HPO-CLD: HP Omnicept - Cognitive Load Dataset

1. Introduction

This file describes the data released in the HP Omnicept – Cognitive Load Dataset (HPO-CLD). Please see the accompanying technical report (HPO-CLD Technical Report) for more details. This dataset contains a subset of the data collected as part of a larger, international research effort to develop a commercial, AI "inference engine" to recognize and assess real-time mental effort (i.e., cognitive load) in virtual reality (VR). To that end, we have developed a scientifically validated solution that reliably predicts cognitive load in the general population (HP Reverb G2 Omnicept Edition) and as part of that effort, we are releasing this dataset for non-commercial use.

Physiological measures such as eye-tracking and cardiovascular data strongly correspond to different cognitive (mental) load levels of individuals and can help in inferring the cognitive load. This dataset was created to support researchers and engineers interested in using physiological data to understand human behavior and cognition.

HPO-CLD contains sensor data collected from VR-HMDs retrofitted with eye-tracking and pulse plethysmography (PPG) sensors while participants were immersed in VR performing tasks of varying difficulty that required different amounts cognitive effort to complete. The data repository contains 100 folders with data corresponding to 100 users. The 100 users were randomly selected from an overall pool of 738 users in the data collection protocol. [**Insert consent info**].

Data includes: sensor data (Bitalino PPG for cardiac activity, Tobii for eye-tracking and pupillometry), task data (difficulty levels, performance), self-report data (NASA TLX scale of mental demands), and demographic data (e.g., age, education, ethnicity).

2. File Ordering and Descriptions

The main unzipped file contains 100 folders corresponding to sensor data of 100 users and a license file. Each folder contains the files as listed below with the highlighted naming conventions:

1. **Participant_number**-tobii **date**.csv
2. **Participant_number**-bitalino **date**.csv
3. **Participant_number** -testlog **date**.csv
4. **Participant_number** -**date-**event_log.csv
5. **Participant_number-**Demographics.csv
6. **Participant_number**-labels.csv

1. **Participant_number**-tobii **date**.csv

- Contains the timestamps and raw data from the Tobii eye tracking sensor. The sensor output has 28 columns and brief descriptions of columns are given in appendix (6a).
- Multiple Tobii files inside a folder should be concatenated together to maintain chronology of the timestamps.

2. **Participant_number**-bitalino **date**.csv
   - Contains timestamps and raw data from the Bitalino Photoplethysmography single-channel sensor. Columns are named A1 through A6 with A1 carrying the sensor data along with a separate column for timestamp. Please refer appendix (6b) for more details.
   - Some folders include two Bitalino sensor data – one sensor at the forehead and one sensor at the index finger. E.g. Bitalino-1, Bitalino-2.
   - Multiples files of the same Bitalino type (e.g. Bitalino-1) should be concatenated together to maintain chronology of the timestamps.

3. **Participant_number** -testlog **date**.csv
   - Logs of transactions and communications that occurred within the system during the data collection protocol (includes timestamp information).
   - Note: If 'pandas read_csv method' is used to read this file the error_bad_lines parameter should be set to false to ensure proper loading. E.g. pd.read_csv(path, error_bad_lines=False).

4. **Participant_number**-**date-**event_log.csv
   - Tracks the events that occurred throughout the protocol. These include task label, (i.e., low, medium or high cognitive load), participant's self-reported mental load (i.e., TLX scored from 0 - 21), along with timestamp, and other variables (described below).

5. **Participant_number-**Demographics.csv
   - Includes time and date of test session, participant's age, ethnicity, ancestry information, and other variables (described below).

6. **Participant_number**-labels.csv
   - These data were used as "ground truth" in our modeling (i.e., the labels our models predicted). This file contains the participant ID*, time start, time end, task difficulty (for classification), and self-reported TLX information (for regression). Please refer to appendix (6c) for more details.

3. Labeling Information

- The raw sensor data corresponding to different states of cognitive load can be aligned with the help of **Participant_number**-labels.csv file inside every folder.

- Each task is marked by a time_start and time_end pair that can be matched with the time stamps of raw sensor data to select data between the time interval specified.

*Note: The original Participant IDs have been replaced with randomly generated HP-CLD IDs as an added level of security for research participants.

