

# Physiological Differences Between TASK and RELAX Conditions in Immersive Virtual Reality Using a Smartwatch

**Giulia-Ana Pop**

Technical University of Cluj-Napoca  
Faculty of Automation and Computer Science  
Cluj-Napoca, Romania  
popgiuliaana@yahoo.com

**Teodor Traian Ștefănuț**

Technical University of Cluj-Napoca  
Faculty of Automation and Computer Science  
Cluj-Napoca, Romania  
stefanutt@gmail.com

## ABSTRACT

This paper presents an experimental system for comparing physiological responses during cognitively demanding and relaxing phases in immersive virtual reality. The system combines a Unity application running on a Meta Quest 2 headset with physiological measurements collected using a Samsung Galaxy Watch 5. Three alternating TASK–RELAX cycles are automatically timestamped and synchronised offline with BPM, RR, RMSSD, and SDNN measurements. A pilot study with ten participants showed higher BPM and lower RR and SDNN during TASK phases, with the same direction observed for all participants. RMSSD showed a smaller and less consistent difference. Paired statistical analysis identified significant TASK–RELAX differences for BPM, RR, and SDNN, but not for RMSSD. The results demonstrate the feasibility of the proposed workflow while emphasising the influence of movement, individual variability, and the limitations of a small pilot study using wrist-worn sensing.

## AUTHOR KEYWORDS

virtual reality; cognitive state; smartwatch; heart rate variability; wearable sensing; physiological monitoring

**DOI:** 10.37789/icusi.2026.32

## INTRODUCTION

Immersive virtual reality (VR) is increasingly used in education, professional training, simulation, healthcare, and interactive applications because it enables users to perform complex activities inside controlled three-dimensional environments. Compared with conventional screen-based systems, immersive VR offers a stronger sense of presence and allows users to interact directly with virtual objects, spatial environments, and simulated scenarios [9]. These characteristics make VR particularly suitable for applications that require engagement, repeated practice, spatial orientation, and decision making.

However, immersive interaction may also place considerable cognitive demands on the user. A person using a VR application must process the content of the task while simultaneously interpreting visual and auditory stimuli, navigating through the environment, controlling virtual actions, and responding to events. The resulting cognitive state may vary according to task difficulty, exposure duration, interaction complexity, previous VR experience, and individual characteristics.

Cognitive load generally refers to the amount of mental effort required to perform a task [10]. It can be investigated through subjective questionnaires, behavioural performance, and physiological signals [6]. Questionnaires provide information about the user's perceived effort, while behavioural

measurements can include response time, errors, or task completion. Physiological signals offer a complementary perspective because they can be collected continuously throughout the interaction and compared across different moments of the VR experience.

Heart rate and heart rate variability (HRV) are frequently considered in studies involving cognitive demand, stress, attention, and relaxation. Heart rate is commonly expressed as beats per minute (BPM), while HRV can be described through indicators such as the interval between successive heartbeats (RR), the root mean square of successive differences (RMSSD), and the standard deviation of normal-to-normal intervals (SDNN) [7, 11]. Changes in these indicators may provide indirect information about the user's physiological response under different experimental conditions.

The development of commercial wearable devices has made this type of monitoring more accessible. Smartwatches can collect physiological measurements while allowing users to interact naturally with a VR environment, without requiring complex laboratory equipment. Although wrist-based measurements may be influenced by movement and sensor placement, they remain useful for exploratory systems that compare relative variations between clearly defined phases of an experiment [8].

Previous studies have investigated cognitive load in immersive VR using physiological measurements, eye tracking, behavioural data, and machine-learning models. Wei et al. used physiological and behavioural markers to estimate cognitive load in virtual reality [12]. Nasri proposed a physiologically adaptive VR framework in which task difficulty could be adjusted according to the estimated state of the user [5]. Han et al. reviewed the relationship between highly immersive VR environments and cognitive load, highlighting the importance of combining multiple evaluation methods [1]. These studies demonstrate the potential of using physiological information to better understand how users respond to immersive experiences.

Building on this direction, the present work investigates differences between two contrasting conditions created within the same VR experiment: a cognitively demanding TASK condition and a RELAX condition. Rather than implementing an automatic cognitive-state classifier, physiological indicators are used as an indirect means of observing how the user's response changes between the two experimental conditions.

The proposed system combines a VR application developed in Unity and executed on a Meta Quest 2 headset with physiological measurements recorded using a Samsung Galaxy Watch 5 and the *Nami – Biofeedback & HRV* smartwatch application. The VR scenario consists of three alternating TASK–RELAX cycles. During the TASK phases, participants

navigate through a virtual maze, interact with the environment, and answer questions with different difficulty levels. During the RELAX phases, the interactive scenario is replaced with a calm audiovisual environment designed to reduce cognitive and interaction demands.

