp.363-380, 2024, 978-3-031-66655-1. 10.1007/978-3-031-66656-8_15 . hal-04883459

HAL Id: hal-04883459

<https://hal.science/hal-04883459v1>

Submitted on 13 Jan 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Electrophysiological Measures for Human-Robot Collaboration Quality Assessment

Mathias Rihet, Guillaume Sarthou, Aurélie Clodic and Raphaëlle N. Roy

Abstract Electrophysiological signals offer invaluable insights for human monitoring, providing real-time, objective measures of mental states. Among the sensors used for signal acquisition, electrocardiography (ECG) and electrodermal activity (EDA) sensors stand out for their affordability and wearability. Electroencephalography (EEG) is also gaining popularity, particularly with the advent of dry electrode devices that reduce setup times, albeit at the expense of signal quality. These signals are increasingly employed to close the loop and develop adaptive systems, a discipline known as physiological computing [1]. For example, cerebral activity can be used as an active brain-computer interface to control a robotic arm [2]. Yet, the field of electrophysiological computing extends beyond mere control. It facilitates the monitoring of human mental states during tasks, which is of significant interest in Human-Robot Collaboration [3]. A robot equipped to monitor human mental states could dynamically adjust its behavior to uphold an optimal quality of interaction, addressing various aspects such as performance, engagement, satisfaction, or security. Moreover, it could learn from the impact of each action on the user's mental state, enabling more informed decision-making. Mental state monitoring has already been applied to several kinds of Human-Robot Interaction (HRI), including teleoperation and error correction [4]. Nevertheless, integrating physiological computing with human-robot collaboration presents unique challenges. Thus, the aim of this article is to review advanced applications of physiological computing in the domain of human-robot collaboration and offer guidance to researchers interested in delving into this interdisciplinary domain. It will develop the advantages of mental state monitoring in human-robot collaboration, examine cutting-edge studies employing such setups, as well as offer practical guidelines based on these findings.

M. Rihet · Raphaëlle N. Roy
Fédération ENAC ISAE-SUPAERO ONERA, Université de Toulouse, Toulouse, France
e-mail: mathias.rihet@isae-supaero.fr

A. Clodic · Guillaume Sarthou
LAAS-CNRS, Université de Toulouse, CNRS, Toulouse, France

Key words: HRI, HRC, physiology, neuroergonomics, cognitive state, mental state monitoring, BCI, guidelines, cognitive effort, stress, surprise

Electrophysiological Measures for Human-Robot Collaboration Quality Assessment	1
Mathias Rihet, Guillaume Sarthou, Aurélie Clodic and Raphaëlle N. Roy	
1 Assessing the quality of human-robot collaboration	3
1.1 Methodological Principles	3
1.2 Measurement Techniques	4
1.3 Mental States of Interest	5
2 Electrophysiological activities	6
2.1 Brain activity	6
2.2 Ocular activity	8
2.3 Cardiac activity	8
2.4 Electrodermal activity	9
3 Mental State Monitoring	10
3.1 Signal preprocessing	10
3.2 Offline and online analyses	11
3.3 Relevant biomarkers	11
4 Guidelines for HRC quality assessment	13
4.1 Considerations for experimental equipment	13
4.2 Considerations for experimental design	14
4.3 Considerations for HRC adaptation	15
5 Conclusion	16
References	16
References	16

1 Assessing the quality of human-robot collaboration

1.1 Methodological Principles

In the future, robots hold the promise of offering assistance in numerous aspects of our daily lives and fulfilling key societal functions. Their potential as social agents has already been tested in various ecological settings where they undertake roles as diverse as waiters in restaurants [5], shopkeepers selling hats [6] or guides in a shopping mall [7]. These examples underscore the complexity of human-robot interaction (HRI) in real-world scenarios and the diversity of subtasks that robots must perform to collaborate smoothly with humans and gain acceptance within society.

As the field of HRI evolves, it necessitates a deep consideration of human factors in robot design and evaluation. Replicable methodological approaches are expected not only to ensure robustness but also to motivate active debates within the research community [8, 9]. Ideally, user studies aiming to characterize the everyday use of robots by average users would require inclusivity across a broad spectrum of participants and multiple sessions to mitigate biases associated with user-specific preferences and session effects. However, such ecological studies require significant

resources. Thus, they are often preceded by more focused investigations targeting specific aspects of the interaction in laboratory settings. Yet, even within these narrower scopes, explicit formulation of the research questions and limitations of the study remains necessary to guarantee transparency and robustness of the research process.

In the context of human-robot collaboration (HRC), additional constraints come into play. According to the joint action paradigm [10], effective collaboration rely on various coordination mechanisms such as joint attention, intentional action understanding, shared task representation. Robot can take advantage of such mechanisms to manage an interaction, not to mimic but to be understood and to understand humans [11]. While performance metrics remain pertinent, fluency emerges as a key factor in optimizing HRC interactions [12]. This concept defined as a well-synchronized meshing of agents actions allow to couple precise and efficient action timing with appropriate and dynamic adaptations of both agents without need for verbal communication.

Thus, collaboration between humans and robots is, at least, as much about world representation than task solving. Addressing the disparities between human expectations and robot actions is essential for achieving effective collaboration. Hence, understanding variations in human mental state during human-robot collaboration and learning how to adapt accordingly seems crucial [13, 14].

1.2 Measurement Techniques

Among the methods commonly used by the HRC community [15], subjective metrics like questionnaires or interviews are frequently employed for offline analysis of human mental states. These measures are often acquired before and after a given task, or during breaks within the task at hand, hence generating task interruption. Moreover, these methods lack the capability for dynamic adaptation during interaction, limiting the assessment of the quality of the interaction (QoI) to session level. Yet, QoI should also be evaluated regarding each task and each action to offer deeper insights into the interaction dynamics [16, 17]. In addition, conducting this evaluation solely from the robot's perspective, without external assistance, would ensure its applicability across a wider range of scenarios.

To address these limitations, the addition of objective metrics is of great interest. Behavioral metrics such as facial recognition or idle time enable online computation but offer limited insights into human mental states. Hence, we perceive physiological measures as a promising complementary approach.

