



**HAL**  
open science

# Robot Noise: Impact on Electrophysiological Measurements and Recommendations

Mathias Rihet, Aurélie Clodic, Raphaëlle Roy

► **To cite this version:**

Mathias Rihet, Aurélie Clodic, Raphaëlle Roy. Robot Noise: Impact on Electrophysiological Measurements and Recommendations. International Conference on Human-Robot Interaction, Mar 2024, Boulder, France. pp.888-891, 10.1145/3610978.3640708 . hal-04844682

**HAL Id: hal-04844682**

**<https://hal.science/hal-04844682v1>**

Submitted on 18 Dec 2024

**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.

# Robot Noise: Impact on Electrophysiological Measurements and Recommendations

Mathias Rihet  
ISAE-SUPAERO, Université de  
Toulouse  
Toulouse, France  
mathias.rihet@isae-supaero.fr

Aurélie Clodic  
LAAS-CNRS, Université de Toulouse,  
CNRS  
Toulouse, France  
aurelie.clodic@laas.fr

Raphaëlle N. Roy  
ISAE-SUPAERO, Université de  
Toulouse  
Toulouse, France  
raphaelle.roy@isae-supaero.fr

## ABSTRACT

Evaluating Human-Robot Interaction (HRI) is crucial for understanding the value that robots would bring to daily life. This paper investigates the robustness of machine learning classification techniques in interpreting physiological signals during HRI, considering potential artifacts induced by robot behavior. This phenomenon was explored with a 30 participants user study involving three cognitive efforts levels. This study used various physiological sensors, including Electroencephalography (EEG), Photoplethysmography (PPG), and Electrodermal Activity (EDA). Results reveal that EEG and PPG signal were impacted by robot-induced noise while EDA was not. By changing preprocessing parameters, EEG was also cleaned from robot noise and revealed better performances than EDA. The study highlights the importance of careful signal selection, balancing robustness and informativeness, and underscores the significance of preprocessing to ensure accurate classification aligned with users' mental states.

## CCS CONCEPTS

• **Human-centered computing** → **User models; User studies; Laboratory experiments; Interaction techniques;** • **Computer systems organization** → **Robotics.**

## KEYWORDS

HRI, evaluation, physiology, data acquisition, cognitive state

### ACM Reference Format:

Mathias Rihet, Aurélie Clodic, and Raphaëlle N. Roy. 2024. Robot Noise: Impact on Electrophysiological Measurements and Recommendations. In *Companion of the 2024 ACM/IEEE International Conference on Human-Robot Interaction (HRI '24 Companion)*, March 11–14, 2024, Boulder, CO, USA. ACM, New York, NY, USA, 4 pages. <https://doi.org/10.1145/3610978.3640708>

## 1 INTRODUCTION

Evaluating Human-Robot Interaction (HRI) is essential to determine the real value that robots can add to our daily life. Extensive research in HRI over the years has resulted in a diverse range of methods, metrics, and levels of observation. Among these metrics,

physiological signals are increasingly popular due to their ability to offer real-time and objective data derived from human mental state [6, 11]. These data can be decoded automatically with the adjunction of machine learning techniques, a field known as physiological computing [4], or brain-computer interface (BCI) when specifically used to deal with brain signal [8]. Such automatic decoding can lead to online adaptation of robot behavior during the interaction based on the collected physiological measures [1].

In this context, it becomes crucial to ensure that machine learning techniques used to interpret physiological signals are robust to potential artifacts induced by robots' behavior. Indeed, robot actions and movements can generate electrical or mechanical interferences that may introduce noise or distortions in the recorded physiological signals. By testing the robustness of machine learning techniques to such artifacts, this paper aims to ensure that the interpretation of physiological data aligns more accurately with the user's actual mental state. To the best of our knowledge, this important phenomenon has never been reported in the literature. Hence, we hypothesize that temporal windows containing artifacts induced by the robot will result in lower classification accuracy than temporal windows clear of robot noise.

## 2 METHODS



Figure 1: A typical HRI experimental setup, with a participant equipped with electrophysiological measurement devices sat in front of a PR2 robot.

### 2.1 Materials

#### Robot

The PR2 (Personal Robot 2) was chosen for its compatibility with ROS along with its ability to grasp objects and play sounds.