To support the comparison between the two conditions, the VR application automatically records the beginning and end of every TASK and RELAX phase. These temporal markers are stored in a timeline file and are subsequently synchronised with the physiological data exported from the smartwatch application. The synchronised dataset is then segmented according to the experimental phases, allowing BPM, RR, RMSSD, and SDNN values to be compared between TASK and RELAX.

The central research question addressed by this work is:

Can physiological indicators collected using a commercial smartwatch reveal observable and consistent differences between TASK and RELAX conditions during an immersive VR experience?

The main contributions of this work are the design and implementation of a complete VR experimental scenario with alternating TASK and RELAX phases, the integration of smartwatch-based physiological monitoring, a timestamp-based method for synchronising VR events with the recorded measurements, and an analysis of BPM and HRV indicators collected from ten participants.

The results indicate distinct tendencies between the two experimental conditions, including a higher average heart rate and lower average RR, RMSSD, and SDNN values during TASK phases. Statistical analysis further shows significant TASK-RELAX differences for BPM, RR, and SDNN, while the smaller RMSSD difference does not reach statistical significance.

In the longer term, this type of system could contribute to the development of adaptive VR applications. Physiological information could be used together with behavioural and interaction data to help an application adjust task difficulty, reduce the intensity of the experience, introduce relaxation periods, provide supportive feedback, or modify the duration of an activity according to the estimated state of the user.

## RELATED WORK

### Cognitive Load in Immersive Virtual Reality

Immersive virtual reality has been investigated not only as a means of increasing presence and engagement, but also as a context in which the user may experience substantial cognitive demands. The user must often process task-related information while simultaneously navigating a three-dimensional environment, interpreting audiovisual stimuli, and operating unfamiliar interaction mechanisms. Consequently, the level of immersion alone does not determine the effectiveness of a VR experience; task complexity, interface design, exposure duration, and the user's previous experience also influence the resulting cognitive load.

Han et al. reviewed 46 studies published between 2010 and 2021 that investigated cognitive load in highly immersive VR environments [1]. Their review showed considerable methodological diversity. Cognitive load was assessed using subjective questionnaires, behavioural measures, physiological signals, or combinations of these methods. The authors noted

that immersive VR may improve presence and engagement, but that additional sensory and interaction demands can also increase mental effort, particularly when users are unfamiliar with VR technology.

The Cognitive Affective Model of Immersive Learning proposed by Makransky and Petersen describes how immersion, presence, motivation, emotion, and cognitive processing interact in VR-based learning experiences [2]. According to this perspective, immersion may support learning by increasing engagement and presence, but it may also introduce additional processing demands that are not directly related to the learning task.

These findings support the need for methods that can observe the user throughout an immersive experience rather than relying only on a questionnaire completed after the session. Continuous measurements can help identify how the user's response changes between individual phases of an application.

### Physiological Indicators for Cognitive-State Analysis

Physiological signals have increasingly been used to complement subjective and behavioural evaluations of cognitive load. Heart rate, HRV, electrodermal activity, pupil diameter, eye movements, and brain activity can provide different perspectives on the user's response during an interactive task. Because these signals can be recorded continuously, changes can be associated with specific events or intervals within a VR session.

Wei et al. investigated cognitive-load inference using physiological and behavioural markers collected from 738 participants performing VR tasks of different complexity levels [12]. Their dataset included pupillometry, eye tracking, photoplethysmography, task performance, and self-reported mental effort. A dual-branch attention network combined subjective and objective information to estimate a continuous cognitive-load score. The reported mean absolute error of 0.11 demonstrated the potential of combining several data sources for automatic cognitive-load inference.

The scale and multimodal design of that study provide strong evidence that physiological information can contribute to understanding cognitive demand in VR. However, such an approach requires a large dataset, several sensing modalities, and a machine-learning pipeline. The work presented here examines a smaller and more accessible setup. Instead of predicting a continuous cognitive-load score, it compares physiological indicators between two clearly defined experimental conditions.

Heart-rate variability is particularly relevant because it describes the variation between consecutive cardiac cycles. RR intervals, RMSSD, and SDNN capture different properties of this variability [7, 11]. In combination with BPM, these measures can be used to compare the physiological response observed during demanding and less demanding phases. Their interpretation must nevertheless consider individual variability, movement, breathing, signal quality, and the duration of the analysed intervals.

### Physiologically Adaptive Virtual Reality

A major motivation for monitoring the user during VR interaction is the possibility of creating adaptive applications. Instead of presenting every user with the same fixed sequence, an adaptive system could modify the experience according to physiological, behavioural, or interaction-related information.

Nasri proposed a framework for intelligent VR training that uses HRV and eye-tracking measurements to identify cognitive load and stress [5]. The proposed approach first uses controlled tasks, including a Stroop task, to collect signals associated with different levels of demand. The resulting models can then support adaptations within a VR environment. When high cognitive demand is detected, the application may reduce task complexity, slow the interaction, or provide supportive audiovisual feedback. When estimated demand is low, the system may gradually increase the challenge.

