Mental Demand without brain signals: A practical approach to modeling the N-Back task

Luis Alberto Barradas-Chacón[1] Miriam de Jesús Sánchez Gama[2] Leticia Chacón Gutiérrez[2]

[1]Technische Universität Graz, Institute of Human-Centred Computing [2]Universidad La Salle Bajío, Centro de Neurociencias

Introduction

Cognitive workload is the measure of cognitive resources used by an individual to perform a task. Common methods for assessing it include behavioral response times and reported subjective states. Integrating physiological responses to its evaluation is not only useful, but crucial for understanding cognitive processes. However, the cost, complexity, and invasiveness of collecting physiological data can be a barrier to its widespread use in cognitive modeling. We propose a practical, cost-effective, and reproducible framework for cognitive modeling of mental demand using the N-Back task. We investigate the robustness of various methodologies for analyzing data from subjective reports, pupillometry, and heart rate variability (HRV). This research aims to show how these diverse data sources can provide a comprehensive view of cognitive processes, with a focus on implementing a complete, cost-effective pipeline adaptable to different size of experiments.

Methods

Thirty-three participants performed the N-Back task across three difficulty levels, with response timings and biosignals recorded. Data collection was streamlined using open-source tools like PsychoPy for task presentation and LabStreamingLayer for synchronization, ensuring methodological reproducibility. Subjective reports were gathered via the NASA-TLX, directly integrated into the task for practical assessment. Physiological measurements, heart rate variability (HRV) using a sports heart rate band and pupil diameters with a Tobii-Pro® eyetracker, were chosen for their minimal-to-noninvasive nature and potential as reliable indicators of mental demand.

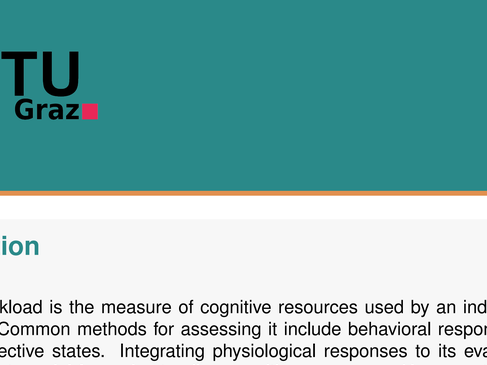

Figure 1. The N-Back task

Participants performed nine blocks of the Letter N-Back task, each lasting 90 seconds. It required them to monitor a sequence of letters and respond when the current one matched the one presented *n* trials back. This task is widely used to assess working memory and cognitive workload.

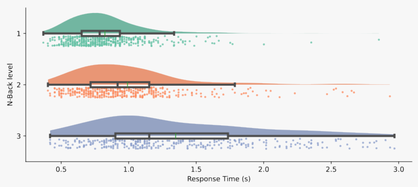

Figure 2. Distribution of the response times of the three different N-Back levels.

Response times (RTs) are recorded for each trial, and the task is designed to increase in difficulty as the N-Back level increases. The RTs are used as a measure of cognitive load, with longer RTs indicating higher cognitive demand.

Tools

To evaluate informational content of different data sources, we used four common machine learning models: Logistic Regression (**LR**), Naive Bayes (**NB**), K-Nearest Neighbors (**KNN**), and Random Forest Classifier (**RFC**). These models were trained to predict the N-Back level based on three different data inputs: the participant's response time (**RT**), NASA-TLX responses, or both. The models were evaluated using Cohen's Kappa score, which allows a fair comparison of model performance across different class distributions.