Advancements in neuroscience and the availability of portable physiological monitoring devices have broadened the applications of physiological measures [2]. For instance, the emerging field of neuroergonomics [18, 19] allows researchers to assess the mental states of human operators engaged in continuous interaction with complex systems through the use of neuroscientific paradigms, and more recently with the adjunction of machine learning models, through physiological computing

techniques [1]. By providing comprehensive insights, these monitoring methods facilitate the system's adaptation, leading to smoother and more effective interactions and enhancing the overall user experience (see Figure 1).

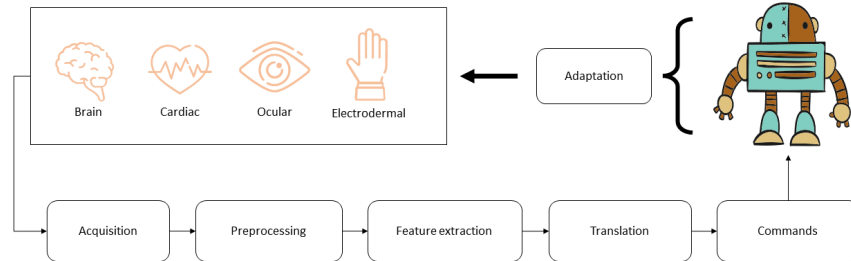


Fig. 1 Illustration of a physiological loop between human and robot.

1.3 Mental States of Interest

Ever since the steep rise of neuroscience and neuroergonomics research, several mental states of interest, both cognitive and affective, have been singled out as being of major importance for enhancing user experience, task performance and safety. A few of them have been studied for HRC.

One critical mental state of interest is cognitive effort, also referred to as mental engagement or mental workload [20]. It is related to the tendency of human performance to decline within a task particularly challenging or, in contrast, overly simple, following a U-shape curve [21]. Detecting fluctuations in cognitive effort is essential for identifying instances of mind wandering, where individuals disengage from the interaction, as well as periods of cognitive overload, characterized by a lack of fluency in the interaction or the need for a cognitive break [22]. By monitoring cognitive effort levels, robots can dynamically adjust their behavior to maintain an optimal level of engagement and support users in completing tasks effectively [14, 23].

Moreover, stress and arousal levels offer insights into human safety when sharing a workspace with robots. This can cover a wide range of parameters related to the prevention of any psychological or physical discomfort in HRC [24]. For instance, several factors such as increased movement speed, dominant movements, unpredictability, or inadaptability of the robot can decrease trust perceptions or increase anxiety [25]. A better comprehension of these factors can improve robotic design and facilitate robot integration into industry 5.0 workspaces.

Additionally, surprise detection plays a crucial role in HRC scenarios, as in other dynamic and unpredictable human-machine interactions [26]. Unexpected events, which can be qualified as novel or erroneous depending on the context [27], may

disrupt the fluency of collaboration and challenge users' beliefs about the robot's capabilities or intentions. By detecting instances of surprise, systems can adapt their behavior to address discrepancies and align with human beliefs during the ongoing process [28]. This may involve providing explanations for unexpected events or adjusting future actions to align with users' revised expectations. Surprise can also be used as a human implicit feedback to train robots to select the most suitable behaviour through reinforcement learning techniques [29].

2 Electrophysiological activities

In order to navigate more easily between the described signals, sensors and features the Table 1 summarize the sensors and their characteristics while Figure 2 provide a structured glossary of the various abbreviations.

Acronym	Recorded Activity	Localisation	Event Time Scale	Sensors
EEG	Brain	Head	millisecond	4 to 128
EOG	Ocular	Face	millisecond	2
ECG	Cardiac	Torso	second	2 to 5
PPG	Cardiac	Finger, Ear, Wrist	second	1
EDA	Electrodermal	Finger, Palm	second	2

Table 1 Summary of discussed electrophysiological sensors.

2.1 Brain activity

2.1.1 Sensors

Among the available electrophysiological signals, brain activity stands as the richest, which also makes it the more complex to process. While techniques like functional magnetic resonance imaging (fMRI) offer superior spatial resolution and magnetoencephalography (MEG) provide both spatial and temporal precision, their immobility and price constraints limit their utility in ecological settings. In contrast, electroencephalography (EEG) presents a promising solution due to its high temporal resolution and portability, making it well-suited for scenarios where participants need mobility [30].

However, the selection between wet and dry electrodes requires careful considerations, balancing ease of use with signal quality. Wet electrodes, though offering higher signal fidelity, require time-consuming setup procedures, while dry electrodes, despite their easier setup, often come with a compromise in signal quality.

In addition, the number of electrodes plays a crucial role, with more electrodes enabling the use of more extensive preprocessing techniques, yet potentially increasing setup time for wet EEG and participant discomfort for both [31].

2.1.2 Features

Brain activity can be processed within two domains : the time domain and the frequency domain.

In the time domain, event related potentials (ERP) offer insights into neural responses to specific events. ERPs are characteristic peaks or deflections in the EEG signal that occur in response to specific events or stimuli. These components, such as the P300 or N200, reflect neural processes related to perception, attention and resource allocation [32]. Key features extracted from ERPs include peak amplitude, latency, duration, and waveform morphology. While the overall ERP shape remains characteristic, these features can be impacted by human current mental state. For example, the amplitude of the P300 component, elicited by a rare stimuli, can vary according to task difficulty [33].

In the frequency domain, brain activity is analyzed by examining oscillatory patterns across different frequency bands. Common frequency bands of interest include delta (0.5 – 4Hz), theta (4 – 8 Hz), alpha (8 – 12 Hz), beta (12 – 30 Hz), and gamma (> 30 Hz). These frequency bands have been associated with distinct cognitive processes and mental states. For example, delta oscillations have been linked to attention processes, theta oscillations to memory encoding and inhibition, while alpha oscillations have been associated with states of relaxation, inhibition, or sensory attenuation. Beta oscillations seem involved in motor control and sensorimotor integration, while gamma oscillations have been associated with higher cognitive functions such as information processing, conscious perception and memory retrieval [34]. The most common feature extracted from frequency domain analysis is bandpower. Depending on the context, it can be used absolute, relative or as a ratio, such as the engagement ratio $\frac{\beta}{\alpha+\theta}$ [35].