This direction is relevant to educational and professional training applications because a fixed level of difficulty may not be appropriate for all users. However, before real-time adaptation can be implemented, it is necessary to establish whether the available sensors can reveal differences between relevant experimental conditions.

The present study addresses this earlier stage. Physiological data are analysed offline after being synchronised with the TASK and RELAX intervals generated by the VR application.

### Wearable Monitoring and VR Relaxation

Commercial smartwatches provide a practical alternative to complex laboratory sensing systems. They are relatively affordable, familiar to users, and can record physiological data without significantly restricting interaction. Morales et al. reviewed the use of smartwatches and physiological biomarkers for occupational stress monitoring [3]. Their review identified HRV as one of the frequently investigated indicators and highlighted the potential of wearable devices for continuous monitoring.

The use of wrist-worn sensors also introduces limitations. Optical measurements can be affected by hand and arm movement, device placement, skin contact, and differences between users. These factors are particularly important in VR because interaction with controllers may generate motion artefacts.

VR has also been investigated as a method for supporting relaxation. Naef et al. compared audiovisual, auditory-only, visual-only, and control conditions during a VR relaxation intervention [4]. Their results showed measurable changes in physiological variables such as heart rate, respiration rate, and blood pressure. The study supports the use of a calmer audiovisual VR condition as a contrast to cognitively demanding tasks.

The current experiment combines these two directions. A commercial smartwatch records BPM and HRV-related indicators, while the VR application alternates between an interactive TASK condition and a calmer RELAX condition.

### Positioning of the Present Work

Previous research demonstrates that cognitive demand in VR can be investigated using physiological measurements and that these measurements may eventually support adaptive immersive applications. Nevertheless, many existing approaches rely on specialised sensors, eye-tracking systems, multiple simultaneous data streams, or large-scale machine-learning datasets.

The present work focuses on a simpler and more accessible configuration. It combines a Meta Quest 2 headset, a Samsung Galaxy Watch 5, and the *Nami – Biofeedback & HRV* application. The main distinction is the use of automatically generated VR timestamps to synchronise wearable measurements with three alternating TASK–RELAX cycles. Rather than

building an automatic classifier, the study evaluates whether BPM, RR, RMSSD, and SDNN exhibit observable and consistent differences between the two conditions.

## SYSTEM DESIGN AND IMPLEMENTATION

### System Overview

The proposed system was designed to support the comparison of physiological responses recorded during cognitively demanding and relaxing phases of an immersive VR experience. It consists of two independent data-acquisition components: a VR application that controls the experimental scenario and records its temporal structure, and a smartwatch application that records physiological measurements.

The VR component was developed in Unity and deployed on a Meta Quest 2 headset. It provides an interactive virtual environment containing alternating TASK and RELAX phases. During the experiment, the application records the start and end of each phase in a timeline CSV file.

The physiological component consists of a Samsung Galaxy Watch 5 running the *Nami – Biofeedback & HRV* application. The smartwatch records heart-rate and heart-rate-variability measurements throughout the session. After the experiment, the recorded data are exported as a CSV file containing timestamps and the BPM, RR, RMSSD, and SDNN indicators.

The two components operate independently during data collection. Their outputs are combined after the session using the timestamps contained in both CSV files. This offline architecture avoids the need for direct real-time communication between the VR application and the smartwatch while still allowing physiological measurements to be associated with individual TASK and RELAX intervals.

The separation between data acquisition and data analysis also makes the system easier to reproduce and extend. The VR application is responsible for controlling the experimental scenario and recording its temporal events, while the smartwatch application operates independently as the physiological sensing component.

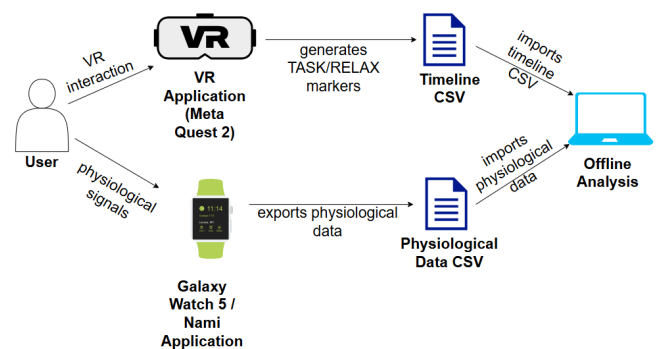


Figure 1: Architecture of the proposed system.

As illustrated in Figure 1, the participant simultaneously interacts with the VR application and wears the smartwatch. The VR application generates the experimental markers, whereas the smartwatch records the physiological response. The timeline and physiological files are later imported into the same analysis pipeline, where the measurements are segmented according to the TASK and RELAX intervals.