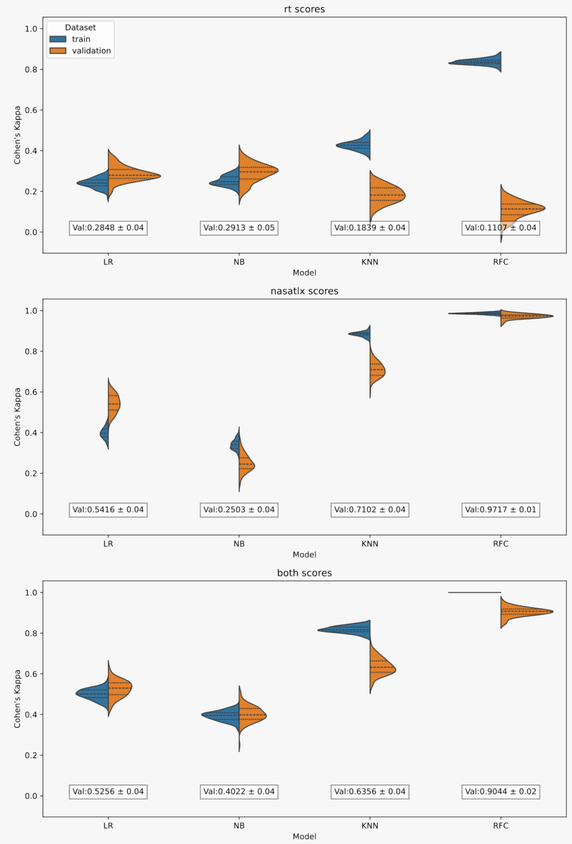

Figure 3. Distribution of Cohen's Kappa scores for four ML models predicting the N-Back level given different data inputs: the participant's response time, NASA-TLX responses, or both.

Biosignals

A gaussian optimization Tree-structured Parzen Estimator algorithm was used to evaluate the best performing features for each biosignal. For HRV, the maximum heart rate and the normalized proportion of low-frequency power were selected as features. For pupillometry, the peak pupil response in the 300 to 1000 ms window after stimulus onset during the N-Back task was used as a feature. The best performing model from the previous step was trained using these features to predict the N-Back level, and their performance was evaluated using accuracy and Cohen's Kappa score.

Results

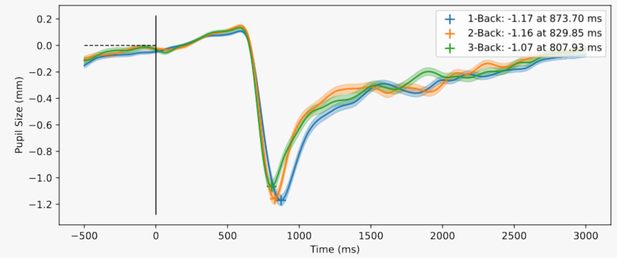

Figure 4. Grand Average Event-Related Pupillary Response for the three N-Back levels.

Although average pupillary responses show a trend across N-Back levels, individual responses show no significant statistical differences. A data loss of about half of the trials was observed. This is consistent with open datasets of similar experiments.

Biosignal	Accuracy (std)	Kappa (std)
HRV	93.13% (.01)	0.8967 (.01)
Pupillometry	41.83% (.02)	0.0999 (.02)

Table 1. Performance Metrics for RFC models predicting N-Back level using different biosignals.

HRV features show a high accuracy and Kappa score when predicting the N-Back level, indicating a strong relationship between heart rate variability and cognitive workload.

In short

We are able to:

- Characterize the certainty of each data source in modeling Cognitive Load for the N-Back task.
- Contrast **subjective reports**, **pupillary response**, **heart rate variability** and **response timings** as contextual indicators for a cognitive variable.
- Gauge and compare the informational content of diverse data sources.
- Demonstrate cost-effectiveness and reproducibility of different experimental methods.
- Provide a **practical**, **open**, and **reproducible** framework for modeling of cognitive mental workload.

Results show that the extra cost of collecting pupillometry data and the amount of noise in it justify their replacement with the use of low-cost HR bands for measuring heart rate variability, in the context of the N-Back task. The use of response timings and subjective reports is sufficient to model cognitive load, with a high degree of certainty.

Acknowledgements

This research was supported by the Technische Universität Graz's International Office as a Research Abroad Grant, the Institute of Human-Centred Computing at Technische Universität Graz, and by the Universidad La Salle Bajío's Center for Neuroscience. We would like to thank to the Center's students; Alexa, Ali, Emi, and Pame, for their help during data collection.

We would like to thank **Dr. Leticia Chacón Gutiérrez** and **M.Sc. Carlos Barradas-Chacón** for their support and guidance throughout this project.