It should be noted that such temporal and frequency features can be extracted either directly at the electrode level, or at a later stage at a region of interest (ROI) level (e.g. focusing on three zones including a frontal area, a central area and a parieto-occipital area). Moreover, more complex features such as event-related spectral perturbations (ERSPs), connectivity metrics [36] and covariance matrices [37] can also be extracted to allow for a richer and/or more accurate mental state assessment.

2.2 Ocular activity

2.2.1 Sensor

While eye-tracking allows for a complete analysis of participant gaze behavior, electrooculography (EOG) is a lighter technique which can still extract pertinent features related to mental states and does not require calibration [38]. EOG involves a pair of electrodes placed vertically, one above and one behind the eye. Since EOG recordings directly capture eye movements, including blinks, they often contribute to artifact rejection when analyzing EEG signals and are thus collectable with the same frequency.

2.2.2 Features

Blink can be extracted using thresholds, peak detection, filtering or more complex methods [39]. Once detected, the most common feature exploited is blink rate. However, it necessitates careful manipulation of confounding factors as blink rate can be impacted by the degree of visual processing required [40]. Additional analyses are usually performed using blink parameters including blink duration, blink amplitude, opening and closing velocity, as well as the percentage of closure, in particular to assess vigilance states [41].

2.3 Cardiac activity

2.3.1 Sensors

Cardiac activity is commonly monitored using electrocardiography (ECG) due to its reliability, while photoplethysmography (PPG) is chosen for its wearability. Among other wearable heart monitoring techniques like phonocardiography or seismocardiography, PPG is favored due to its reduced susceptibility to motion noise [42].

ECG involves placing at least two electrodes on the skin across the chest to measure the potential difference between them, capturing the electrical activity of the heart. A characteristic waveform of the ECG signal includes two deflections (Q, S) and a peak (R), collectively known as the QRS complex, which represents ventricular depolarization. The R-peak of this complex is commonly used to evaluate the heartbeat period by calculating the RR-intervals [43].

PPG monitors blood volume changes in the microvascular tissues through an optical device placed on areas such as the finger, ear or wrist. Although easier to wear, this location makes the sensor more prone to motion artifacts than ECG. In the signal, peaks related to pulse waves created by ventricular systol activity can be observed. Pulse-to-pulse intervals (PP-intervals) represents the time between consecutive pulse wave rises and are used to monitor heartbeat [42].

There is a delay, known as pulse transit time (PTT), between the QRS complex in ECG and the onset of the pulse wave in PPG, leading to theoretical differences between PP-intervals and RR-intervals. Although this delay is typically minimal in healthy individuals at rest, comparing these two sensors in more ecological settings could help characterize their respective limitations [44].

2.3.2 Features

Both RR-interval and PP-interval allow to derive various metrics [45]. The most common in the time domain is the heart rate mean value (HR) computed as $60/\overline{RR}$.

Heart rate variability (HRV), another usual metrics extracted for mental state monitoring purposes, can be computed in several ways depending on how much overall fluctuations (SDNN) or short-term fluctuations (RMSSD, pNN50) need to be integrated [46].

This signal can also be investigated in the frequency domain by separating very low frequency (VLF, 0–0.04 Hz), low frequency (LF, 0.04–0.15 Hz), and high frequency (HF, 0.15–0.4 Hz). Depending on the context, it can be used absolute, relative or as a LF/HF ratio.

2.4 Electrodermal activity

2.4.1 Sensors

Electrodermal activity (EDA) is typically measured using two electrodes placed on the palm of the hand or the fingertips. It can also be referred to as Galvanic skin conductance (GSC), skin conductance (SC), or galvanic skin resistance (GSR), with conductance being the inverse of resistance. The EDA signal comprises two main components: phasic and tonic [47].

The phasic component, also known as skin conductance response (SCR), reflects rapid changes in EDA which can be related or not to a specific stimulus depending on the experimental design. The tonic component refers to the baseline level of EDA, excluding SCRs, and is known as skin conductance level (SCL).

2.4.2 Features

Several features can be extracted from SCL and SCR, depending on the experimental design [48]. If the experimental setup includes specific stimuli, focusing on event-related skin conductance responses (ER-SCRs) can provide features such as peak amplitude, latency, and duration.

In experiments focusing on monitoring changes over time, the mean SCL over a given time-window is a more suitable metric. Yet, the number of SCR peaks

(typically reported in SCRs/minute) can still provide insights into the overall arousal in this context.

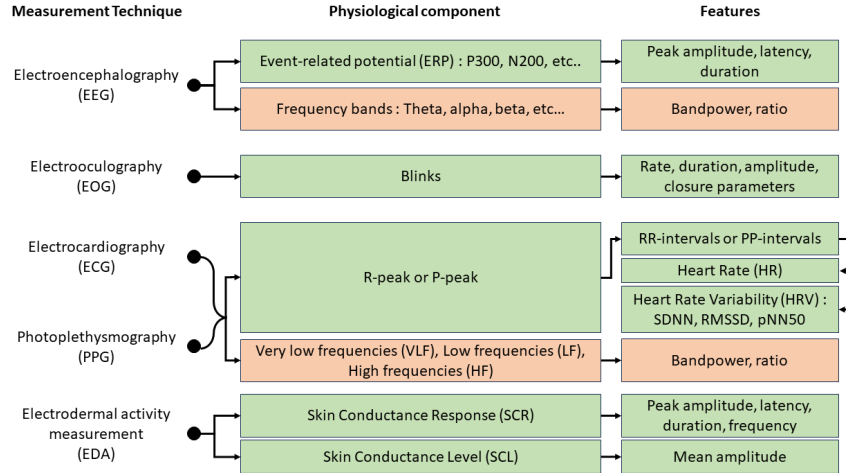


Fig. 2 Comprehensive glossary of discussed electrophysiological features and sensors.