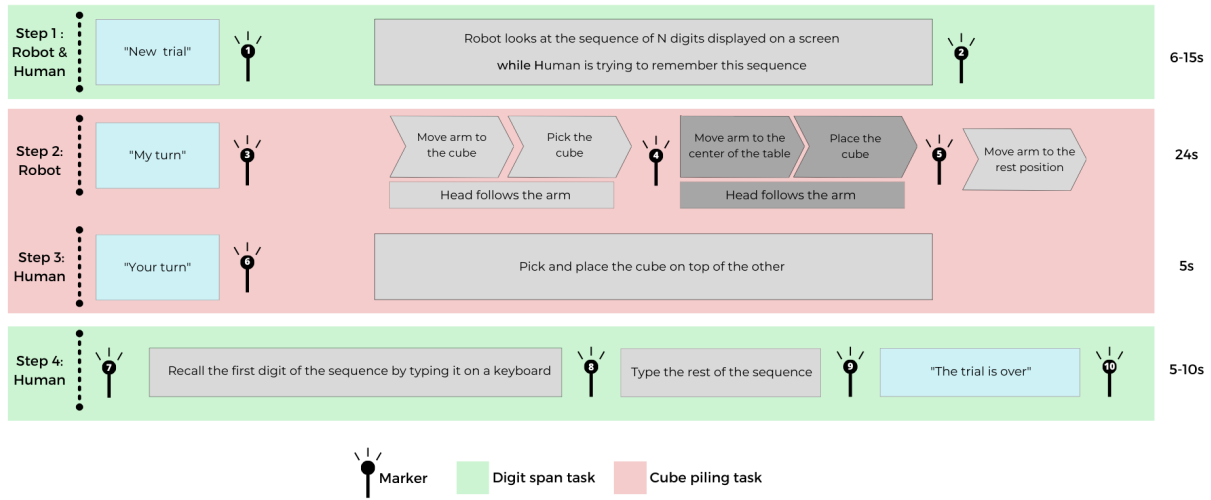


Figure 2: Task overview with markers and the duration of each step.

Sentences allowed to sequence the trial, both verbally and by raising a marker in the recording as presented in Figure 2 (1: "New trial", 3: "My turn", 6: "Your turn", 10: "The trial is over").

**Physiological Sensors**

Electroencephalography (EEG), chosen for its resolution at millisecond level, recorded brain electrical activity using a 20-channel Enobio system by Neuroelectrics. Dry electrodes were positioned on a cap following the 10/20 international system, and data were sampled at 500 Hz. Electrocardiography (ECG), photoplethysmography (PPG), electrodermal activity measurement (EDA), and eye-tracking respectively measured cardiac activity, blood volume changes, electrodermal activity, and ocular behavior. A Faros 360° collected the ECG data at 125 Hz, while EDA and PPG data were obtained using a Shimmer3 GSR+ Unit and sampled at 50 Hz. Eye-tracking was achieved using a Pupil Core at 200 Hz.

To ensure the sensitivity of these sensors, all data were collected in a quiet, stimulus-isolated room. The various collected streams were unified using the open-source ecosystem lab streaming layer (LSL) [5] to address synchronization issues.

**Tasks**

In order to induce different cognitive effort levels, digit sequences of various sizes were sequentially displayed on a wall-mounted screen near the robot (see Figure 2, Step 1). Participants were later asked to recall them on the same screen using a mouse and a virtual keyboard (see Figure 2, Step 4). Such a digit span task was chosen to replicate results from the HRI literature [7]. Hence, sequences of 1, 3, and 7 digits were designed to induce respectively low, medium, and high cognitive effort levels.

The joint manipulation task involved piling cubes with a robot (PR2) and started with participants seated at a table, the robot standing on the opposite side. A scripted dialogue signaled the start of PR2’s turn (see Figure 2, Step 2), during which it picked and placed a cube at the center of the table. This scripted sequence ensured consistency in duration (approximately 25 seconds). Subsequently, another dialogue indicated the beginning of participants’ turn (see

Figure 2, Step 3), instructing them to place a cube on the one placed by the robot.

In addition, a resting state was performed before and after the main task. This paradigm of growing interest in neuroscience [10] can be easily introduced in most neurophysiological experiments thanks to its simplicity and straightforward standardization. It lasts 5 min during which participants alternate 30-second periods with eyes opened and 30-second periods with eyes closed. Such recording of physiological activity at rest provides a useful baseline for data analysis.

Several markers were generated during these tasks, allowing to process physiological signals over various time-windows.