## VR Application

The VR application was implemented in Unity and configured for the Meta Quest 2 headset. Interaction was performed using the headset controllers, which allowed participants to navigate through the virtual environment and interact with the experimental interface.

The TASK environment was designed as an interactive maze. Participants were required to navigate through the virtual space while responding to questions presented during the task. This combination was selected to involve several cognitive processes, including sustained attention, spatial orientation, decision making, and the processing of questions with different difficulty levels.

The application controls the complete experimental sequence automatically. Once the session begins, the participant progresses through three TASK–RELAX cycles without requiring manual intervention from the experimenter. This ensures that the same sequence and phase durations are used for all participants.

## Experimental Phase Sequence

The session consists of three consecutive cycles:

- (1) TASK 1 followed by RELAX 1;
- (2) TASK 2 followed by RELAX 2;
- (3) TASK 3 followed by RELAX 3.

During each TASK phase, the participant navigates through the maze and answers questions. The questions are introduced at predefined moments and use different difficulty levels. The participant selects an answer using the VR controllers, after which the application continues the experimental sequence.

During each RELAX phase, the interactive maze and its task-related stimuli are replaced by a calm audiovisual experience. The relaxation environment reduces the interaction requirements and provides a contrasting condition against which the physiological response recorded during the TASK phase can be compared.

The alternating design also enables a within-participant comparison. Each participant experiences both conditions during the same session, which reduces the influence of differences in resting heart rate and other participant-specific characteristics on the comparison.

## Physiological Data Collection

Physiological measurements were recorded using a Samsung Galaxy Watch 5. Before the VR session, the smartwatch was placed securely on the participant's wrist and recording was started in the *Nami – Biofeedback & HRV* application. The watch remained in place throughout the complete experiment. The application provided continuous monitoring during the session and supported the export of the recorded measurements after completion. The exported CSV file contained timestamps and values for BPM, RR, RMSSD, and SDNN. BPM represents heart rate in beats per minute, while RR represents the duration between successive heartbeats. RMSSD and SDNN describe short-term and overall variations in the recorded heartbeat intervals.

The smartwatch recording was started shortly before the VR experiment and stopped after the final RELAX phase. Consequently, the physiological file could contain measurements recorded immediately before or after the experimental interval. These samples were excluded during preprocessing.

## Event Logging

The VR application automatically logs the principal events of the experimental session. Each event is stored in a timeline CSV file together with its timestamp and elapsed session time. The recorded events include:

- the beginning of the complete session;
- the start and end of every TASK phase;
- the start and end of every RELAX phase;
- the end of the complete session.

The timeline therefore defines the temporal boundaries required to separate the physiological measurements into the six experimental phases. Automatically generated markers were preferred to manually recorded timestamps because they provide a consistent and reproducible description of the session.

## Offline Synchronisation and Preprocessing

After each session, the timeline CSV generated by Unity and the physiological CSV exported from the smartwatch application were processed together. Both files contained temporal information that allowed physiological samples to be associated with the experimental phase active at that moment. Before segmentation, the files were checked for the required columns and valid timestamps. Missing, zero, or non-numeric physiological values were excluded. Physiologically implausible heart-rate values were also checked, while only positive and numerically valid RR, RMSSD, and SDNN measurements were retained. No additional artefact-correction algorithm was applied in the offline analysis beyond these validity checks. RMSSD and SDNN values were provided directly in the data exported by the Nami application and were not recalculated from raw RR intervals in the present analysis.

Samples recorded before the start of the VR session or after its end were removed. For every TASK and RELAX interval, the remaining physiological samples were selected according to the actual start and end markers stored in the timeline.

The use of actual event markers rather than only the theoretical phase durations ensured that every physiological value was assigned to the experimental interval that was active during execution.

The synchronised data formed the input for the subsequent analysis. For each participant, the three TASK phases were compared with the three RELAX phases using BPM, RR, RMSSD, and SDNN. The same processing procedure was applied to all participants.

## EXPERIMENTAL STUDY AND DATA ANALYSIS

### Participants and Protocol

Ten participants, identified as P1–P10, completed the experiment. Their ages ranged from 22 to 25 years; seven participants were female and three were male. Five participants reported previous experience with VR, while none reported known cardiac conditions. Only anonymised participant identifiers were used in the analysis.

Each participant completed the same 12-minute experimental sequence. The session contained three TASK–RELAX cycles, with each TASK and RELAX phase lasting 120 seconds. During TASK phases, participants navigated through the maze

and answered questions, whereas the RELAX phases presented a calm audiovisual environment. The order and duration of all phases were controlled automatically by the Unity application.

All participants provided informed consent before taking part in the study.

### Data Processing and Comparison

The analysis was implemented in Python. For every participant, the timeline generated by Unity was processed together with the physiological CSV exported from Nami. Each valid physiological sample was assigned to one of the six experimental intervals according to its timestamp.