3 Mental State Monitoring

3.1 Signal preprocessing

Physiological signals are prone to contamination by various sources of noise, and often require a preprocessing step before any data analysis, especially in dynamic settings. Each one of the signals described in the previous section can be impacted by line noise and motion-related artifacts. In the specific case of human-robot collaboration, recent work showed that the robot itself can be a source of artifact [49].

Basic filtering with Notch, Butterworth or Chebyshev filters can be sufficient to remove electromagnetic noise from the sensors. More complex techniques can also be used, however they require more data and are thus mainly used for brain signal preprocessing. Among these techniques we can cite independent component analysis (ICA) which decomposes the signal in independent components with the underlying assumption that some of these components will isolate noise sources [50]. Another commonly used method in recent studies is artifact subspace reconstruction (ASR) [51]. These filtering methods are based on blind source separation which is akin to spatial filtering.

Spatial filtering techniques exploit spatial information inherent in brain activity for enhanced signal quality and artifact mitigation. They can include electrode rereferencing, laplacian filtering and common spatial pattern analysis (CSP) [36].

It should be noted that the available preprocessing techniques may also vary depending on the type of analysis (offline or online) as discussed further below.

3.2 Offline and online analyses

Physiological signal analyses can be performed offline or online depending on the goal of the study and/or product development [52]. Offline analyses allow the use of more powerful techniques because they are applied in a finite space. However, if the end-goal is an online QoI evaluation from the robot's perspective, using pseudo-online pipeline for testing may be more suitable than using offline techniques which will not guarantee online performances. Hence, pseudo-online pipelines will load data in a buffer of realistic duration before applying any preprocessing method and use techniques adapted to this time constraint [51]. Yet, offline techniques are still relevant if the goal is rather to study thoroughly mental state physiological characteristics.

3.3 Relevant biomarkers

Depending on the mental states of interest, several biomarkers can be used. Ideally, these biomarkers are derived from prior research and hold recognition within the field.

While the mental states discussed here align with the authors' backgrounds, the terminology used to articulate these mental states may lack universality. Indeed, different researchers may employ diverse terminologies or conceptual frameworks to characterize similar mental phenomena, further complicating efforts to achieve a standardized classification.

Consequently, studies were categorized based on their experimental designs rather than on the mental states explicitly reported. This strategy aims to gather related studies irrespective of the specific terminology used by researchers, thereby facilitating a more comprehensive understanding of the underlying phenomena in HRI.

3.3.1 Cognitive effort

This section gather experiments where human and robot collaborate on tasks of varying difficulty levels. Depending on the study, the monitored mental state can be reported as workload [53–55], cognitive load [56], stress [57], or even fatigue [58,59].

In these designs, mental state monitoring has been achieved using a variety of physiological measures. For instance, EEG frequency features such as alpha and low gamma bandpower from frontal channels have been utilized [55]. Cardiac features, including RR intervals from chest straps, wristbands, or smartwatches [58, 59], as well as the LF/HF ratio from chest straps [53], have also been employed. Furthermore, EDA features like the SCR mean peak amplitude have been processed [57]. In addition, some studies have taken a more data-driven approach, extracting an extensive range of EEG features to input into machine learning models [54, 56].

However, while most of these studies focus on detecting the mental state of interest, only a few implement [56, 59] or propose [58] explicit adaptations. These adaptations rely on robot pace [56] and task allocation [58, 59].

3.3.2 Stress

This section compiles experiments where human and robot shared a workspace without explicitly performing a task together. Stress is the most reported mental state [53, 60–62] but it can also be labeled as affective responses [63]. EDA features are the most common, using both SCR peak amplitude [53, 60–63] and mean SCL level [53]. Yet, cardiac features such as HR have also been collected via ECG [63].

While most of these studies propose ideal robot speed, trajectory and distance to human based on the observed values, one of them discusses the opportunity for online adaptation, providing user-specific robot-behaviour [63].

3.3.3 Surprise

This section gathers experiments where physiological response is used to detect human reaction to an unexpected action from the robot. Across all these experiments, the primary aim was to detect error-related potentials (ErrPs) through EEG measurements. This particular ERP is characterized by two distinctive components: an error-related negativity observed approximately 200 milliseconds following the stimuli (referred to as N200 or ERN), and a subsequent positive peak occurring around 300 milliseconds post-event (known as P300). These components are involved in different neural processing steps of error recognition and suggest together a stimuli perceived as erroneous.

Given the necessity to utilize the overall shape of ErrPs for detection, most experiments employed all channel values at each time point within a specified time window as the feature of interest [4, 29, 64]. However, two studies employed data reduction techniques: a principal component analysis (PCA) [4] and the XDawn spatial filter [29], to decrease the number of channels. Another study opted for covariance matrices among a specific subset of channels, namely those located at frontal sites [65].

These features were then given to machine learning models to detect error occurrences. These outputs were then used to train reinforcement learning algorithms [29] or correct a simulated erroneous behavior of the robot [4, 65].

4 Guidelines for HRC quality assessment

4.1 Considerations for experimental equipment

Selecting appropriate experimental equipment necessitates a careful balance between factors such as wearability and robustness to artifacts. Portable and compact devices like dry EEG systems, bioharnesses, and wristbands offer convenience to participants by eliminating the need for adhesive electrodes and enabling quick setup. However, the information provided is comparatively limited when contrasted with medical-grade recording devices such as 64-channel wet EEG or five-electrode ECG systems [31].

The research context significantly influences sensor choice, emphasizing the importance of prior understanding of the mental states to be monitored and its related features. Strong physiological responses facilitate the use of more wearable sensors and ecological settings. Conversely, deeply investigating mental states with current sensor capabilities often necessitates controlled laboratory environments and offline data processing methods. Both approaches hold relevance in the field, as studying mental states with a high resolution can allow to better identify robust biomarkers. Thus, advancements in mental state monitoring require a seamless transition between laboratory-based and real-world settings as proposed within the neuroergonomic field [66].