**2.2 Window segmentation**

The protocol was run once in the absence of participants to gather a baseline signal containing artifact produced by robot pick and place actions but no brain activity (Fig. 3).

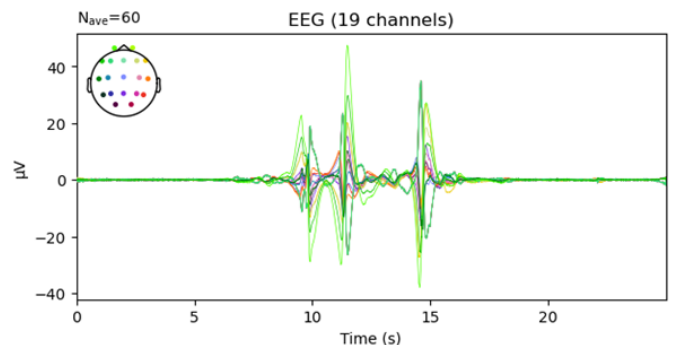
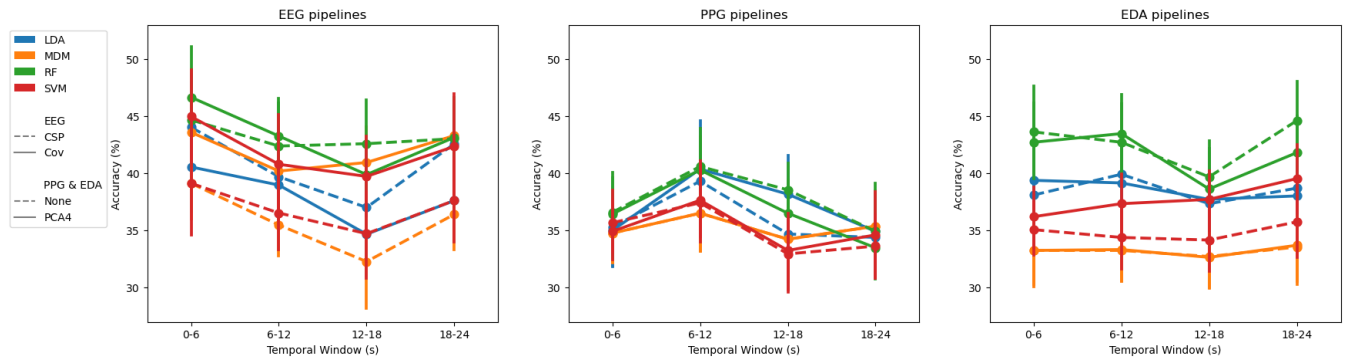


Figure 3: Mean EEG signal collected by an empty cap during an experimental session that includes several robot actions (average across trials).



**Figure 4: Classification accuracies for each signal type and pipeline (3-class estimation, adjusted chance level at 43.3% [3]).**

Based on this recording, the 24 seconds between step 2 and step 3 (see Figure 2) were split into four windows of 6 seconds each. The first and last window contained no robot noise while the second window contained two artifacts and the third window contained one artifact.

### 2.3 Preprocessing

Each signal underwent a minimal preprocessing step :

- The EEG signal was filtered using a FIR bandpass filter from 1 to 40 Hz and re-referenced by Common Average Reference (CAR).
- The PPG signal was preprocessed with a FIR bandpass filtering process within the range of 1 to 20 Hz.
- The EDA signal underwent a lowpass filter with a cutoff frequency at 3 Hz.

### 2.4 Classification

A classification analysis was performed on physiological signals to explore the robustness of classification under several conditions of robot noise. Various classifiers, including Random Forest (RF), Linear Discriminant Analysis (LDA), Support Vector Machine (SVM), and Minimum Distance to Mean (MDM), were trained on the previously defined windows to predict the associated digit condition, representing the cognitive effort level. In addition, spatial information was considered, with spatial filtering methods to reduce the dimensionality and enhance the signal to noise ratio (CSP [2] for EEG, and PCA for PPG and EDA), and EEG covariance matrices used in a Riemannian framework [9] as classically done in state-of-the-art EEG pipelines.

## 3 USER STUDY

### 3.1 Participants

Thirty volunteers took part in the experiment. Technical issues led to the exclusion of data from four participants, while four others were excluded due to poor signal quality. The remaining 22 participants (11 females, 11 males) had an average age of  $28 \pm 5.8$  years. Their prior knowledge with robots scored  $2.2 \pm 1.1$  on a scale of 1 “not at all” to 5 “very much”. On the same scale, prior knowledge with physiological sensors scored  $2.1 \pm 1.5$ . Participants received instructions about the experimental protocol and provided informed

consent. The study received approval from the institutional ethics review board of the University of Toulouse (project n°2022\_525).