For each physiological indicator, the values from the three TASK phases were grouped together and compared with the values from the three RELAX phases. The analysis used the mean values of BPM, RR, RMSSD, and SDNN for each condition.

The difference between conditions was calculated as

$$\Delta X = \bar{X}_{\text{TASK}} - \bar{X}_{\text{RELAX}}, \quad (1)$$

where  $X$  represents the analysed physiological indicator. A positive BPM difference indicates a higher average heart rate during TASK. Negative differences for RR, RMSSD, and SDNN indicate lower values during TASK than during RELAX.

The analysis was performed both at the participant level and at the aggregate group level. In addition to the mean differences, the consistency of the direction of change was calculated as the number of participants whose results followed the overall TASK–RELAX direction.

To provide inferential analysis in addition to the descriptive comparison, paired-sample  $t$ -tests were applied to the participant-level TASK–RELAX differences. For each indicator, a 95% confidence interval and the paired-samples effect size  $d_z$  were also calculated.

## RESULTS AND DISCUSSION

### Aggregated TASK–RELAX Comparison

Table 1 summarises the mean values obtained for the TASK and RELAX conditions. The difference was calculated as the TASK mean minus the RELAX mean. The consistency column indicates the number of participants whose individual results followed the same direction as the aggregate difference.

| Metric     | TASK   | RELAX  | Difference | Consistency |
|------------|--------|--------|------------|-------------|
| BPM        | 100.81 | 96.30  | +4.51      | 10/10       |
| RR (ms)    | 596.46 | 624.33 | −27.87     | 10/10       |
| RMSSD (ms) | 7.55   | 7.89   | −0.34      | 7/10        |
| SDNN (ms)  | 12.28  | 14.58  | −2.30      | 10/10       |

**Table 1: Aggregated physiological results for TASK and RELAX phases.**

The mean heart rate was 4.51 BPM higher during TASK phases than during RELAX phases. The same direction was observed for all ten participants, although the magnitude of the difference varied between individuals. Participant-level BPM differences ranged from +0.59 to +8.73 BPM.

The RR interval was, on average, 27.87 ms shorter during TASK phases. This result is consistent with the higher BPM

values because an increased heart rate corresponds to shorter intervals between successive heartbeats. All participants exhibited this direction of change.

SDNN was also lower during TASK phases, with an aggregate difference of −2.30 ms. The same direction was observed for all participants, indicating consistently lower SDNN values during the TASK condition.

RMSSD showed a smaller aggregate difference of −0.34 ms. Seven of the ten participants had lower RMSSD values during TASK phases, while three exhibited positive differences. Therefore, RMSSD followed the overall TASK–RELAX tendency less consistently than BPM, RR, and SDNN.

Figure 2 provides a visual comparison of the aggregate values reported in Table 1. The four plots illustrate the direction and relative magnitude of the differences between TASK and RELAX conditions.

Paired-sample tests were conducted using the participant-level TASK–RELAX differences.

BPM was significantly higher during TASK phases, with a mean difference of 4.51 BPM, 95% CI [2.79, 6.22],  $t(9) = 5.95$ ,  $p < .001$ , and a large paired-samples effect ( $d_z = 1.88$ ).

RR was significantly lower during TASK phases ( $M_\Delta = -27.87$  ms, 95% CI [-38.32, -17.41],  $t(9) = -6.03$ ,  $p < .001$ ,  $d_z = -1.91$ ).

SDNN was also significantly lower during TASK phases ( $M_\Delta = -2.30$  ms, 95% CI [-3.22, -1.37],  $t(9) = -5.62$ ,  $p < .001$ ,  $d_z = -1.78$ ).

The RMSSD difference was smaller and did not reach statistical significance ( $M_\Delta = -0.34$  ms, 95% CI [-0.96, 0.28],  $t(9) = -1.25$ ,  $p = .244$ ,  $d_z = -0.39$ ). This result is consistent with the participant-level analysis, in which seven of the ten participants showed lower RMSSD during TASK.

Because the study included only ten participants, these inferential results should be interpreted together with the descriptive and participant-level analyses rather than as strong evidence of generalisability to a broader population.

### Interpretation and Discussion

The participant-level analysis supports the aggregate findings. BPM, RR, and SDNN followed the same TASK–RELAX direction for all ten participants, showing that the group-level differences were not produced only by a small number of participants with unusually large responses. RMSSD followed the same direction for seven participants and therefore showed greater individual variability.

The complementary relationship between BPM and RR strengthens the interpretation of the cardiovascular response. During TASK phases, the increase in heart rate was accompanied by shorter intervals between successive heartbeats. SDNN also decreased consistently during TASK, while RMSSD showed a smaller and less uniform difference.

The statistical results support this distinction. BPM, RR, and SDNN showed significant paired differences with large paired-samples effect sizes, whereas RMSSD did not show a statistically significant difference. This result is compatible with the descriptive finding that RMSSD had the lowest participant-level consistency.