Irrespective of experimental settings, the importance of familiarizing oneself with new devices cannot be overstated. Sensors, in particular, may technically "work" while producing signals with poor signal-to-noise ratios. Therefore, it is imperative to develop the ability to differentiate between clean and noisy signals prior to any further analysis.

Therefore, testing new devices or a novel combination of existing ones with a non-robotic task can be of great interest. Established paradigms such as the N-back task [67] for cognitive effort or Go/noGo [68] tasks and antissacades [69] for ErrPs can reliably elicit the expected biomarkers. They present valuable opportunities for researchers to assess sensor configuration, data synchronization, and feature extraction simultaneously before starting the HRC user study. Although such pre-studies may lengthen the overall experimental design process, their significance lies in preventing the wastage of both participant and researcher time by avoiding the collection of irrelevant data. From a methodological perspective, they ensure that both processing pipeline and sensor setup have the capacity to extract the expected biomarkers in perfect settings. Thus, when switching to HRC, any variation in the data will be due to the robotic task.

In contrast, a well-established setup should not prevent from prior testing with the robot, as the task itself can introduce new constraints such as synchronization issues [70] or additional artifacts, generated for instance by the robot itself [49].

4.2 Considerations for experimental design

Developing experiments to evaluate the quality of HRC necessitates a careful protocol design. It is imperative to address, or at least discuss, the limitations imposed by well-known biases, such as the diversity of participant expertise with robots and the session effect [8].

Even for a proof of concept, researchers must anticipate the need for adaptation and ensure that the studied scenario closely aligned with the intended use case. For instance, employing a double-task design is only advisable if the use case genuinely involves dual-tasking (or multi-tasking), as in some HRC studies for human-robot teaming. In such scenarios, humans perform several tasks prior to robotic intervention and some tasks are reallocated to the robot in order to reduce human burden [3]. However, if the objective is to examine the mental states induced by the robot or the collaborative task, it is advisable to refrain from using a double-task design. Although double-task designs are popular for studying cognitive effort due to their ability to seamlessly integrate validated neuroscientific or psychological tasks into any HRC scenario, they lack generalizability. Ultimately, they tend to turn the inductive task into the main task and make the HRC peripheral, failing to provide insights into mental states specific to the interaction. Moreover, they do not facilitate any adaptation of robot behavior to influence cognitive effort, since this effort is not instigated by the robot itself.

Presently, there exists a misalignment between current techniques in robotic and physiological computing which lead to compromise when trying to merge these two fields. On the one hand, robots continue to face challenges in performing tasks that humans execute effortlessly, particularly in shared workspaces where safety concerns add to the technical limitations. Furthermore, a certain level of interdependence between both agents is essential to underscore the necessity for collaboration in HRC scenarios. Taking these factors into account, an HRC trial lasting less than a few minutes proves challenging.

In contrast, physiological computing demands substantial data to address physiological variability, which poses limitations on mental state monitoring at multiple levels [71]. Intra-session variability makes repeated trials on a single individual a necessity to draw robust conclusions about the impact of studied conditions on the target mental state. Inter-individual variability leads to user-specific analyses, avoiding to tackle intra-session variability by group scale analysis. Additionally, it prohibits the use of global models for accurate predictions across users. Inter-session variability further restricts user-specific studies to a single session, hindering the long-term viability of such models. Fortunately, transfer learning techniques are emerging to

address these challenges robustly, offering potential solutions for both inter-session and inter-individual variability [72].

In summary, achieving robust mental state monitoring for HRC involves conducting tens or even hundreds of few minutes trials with current techniques, resulting in recording sessions of several hours. In this context, the comfort of participants must be carefully considered, both for ethical reasons and to mitigate potential biases in the results. Discomfort with sensors can introduce noise into the signal, while task repetitiveness can impair participant performance or introduce confounding mental states such as fatigue. Thus, striking a balance between the quality of the HRC scenario, the robustness of physiological analysis, and participant comfort is essential. This balance is contingent on the study's objectives, and the highlighted limitations should be duly acknowledged.

4.3 Considerations for HRC adaptation

Due to the inherent challenges arising from the interaction between physiological computing and HRC, there are few examples of online robotic adaptation based on mental state monitoring. Nevertheless, given that physiological computing essentially involves modeling human behavior, it holds promise for adaptation at three distinct levels: physical, cognitive, and social levels [13].

At the physical level, robots can modulate the speed and the trajectory during navigation or manipulation tasks. Such adaptations may address safety concerns by modulating biomarkers associated with stress and trust. Additionally, they can serve as cognitive effort regulators, especially when the speed of the robot directly influences the pace of the human worker [56].

At the cognitive level, robots can adapt their planning strategies based on the current mental state of humans. For instance, they can dynamically reallocate tasks based on workload or fatigue levels, as seen in mixed-initiative systems [3]. They can also rectify a plan through surprise detection. In addition, leveraging reinforcement learning techniques enables robots to evolve their decision-making processes over time, attaining a metacognitive level of adaptation. By establishing a feedback loop between humans and robots, physiological computing empowers robots not only to perceive the present mental state of humans but also to gauge the impact of their actions on this mental state, thereby enhancing decision-making capabilities in the long term.

At the social level, robots can adjust their communication strategies based on human mental state. This entails employing various social signals, including non-verbal cues, explicit or redundant information transmission, and even personality adjustments. For example, considering that human preferred interaction styles can vary depending on the task [73], it is plausible to assume that they may also fluctuate based on certain mental states.

Ultimately, it should not be minimized that such a framework raises significant social and ethical concerns. While these aspects are crucial, they fall beyond the

scope of this article. Readers seeking insights into these areas are encouraged to explore studies addressing the issue of personal data in HRI [74] or discussing the impact of physiological data collection in commercial devices [75].

5 Conclusion

In conclusion, this article has explored the application of electrophysiological measures in the evaluation of HRC. Through detailed feature extraction techniques and comprehensive examples from the literature, the potential of physiological computing techniques to evaluate and improve the quality of interaction between humans and robots in collaborative settings has been highlighted.