### 3.2 Protocol validation

Subjective and behavioral data were used to test whether the established protocol successfully induced the targeted mental state, namely cognitive effort.

A One Way Repeated Measures ANOVA revealed that both perceived effort ( $F(2, 42) = 97.23, p < .001, \eta_p^2 = .82$ ) and answer accuracy ( $F(2, 42) = 22.08, p < .001, \eta_p^2 = .51$ ) varied significantly across digit sequences. On the other hand, response time did not vary significantly ( $F(2, 42) = 2.78, p = 0.07, \eta_p^2 = .12$ ).

While only reflected by variations in perceived effort and accuracy, cognitive effort seemed indeed induced by this protocol.

### 3.3 Classification results

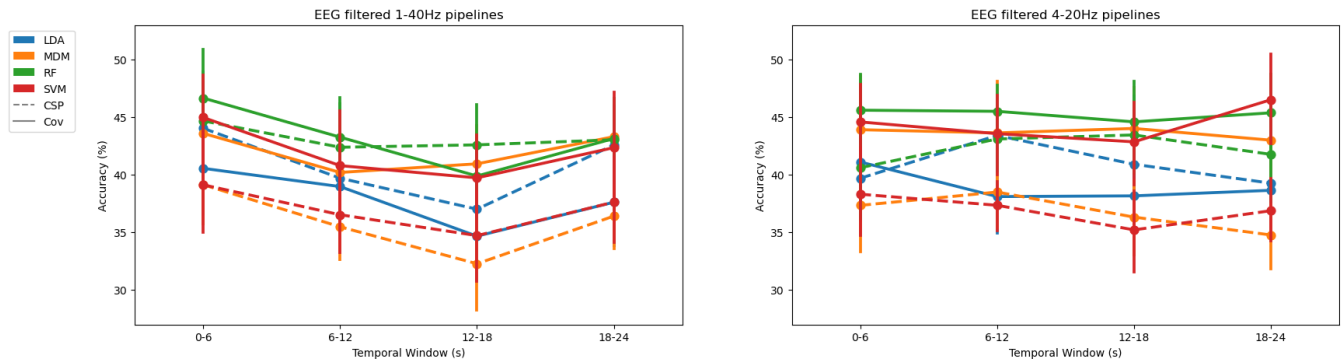
Pipeline performances can be found in Figure 4. A Two Way Repeated Measures ANOVA was used to test both the effect of pipeline and the effect of the window on classification accuracy.

Concerning the pipeline, a significant effect was revealed for EEG classification accuracy ( $F(7, 147) = 14.26, p < 0.001, \eta_p^2 = .09$ ), with the pipeline Cov + RF achieving 43.2% mean accuracy over the four windows, PPG classification accuracy ( $F(7, 147) = 2.79, p < 0.05, \eta_p^2 = .02$ ) with the pipeline RF achieving 37.6% mean accuracy over the four windows, and EDA classification accuracy ( $F(7, 147) = 12.62, p < 0.001, \eta_p^2 = .16$ ) with the pipeline RF achieving 42.6% mean accuracy over the four windows.

Concerning the window, a significant effect was found for EEG classification accuracy ( $F(3, 63) = 3.41, p < 0.05, \eta_p^2 = .04$ ) and PPG classification ( $F(3, 63) = 3.12, p < 0.05, \eta_p^2 = .04$ ) with the window 12-18 giving the worst results in both cases. EDA classification accuracy revealed no significant effect of the window.

### 3.4 Pipeline improvement

Preprocessing is a crucial step for EEG classification and manipulating some parameters can have a huge impact on performances. We addressed robot noise in EEG by narrowing the frequency band of the data. Hence EEG signal went through the same preprocessing step with a bandpass filter from 4 to 20Hz instead of 1 to 40 Hz.



**Figure 5: Classification accuracies depending on EEG pipelines with different preprocessing steps (3-class estimation, adjusted chance level at 43.3%).**

This preprocessing improvement allowed EEG classification to be more stable across windows and to reach 45.2% accuracy for Cov + RF pipeline.