The general direction of the findings is consistent with previous research associating increased cognitive demand with higher heart rate, shorter RR intervals, and changes in HRV-related indicators [7, 12]. However, the present results should

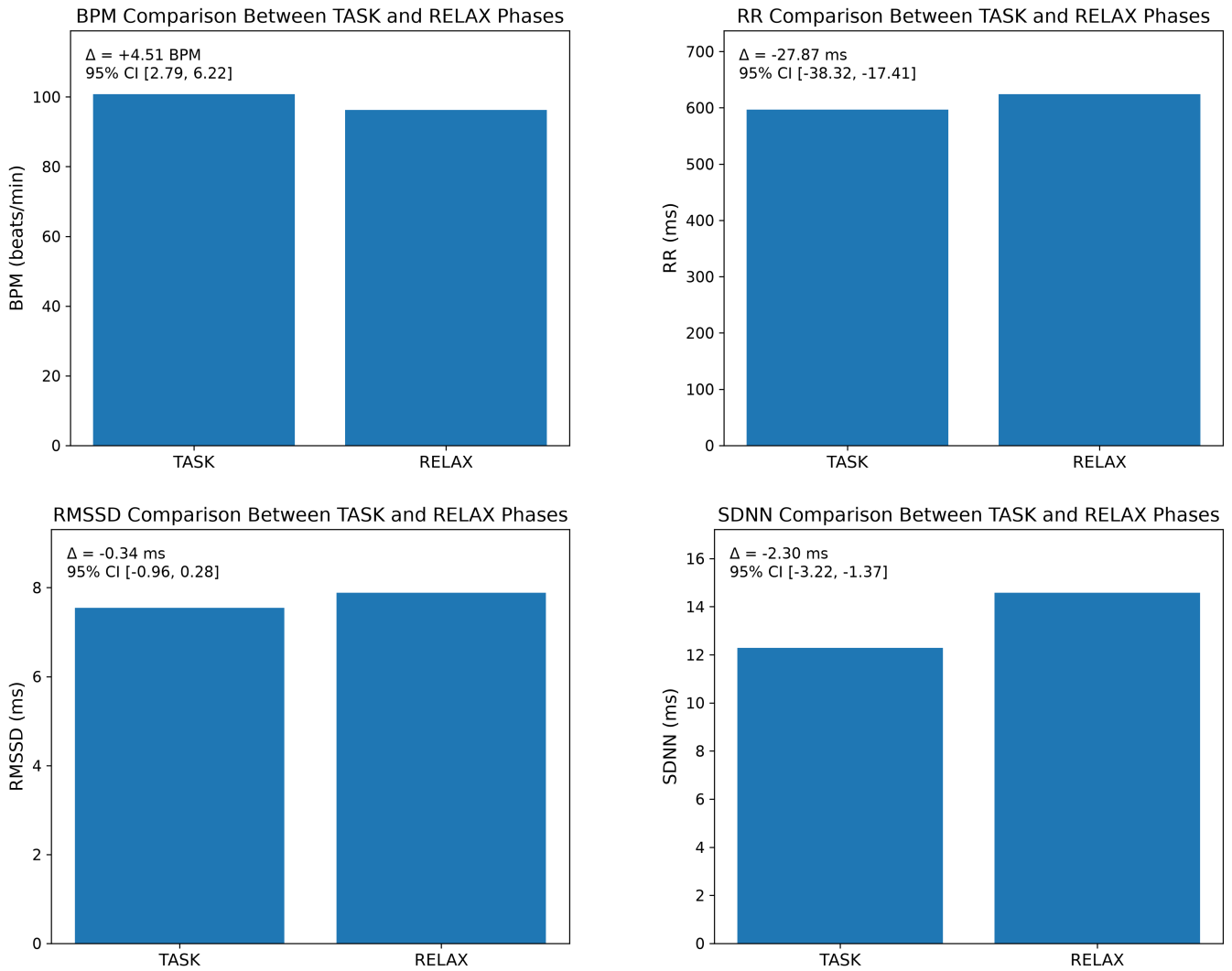


**Figure 2: Aggregated comparison of BPM, RR, RMSSD, and SDNN between TASK and RELAX phases. Each plot reports the participant-level mean TASK-RELAX difference and its 95% confidence interval.**

not be interpreted as showing that cognitive demand alone caused the observed differences.

An important consideration is that the TASK and RELAX conditions differ in more than cognitive demand alone. During TASK phases, participants navigate through the maze, use the VR controllers, process visual information, make decisions, and answer questions, whereas the RELAX condition requires substantially less active interaction. Consequently, physical movement, controller interaction, visual stimulation, or emotional arousal may also have contributed to the recorded physiological response.

This issue is particularly relevant for wrist-based photoplethysmography (PPG), because movement and sensor contact may affect optical measurements. Headset and controller movement data were not recorded in the present study. Future experiments should therefore record headset and controller movement or include a comparison condition with similar movement requirements but lower cognitive demand. Such a design could help distinguish the contribution of cognitive processing from that of physical interaction.