By processing electrophysiological signals such as EEG, EOG, ECG and EDA researchers can gain valuable insights into human mental states. Yet, the examination of available measurement techniques and biomarkers has underscored how understanding mental states beforehand is crucial for choosing appropriate sensors and settings.

While studying cognitive effort, stress and surprise holds significant promise for the future of HRC, their exploration demands careful consideration, as outlined in the proposed practical guidelines. Avoiding user studies' biases, addressing constraints from both physiological computing and HRC, and designing a suitable adaptation require expertise in various domains.

Moving forward, continued advancements in physiological computing and robotics hold great promise for enhancing human-robot collaboration across various domains, including healthcare, manufacturing, and assistive technology. By integrating physiological monitoring with adaptive robotic systems, we can create more intuitive and efficient robotic platforms that prioritize user experience and collaboration quality.

References

1. S. H. Fairclough, "Fundamentals of physiological computing," *Interacting with computers*, vol. 21, no. 1-2, pp. 133–145, 2009.
2. J.-H. Jeong, K.-H. Shim, D.-J. Kim, and S.-W. Lee, "Brain-controlled robotic arm system based on multi-directional cnn-bilstm network using eeg signals," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 28, no. 5, pp. 1226–1238, 2020.
3. R. N. Roy, N. Drougard, T. Gateau, F. Dehais, and C. P. Chanel, "How can physiological computing benefit human-robot interaction?" *Robotics*, vol. 9, no. 4, p. 100, 2020.
4. C. Lopes-Dias, A. I. Sburlea, and G. R. Müller-Putz, "Online asynchronous decoding of error-related potentials during the continuous control of a robot," *Scientific reports*, vol. 9, no. 1, pp. 1–9, 2019.
5. N. Mishraa, D. Goyal, and A. D. Sharma, "Issues in existing robotic service in restaurants and hotels," in *Proceedings of 3rd International Conference on Internet of Things and Connected Technologies (ICIoTCT)*, 2018, pp. 26–27.
6. S. Edirisinghe, S. Satake, D. Brscic, Y. Liu, and T. Kanda, "Field trial of an autonomous shopworker robot that aims to provide friendly encouragement and exert social pressure," in

- Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*, 2024, pp. 194–202.
7. A. Mayima, G. Sarthou, G. Buisan, P.-T. Singamaneni, Y. Sallami, J. Waldhart, K. Belhassein, A. Clodic, and R. Alami, “How to make a robot guide?” in *18th International Symposium on Experimental Robotics (ISER 2023)*, Chiang Mai, Thailand, Nov. 2023. [Online]. Available: <https://hal.science/hal-04243068>
 8. K. Belhassein, G. Buisan, A. Clodic, and R. Alami, “Towards methodological principles for user studies in human-robot interaction,” in *Test Methods and Metrics for Effective HRI in Collaborative Human-Robot Teams Workshop, ACM/IEEE International Conference on Human-Robot Interaction*, 2019.
 9. C. Jost, B. Le Pévédic, T. Belpaeme, C. Bethel, D. Chrysostomou, N. Crook, M. Grandgeorge, and N. Mirnig, *Human-Robot Interaction: Evaluation methods and their standardization*. Cham, Switzerland: Springer Nature, May 2020.
 10. G. Knoblich, S. Butterfill, and N. Sebanz, “Psychological Research on Joint Action,” in *Psychology of Learning and Motivation*. Elsevier, 2011, vol. 54, pp. 59–101. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/B9780123855275000036>
 11. A. Clodic, E. Pacherie, R. Alami, and R. Chatila, “Key elements for human-robot joint action,” *Sociality and normativity for robots: philosophical inquiries into human-robot interactions*, pp. 159–177, 2017.
 12. G. Hoffman, “Evaluating fluency in human–robot collaboration,” *IEEE Transactions on Human-Machine Systems*, vol. 49, no. 3, pp. 209–218, 2019.
 13. S. Rossi, F. Ferland, and A. Tapus, “User profiling and behavioral adaptation for hri: A survey,” *Pattern Recognition Letters*, vol. 99, pp. 3–12, 2017.
 14. E. Coronado, T. Kiyokawa, G. A. G. Ricardez, I. G. Ramirez-Alpizar, G. Venture, and N. Yamanohe, “Evaluating quality in human-robot interaction: A systematic search and classification of performance and human-centered factors, measures and metrics towards an industry 5.0,” *Journal of Manufacturing Systems*, vol. 63, pp. 392–410, 2022.
 15. C. L. Bethel and R. R. Murphy, “Review of human studies methods in hri and recommendations,” *Int. J. Soc. Robot.*, vol. 2, no. 4, pp. 347–359, 2010.
 16. A. Mayima, A. Clodic, and R. Alami, “Towards robots able to measure in real-time the quality of interaction in hri contexts,” *International Journal of Social Robotics*, vol. 14, no. 3, pp. 713–731, 2022.
 17. S. Bensch, A. Jevtic, and T. Hellström, “On interaction quality in human-robot interaction,” in *ICAART 2017 Proceedings of the 9th International Conference on Agents and Artificial Intelligence, vol. 1*. SciTePress, 2017, pp. 182–189.
 18. F. Dehais and H. Ayaz, “Progress and direction in neuroergonomics,” in *Neuroergonomics*. Elsevier, 2019, pp. 3–7.
 19. R. Parasuraman and M. Rizzo, *Neuroergonomics: The brain at work*. Oxford University Press, 2006.
 20. B. Cain, “A review of the mental workload literature,” *DTIC Document*, 2007.
 21. M. Mendl, “Performing under pressure: stress and cognitive function,” *Applied animal behaviour science*, vol. 65, no. 3, pp. 221–244, 1999.
 22. F. Dehais, A. Lafont, R. Roy, and S. Fairclough, “A neuroergonomics approach to mental workload, engagement and human performance,” *Frontiers in neuroscience*, vol. 14, p. 519228, 2020.
 23. C. Carissoli, L. Negri, M. Bassi, F. A. Storm, and A. Delle Fave, “Mental workload and human-robot interaction in collaborative tasks: A scoping review,” *International Journal of Human-Computer Interaction*, pp. 1–20, 2023.
 24. P. A. Lasota, T. Fong, J. A. Shah *et al.*, “A survey of methods for safe human-robot interaction,” *Foundations and Trends® in Robotics*, vol. 5, no. 4, pp. 261–349, 2017.
 25. S. Hopko, J. Wang, and R. Mehta, “Human factors considerations and metrics in shared space human-robot collaboration: A systematic review,” *Frontiers in Robotics and AI*, vol. 9, p. 799522, 2022.
 26. N. B. Sarter and D. D. Woods, “How in the world did we ever get into that mode? mode error and awareness in supervisory control,” *Human factors*, vol. 37, no. 1, pp. 5–19, 1995.