#### 4 DISCUSSION & CONCLUSION

We hypothesized that temporal windows containing artifacts induced by the robot would yield a lower classification accuracy compared to artifact-free windows. Examining this hypothesis across a range of sensors highlighted variations in their sensitivity to robot noise. Both EEG and PPG signals appeared susceptible to these artifacts initially. However, a modification in the preprocessing pipeline effectively mitigated artifacts from the EEG signal. On the contrary, despite efforts, the PPG signal classification remained significantly impacted, underscoring the need for caution in its application in similar studies.

In addition, the EDA signal proved to be robust against robot noise, yet it led to less accurate classification compared to the EEG signal filtered between 4 and 20 Hz. This finding suggests that, despite its resilience to artifacts, the EDA signal is less informative than the EEG signal when it comes to mental state monitoring, at least in the case of cognitive effort.

It is noteworthy that, regardless of the signal type, the random forest classifier consistently demonstrated superior performance compared to other classifiers. This emphasizes the effectiveness of the random forest approach in handling the challenges posed by robot artifacts across different physiological signals.

In conclusion, our research highlighted the need to consider physiological signals carefully for studies involving humans and robots. A balance needs to be found between robustness to noise generated during the interaction and informativeness about the mental state of interest. Great attention should also be given to preprocessing, to ensure that classified data mainly contain physiological information.

#### ACKNOWLEDGMENTS

This work was supported by the French National Research Agency (ANR) with project EPIIC (ANR-21-CE33-0005), AI4HRI (ANR-20-IADJ-0006), Défi Clé Robotique centrée sur l’humain (funded by Région Occitanie, France) and by the Artificial and Natural Intelligence Toulouse Institute - Institut 3iA ANITI ANR-19-PI3A-0004.

#### REFERENCES

- [1] Maryam Alimardani and Kazuo Hiraki. 2020. Passive brain-computer interfaces for enhanced human-robot interaction. *Front. Rob. AI* 7 (2020), 125.
- [2] Benjamin Blankertz, Ryota Tomioka, Steven Lemm, Motoaki Kawanabe, and Klaus-Robert Muller. 2007. Optimizing spatial filters for robust EEG single-trial analysis. *IEEE Signal processing magazine* 25, 1 (2007), 41–56.
- [3] Etienne Combrisson and Karim Jerbi. 2015. Exceeding chance level by chance: The caveat of theoretical chance levels in brain signal classification and statistical assessment of decoding accuracy. *Journal of neuroscience methods* 250 (2015), 126–136.
- [4] Stephen H Fairclough. 2009. Fundamentals of physiological computing. *Interacting with computers* 21, 1-2 (2009), 133–145.
- [5] Christian Kothe, David Medine, Chadwick Boulay, Matthew Grivich, and Tristan Stenner. 2014. Lab streaming layer. <https://github.com/scn/labstreaminglayer>
- [6] Austin Kothig, John Munoz, Sami Alperen Akgun, Alexander M Aroyo, and Kerstin Dautenhahn. 2021. Connecting humans and robots using physiological signals—closing-the-loop in hri. In *IEEE Int Conf Robot & Human Interactive Communication (RO-MAN)*. IEEE, 735–742.
- [7] Yizhi Liu, Mahmoud Habibnezhad, and Houtan Jebelli. 2021. Brainwave-driven human-robot collaboration in construction. *Automation in Construction* 124 (2021), 103556.
- [8] Raphaëlle N Roy, Nicolas Drougard, Thibault Gateau, Frédéric Dehais, and Caroline PC Chanel. 2020. How can physiological computing benefit human-robot interaction? *Robotics* 9, 4 (2020), 100.
- [9] Raphaëlle N Roy, Marcel F Hinss, Ludovic Darnet, Simon Ladouce, Emilie S Jahanpour, Bertille Somon, Xiaoqi Xu, Nicolas Drougard, Frédéric Dehais, and Fabien Lotte. 2022. Retrospective on the first passive brain-computer interface competition on cross-session workload estimation. *Frontiers in Neuroergonomics* 3 (2022), 838342.
- [10] Abraham Z Snyder and Marcus E Raichle. 2012. A brief history of the resting state: the Washington University perspective. *Neuroimage* 62, 2 (2012), 902–910.
- [11] Valeria Villani, Massimiliano Righi, Lorenzo Sabatini, and Cristian Secchi. 2020. Wearable devices for the assessment of cognitive effort for human-robot interaction. *IEEE Sensors Journal* 20, 21 (2020), 13047–13056.