The cognitive manipulation was not independently validated using a subjective or behavioural cognitive-load measure. The present experiment did not include NASA-TLX, a mental-effort rating, response time, navigation errors, or task accuracy as outcome variables. Therefore, the TASK and RELAX labels describe the intended experimental conditions rather than independently verified levels of cognitive load. Future studies should combine physiological measurements with subjective and behavioural measures to provide stronger triangulation. The repeated three-cycle structure provides multiple exposures to both conditions. However, the sequence was fixed, with each TASK phase followed by a RELAX phase. Therefore, order, adaptation, fatigue, and physiological recovery may partly overlap with the experimental condition. A counterbalanced or randomised phase order would provide stronger evidence in future experiments.

The technical objectives of the pilot study were nevertheless achieved: the complete VR scenario was executed correctly, physiological recordings were successfully synchronised with the timeline, and the processed data revealed measurable differences between TASK and RELAX conditions.

These results support the feasibility of the implemented VR-smartwatch workflow as a basis for more controlled future studies.

### Limitations

The study was conducted with ten participants from a relatively narrow age range and should therefore be interpreted as a pilot investigation. The small sample allows the observation of initial tendencies but limits the generalisability of the findings.

The TASK and RELAX conditions also differ in physical interaction. TASK phases required navigation and controller use, while RELAX phases involved substantially less movement. Because wrist-worn measurements may be influenced by movement, sensor contact, and device placement, part of the observed physiological response may reflect physical interaction in addition to cognitive demand.

The fixed TASK-RELAX order introduces another limitation. Every TASK phase was followed by RELAX, meaning that condition effects cannot be fully separated from order, adaptation, fatigue, or physiological recovery effects.

The experiment did not include a resting baseline before the first TASK phase. Consequently, RELAX should be interpreted as the lower-demand comparison condition in this experiment rather than as a direct physiological resting baseline.

The 120-second intervals may also be short for the stable estimation of some HRV indicators. This limitation is particularly relevant to RMSSD, which showed greater participant-level variability and no statistically significant paired difference.

Finally, the experiment did not include an independent subjective or behavioural validation of cognitive load. Future studies should include baseline recordings, movement measurements, counterbalanced conditions, longer HRV analysis intervals, behavioural measures, and subjective instruments such as NASA-TLX or mental-effort ratings.

### ETHICS AND PRIVACY

The study was conducted as part of a master's dissertation under the supervision of Conf. Dr. Eng. Teodor Traian Stefanut. No separate formal ethics committee approval was obtained. All participants provided informed consent before taking part in the study. Participant data were analysed using anonymised identifiers (P1-P10), and no participant names or directly identifying information are reported in this paper.

### CONCLUSION AND FUTURE WORK

This work presented the design and evaluation of an experimental system for comparing physiological responses during TASK and RELAX conditions in an immersive virtual reality environment. The proposed system combines a VR application developed in Unity and executed on a Meta Quest 2 headset with physiological data collected using a Samsung Galaxy Watch 5 and the *Nami – Biofeedback & HRV* application.

The VR application alternates between TASK and RELAX phases. During TASK phases, participants navigate through a virtual maze and answer questions displayed in the VR environment, whereas during RELAX phases they are exposed to a calmer visual and auditory context. This structure enables comparison of the participant's physiological response under two contrasting experimental conditions.

An important contribution of the work is the implementation of the event-logging mechanism. The VR application automatically generates a `timeline.csv` file containing the start and end timestamps of every TASK and RELAX phase. This file provides the connection between the VR component and the offline analysis by allowing the physiological measurements exported from Nami to be synchronised and segmented according to the experimental intervals.

The pilot study included ten participants, and the analysis considered BPM, RR, RMSSD, and SDNN at both participant and aggregate levels. TASK phases showed higher mean BPM and lower mean RR, RMSSD, and SDNN values than RELAX phases. The direction of change was consistent for all participants for BPM, RR, and SDNN and for seven of the ten participants for RMSSD.

Paired statistical analysis showed significant TASK-RELAX differences for BPM, RR, and SDNN, with large paired-samples effect sizes. The smaller RMSSD difference did not reach statistical significance, which is consistent with its greater participant-level variability.

The findings demonstrate that the implemented system can capture physiological differences between TASK and RELAX conditions using a commercial smartwatch. However, because the two conditions also differ in physical interaction and the study did not include an independent measure of cognitive load, the findings should not be interpreted as evidence of automatic cognitive-state detection. Instead, they support the feasibility of the proposed VR-smartwatch workflow for further experimental investigation.

Future work should extend the study to a larger and more diverse participant group in order to reduce the influence of individual variability and examine whether the observed tendencies remain stable across different populations.