27. J. R. Wessel, C. Danielmeier, J. B. Morton, and M. Ullsperger, "Surprise and error: common neuronal architecture for the processing of errors and novelty," *Journal of Neuroscience*, vol. 32, no. 22, pp. 7528–7537, 2012.
28. R. Chavarriaga, A. Sobolewski, and J. d. R. Millán, "Errare machinale est: the use of error-related potentials in brain-machine interfaces," *Frontiers in neuroscience*, vol. 8, p. 86996, 2014.
29. S. K. Kim, E. A. Kirchner, A. Stefes, and F. Kirchner, "Intrinsic interactive reinforcement learning—using error-related potentials for real world human-robot interaction," *Scientific reports*, vol. 7, no. 1, pp. 1–16, 2017.
30. E. Jungnickel, L. Gehrke, M. Klug, and K. Gramann, "Mobi—mobile brain/body imaging," *Neuroergonomics*, pp. 59–63, 2019.
31. R. Hari and A. Puce, *Meg-EEG Primer*. Oxford University Press, 2023.
32. S. Fu and R. Parasuraman, "Event-related potentials (erps) in neuroergonomics." 2006.
33. F. Dehais, A. Duprès, S. Blum, N. Drougard, S. Scannella, R. N. Roy, and F. Lotte, "Monitoring pilot's mental workload using erps and spectral power with a six-dry-electrode eeg system in real flight conditions," *Sensors*, vol. 19, no. 6, p. 1324, 2019.
34. C. S. Herrmann, D. Strüber, R. F. Helfrich, and A. K. Engel, "Eeg oscillations: from correlation to causality," *International Journal of Psychophysiology*, vol. 103, pp. 12–21, 2016.
35. A. T. Pope, E. H. Bogart, and D. S. Bartolome, "Biocybernetic system evaluates indices of operator engagement in automated task," *Biological psychology*, vol. 40, no. 1-2, pp. 187–195, 1995.
36. F. Lotte, "A tutorial on eeg signal-processing techniques for mental-state recognition in brain-computer interfaces," *Guide to brain-computer music interfacing*, pp. 133–161, 2014.
37. M. Congedo, A. Barachant, and R. Bhatia, "Riemannian geometry for eeg-based brain-computer interfaces; a primer and a review," *Brain-Computer Interfaces*, vol. 4, no. 3, pp. 155–174, 2017.
38. A. Bulling, J. A. Ward, H. Gellersen, and G. Tröster, "Eye movement analysis for activity recognition using electrooculography," *IEEE transactions on pattern analysis and machine intelligence*, vol. 33, no. 4, pp. 741–753, 2010.
39. J. Schmidt, R. Laarousi, W. Stolzmann, and K. Karrer-Gauß, "Eye blink detection for different driver states in conditionally automated driving and manual driving using eog and a driver camera," *Behavior research methods*, vol. 50, pp. 1088–1101, 2018.
40. P. Ayres, J. Y. Lee, F. Paas, and J. J. Van Merriënboer, "The validity of physiological measures to identify differences in intrinsic cognitive load," *Frontiers in psychology*, vol. 12, p. 702538, 2021.
41. R. N. Roy, S. Charbonnier, and S. Bonnet, "Eye blink characterization from frontal eeg electrodes using source separation and pattern recognition algorithms," *Biomedical Signal Processing and Control*, vol. 14, pp. 256–264, 2014.
42. P. K. Jain and A. K. Tiwari, "Heart monitoring systems—a review," *Computers in biology and medicine*, vol. 54, pp. 1–13, 2014.
43. U. Rajendra Acharya, K. Paul Joseph, N. Kannathal, C. M. Lim, and J. S. Suri, "Heart rate variability: a review," *Medical and biological engineering and computing*, vol. 44, pp. 1031–1051, 2006.
44. A. Schäfer and J. Vagedes, "How accurate is pulse rate variability as an estimate of heart rate variability?: A review on studies comparing photoplethysmographic technology with an electrocardiogram," *International journal of cardiology*, vol. 166, no. 1, pp. 15–29, 2013.
45. A. Temko, "Accurate heart rate monitoring during physical exercises using ppg," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 9, pp. 2016–2024, 2017.
46. A. J. Camm, M. Malik, J. T. Bigger, G. Breithardt, S. Cerutti, R. J. Cohen, P. Coumel, E. L. Fallen, H. L. Kennedy, R. E. Kleiger *et al.*, "Heart rate variability: standards of measurement, physiological interpretation and clinical use. task force of the european society of cardiology and the north american society of pacing and electrophysiology," *Circulation*, vol. 93, no. 5, pp. 1043–1065, 1996.
47. H. F. Posada-Quintero and K. H. Chon, "Innovations in electrodermal activity data collection and signal processing: A systematic review," *Sensors*, vol. 20, no. 2, p. 479, 2020.

48. K. Kyriakou, B. Resch, G. Sagl, A. Petutschnig, C. Werner, D. Niederseer, M. Liedlgruber, F. Wilhelm, T. Osborne, and J. Pykett, "Detecting moments of stress from measurements of wearable physiological sensors," *Sensors*, vol. 19, no. 17, p. 3805, 2019.
49. M. Rihet, A. Clodic, and R. N. Roy, "Robot noise: Impact on electrophysiological measurements and recommendations," in *Companion of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*, 2024, pp. 888–891.
50. A. Delorme, T. Sejnowski, and S. Makeig, "Enhanced detection of artifacts in eeg data using higher-order statistics and independent component analysis," *Neuroimage*, vol. 34, no. 4, pp. 1443–1449, 2007.
51. S. Blum, N. S. Jacobsen, M. G. Bleichner, and S. Debener, "A riemannian modification of artifact subspace reconstruction for eeg artifact handling," *Frontiers in human neuroscience*, vol. 13, p. 141, 2019.
52. M. Rodríguez-Ugarte, E. Iáñez, M. Ortíz, and J. M. Azorín, "Personalized offline and pseudo-online bci models to detect pedaling intent," *Frontiers in neuroinformatics*, vol. 11, p. 45, 2017.
53. M. Lagomarsino, M. Lorenzini, P. Balatti, E. De Momi, and A. Ajoudani, "Pick the right co-worker: Online assessment of cognitive ergonomics in human-robot collaborative assembly," *IEEE Transactions on Cognitive and Developmental Systems*, 2022.
54. A. H. Memar and E. T. Esfahani, "Objective assessment of human workload in physical human-robot cooperation using brain monitoring," *ACM Transactions on Human-Robot Interaction (THRI)*, vol. 9, no. 2, pp. 1–21, 2019.
55. A. Buerkle, H. Matharu, A. Al-Yacoub, N. Lohse, T. Bamber, and P. Ferreira, "An adaptive human sensor framework for human–robot collaboration," *The International Journal of Advanced Manufacturing Technology*, pp. 1–16, 2022.
56. Y. Liu, M. Habibnezhad, and H. Jebelli, "Brainwave-driven human-robot collaboration in construction," *Automation in Construction*, vol. 124, p. 103556, 2021.
57. E. Verna, S. Puttero, G. Genta, M. Galetto *et al.*, "A novel diagnostic tool for human-centric quality monitoring in human-robot collaboration manufacturing," *Journal of Manufacturing Science and Engineering*, vol. 145, no. 12, p. 121009, 2023.
58. A. Bettoni, E. Montini, M. Righi, V. Villani, R. Tsvetanov, S. Borgia, C. Secchi, and E. Carpanzano, "Mutualistic and adaptive human-machine collaboration based on machine learning in an injection moulding manufacturing line," *Procedia CIRP*, vol. 93, pp. 395–400, 2020.
59. V. Villani, B. Capelli, C. Secchi, C. Fantuzzi, and L. Sabattini, "Humans interacting with multi-robot systems: a natural affect-based approach," *Autonomous Robots*, vol. 44, pp. 601–616, 2020.
60. R. Gervasi, K. Aliev, L. Mastrogiacomo, and F. Franceschini, "User experience and physiological response in human-robot collaboration: a preliminary investigation," *Journal of Intelligent & Robotic Systems*, vol. 106, no. 2, p. 36, 2022.
61. T. Arai, R. Kato, and M. Fujita, "Assessment of operator stress induced by robot collaboration in assembly," *CIRP annals*, vol. 59, no. 1, pp. 5–8, 2010.
62. 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.
63. D. Kulić and E. Croft, "Physiological and subjective responses to articulated robot motion," *Robotica*, vol. 25, no. 1, pp. 13–27, 2007.
64. S. K. Ehrlich and G. Cheng, "Human-agent co-adaptation using error-related potentials," *Journal of neural engineering*, vol. 15, no. 6, p. 066014, 2018.
65. A. F. Salazar-Gomez, J. DelPreto, S. Gil, F. H. Guenther, and D. Rus, "Correcting robot mistakes in real time using eeg signals," in *2017 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2017, pp. 6570–6577.
66. F. Dehais, W. Karwowski, and H. Ayaz, "Brain at work and in everyday life as the next frontier: grand field challenges for neuroergonomics," p. 583733, 2020.
67. A.-M. Brouwer, M. A. Hogervorst, J. B. Van Erp, T. Heffelaar, P. H. Zimmerman, and R. Oostenveld, "Estimating workload using eeg spectral power and erps in the n-back task," *Journal of neural engineering*, vol. 9, no. 4, p. 045008, 2012.

68. D. Van't Ent and P. Apkarian, "Motoric response inhibition in finger movement and saccadic eye movement: a comparative study," *Clinical Neurophysiology*, vol. 110, no. 6, pp. 1058–1072, 1999.
69. S. Nieuwenhuis, K. R. Ridderinkhof, J. Blom, G. P. Band, and A. Kok, "Error-related brain potentials are differentially related to awareness of response errors: evidence from an antisaccade task," *Psychophysiology*, vol. 38, no. 5, pp. 752–760, 2001.
70. C. Savur, S. Kumar, and F. Sahin, "A framework for monitoring human physiological response during human robot collaborative task," in *2019 IEEE international conference on systems, man and cybernetics (SMC)*. IEEE, 2019, pp. 385–390.
71. S. Saha and M. Baumert, "Intra-and inter-subject variability in eeg-based sensorimotor brain computer interface: a review," *Frontiers in computational neuroscience*, vol. 13, p. 87, 2020.
72. Z. Wan, R. Yang, M. Huang, N. Zeng, and X. Liu, "A review on transfer learning in eeg signal analysis," *Neurocomputing*, vol. 421, pp. 1–14, 2021.
73. R. Schulz, P. Kratzer, and M. Toussaint, "Preferred interaction styles for human-robot collaboration vary over tasks with different action types," *Frontiers in neurorobotics*, vol. 12, p. 36, 2018.
74. K. Pollmann, W. Loh, N. Fronemann, and D. Ziegler, "Entertainment vs. manipulation: Personalized human-robot interaction between user experience and ethical design," *Technological Forecasting and Social Change*, vol. 189, p. 122376, 2023.
75. C. A. Fontanillo Lopez, G. Li, and D. Zhang, "Beyond technologies of electroencephalography-based brain-computer interfaces: A systematic review from commercial and ethical aspects," *Frontiers in Neuroscience*, vol. 14, p. 611130, 2020.