Future experiments should also use counterbalanced or randomised conditions and record movement from the headset and controllers. With a larger dataset, repeated-measures or mixed-effects models could also be used to examine differences between the three TASK-RELAX cycles and possible adaptation effects. Subjective and behavioural measures, including NASA-TLX, mental-effort ratings, response accuracy, response time, and navigation performance, could provide independent validation of the experimental manipulation.

The system could also be extended by integrating additional physiological information, including skin conductance, temperature, and movement data. Longer analysis windows and the inclusion of a resting baseline could further strengthen the interpretation of HRV indicators.

Another important direction is the transition from offline analysis to real-time processing. In a future version, physiological measurements could be transmitted directly to the VR application so that the virtual environment adapts automatically according to the estimated state of the user. For example, task difficulty, phase duration, stimulus intensity, or the introduction of relaxation periods could be adjusted according to the detected physiological response.

Overall, the implemented system provides a basis for both more detailed experimental investigations and future adaptive VR applications.

### References

- [1] Jining Han, Qiyu Zheng, and Yunan Ding. 2021. Lost in Virtual Reality? Cognitive Load in High Immersive VR Environments. *Journal of Advances in Information Technology* 12, 4 (2021), 302–310. doi:10.12720/jait.12.4.302-310

- [2] Guido Makransky and Gustav B. Petersen. 2021. The Cognitive Affective Model of Immersive Learning (CAMIL): A Theoretical Research-Based Model of Learning in Immersive Virtual Reality. *Educational Psychology Review* 33, 3 (2021), 937–958. doi:10.1007/s10648-020-09586-2
- [3] Analúcia Morales, Maria Barbosa, Laura Morás, Silvio César Cazella, Livia F. Sgobbi, Iwens Sene, and Gonçalo Marques. 2022. Occupational Stress Monitoring Using Biomarkers and Smartwatches: A Systematic Review. *Sensors* 22, 17 (2022), 6633. doi:10.3390/s22176633
- [4] Andrea C. Naef, Marie-Madlen Jeitziner, Sarah E. J. Knobel, Michael T. Exl, Rene M. Muri, Stephan M. Jakob, Tobias Nef, and Stephan M. Gerber. 2022. Investigating the Role of Auditory and Visual Sensory Inputs for Inducing Relaxation During Virtual Reality Stimulation. *Scientific Reports* 12 (2022), 17073. doi:10.1038/s41598-022-21575-9
- [5] Mahsa Nasri. 2025. Towards Intelligent VR Training: A Physiological Adaptation Framework for Cognitive Load and Stress Detection. In *Proceedings of the 33rd ACM Conference on User Modeling, Adaptation and Personalization*. Association for Computing Machinery, 419–423. doi:10.1145/3699682.3727575
- [6] Fred Paas, Juhani E. Tuovinen, Huib Tabbers, and Pascal W. M. Van Gerven. 2003. Cognitive Load Measurement as a Means to Advance Cognitive Load Theory. *Educational Psychologist* 38, 1 (2003), 63–71. doi:10.1207/S15326985EP3801\_8
- [7] Fred Shaffer and J. P. Ginsberg. 2017. An Overview of Heart Rate Variability Metrics and Norms. *Frontiers in Public Health* 5 (2017), 258. doi:10.3389/fpubh.2017.00258
- [8] Anna Shcherbina, C. Mikael Mattsson, Daryl Waggott, Heidi Salisbury, Jeffrey W. Christle, Trevor Hastie, Matthew T. Wheeler, and Euan A. Ashley. 2017. Accuracy in Wrist-Worn, Sensor-Based Measurements of Heart Rate and Energy Expenditure in a Diverse Cohort. *Journal of Personalized Medicine* 7, 2 (2017), 3. doi:10.3390/jpm7020003
- [9] Mel Slater and Maria V. Sanchez-Vives. 2016. Enhancing Our Lives with Immersive Virtual Reality. *Frontiers in Robotics and AI* 3 (2016), 74. doi:10.3389/frobt.2016.00074
- [10] John Sweller. 1988. Cognitive Load During Problem Solving: Effects on Learning. *Cognitive Science* 12, 2 (1988), 257–285. doi:10.1207/s15516709cog1202\_4
- [11] Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology. 1996. Heart Rate Variability: Standards of Measurement, Physiological Interpretation and Clinical Use. *Circulation* 93, 5 (1996), 1043–1065. doi:10.1161/01.CIR.93.5.1043
- [12] Jishang Wei, Erika Siegel, Prahalathan Sundaramoorthy, Antônio Gomes, Shibo Zhang, Mithra Vankipuram, Kevin Smathers, Sarthak Ghosh, Hiroshi Horii, Jeremy N. Bailenson, and Rafael Tico Ballagas. 2025. Cognitive Load Inference Using Physiological Markers in Virtual Reality. In *2025 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*. IEEE, 759–769. doi:10.1109/VR59515.2025.00098