4. Experiment Setting and Procedure

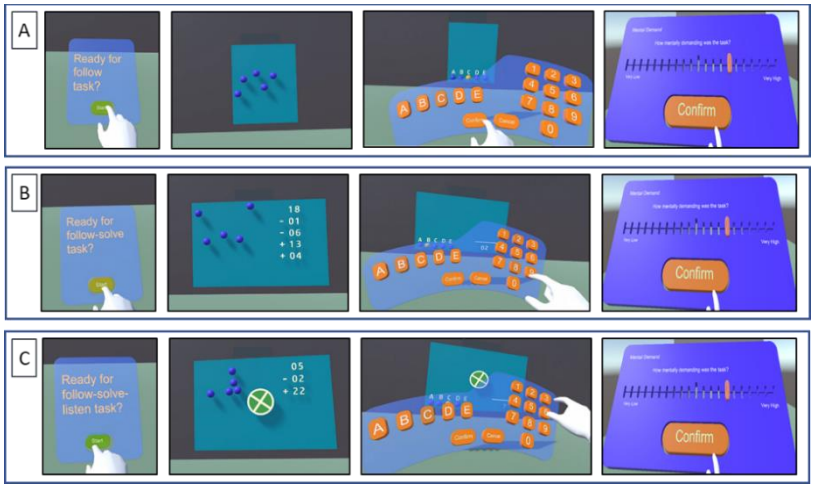

Figure 1. Experiments inducing varied levels of Cognitive Load

Figure 1. represents an Example stimulus from the cognitive load task. On a given trial, participants complete tasks designed to induce different levels of mental effort (low, medium, and high). **Box A** is an overview of a low mental effort trial. In a low trial, five dots move in a random, diverging pattern around the screen. Participants' task is to track a single dot and, when the dots stop moving, indicate which of dots they were tracking. **Box B** is an example of a medium mental effort trial in which participants complete the dot tracking task and a mental math task where numbers and operators appear at random intervals and they are instructed to do the math in their heads. When the dots stop moving and the numbers stop appearing on the screen, participants report which dot they were tracking and their answer to the mental math problem. **Box C** is an example of a high mental effort trial in which participants complete the dot tracking, mental math, and a third, vigilance task, in which they monitor a spinning wheel that changes directions. When they hear a beep, the wheel stops spinning, and they report which direction the wheel was spinning (while simultaneously tracking the dots and doing the math problem). In all three conditions, at the end of each trial, participants then self-report how mentally demanding they found the task on a continuous rating scale from very low to very high

(TLX score). Data from the two sensors were recorded simultaneously as the participant engaged in the described experiment.

5. Appendix

a. The file with 'tobii' contains data from the tobii eye-tracking sensor. The raw data columns and brief descriptions are given below. For more information, please refer http://developer.tobiipro.com/python/python-sdk-reference-guide.html.

Column	Description
t	System aligned timestamp when the data point was recorded.
device_time_stamp	Isolated timestamp of the sensor.
right_pupil_validity	The validity of the right pupil data. 1 if valid. 0 if invalid.
left_pupil_validity	The validity of the left pupil data. 1 if valid. 0 if invalid.
right_gaze_direction_validity	The validity of the right gaze direction data. 1 if valid. 0 if invalid.
left_gaze_direction_validity	The validity of the left gaze direction data. 1 if valid. 0 if invalid.
right_pupil_position_validity	The validity of the right pupil position data. 1 if valid. 0 if invalid.
left_pupil_position_validity	The validity of the left pupil position data. 1 if valid. 0 if invalid.
right_gaze_origin_validity	The validity of the right gaze origin data. 1 if valid. 0 if invalid.
left_gaze_origin_validity	The validity of the left gaze origin data. 1 if valid. 0 if invalid.
right_pupil_diameter	The diameter of the right pupil in millimeters.
left_pupil_diameter	The diameter of the left pupil in millimeters.
right_gaze_origin_position_in_hmd_coordinates_x	The x coordinate that describes the right gaze origin in mm.
right_gaze_origin_position_in_hmd_coordinates_y	The y coordinate that describes the right gaze origin in mm.
right_gaze_origin_position_in_hmd_coordinates_z	The z coordinate that describes the right gaze origin in mm.
left_gaze_origin_position_in_hmd_coordinates_x	The x coordinate that describes the left gaze origin in mm.
left_gaze_origin_position_in_hmd_coordinates_y	The y coordinate that describes the left gaze origin in mm.

left_gaze_origin_position_in_hmd_coordinates_z	The z coordinate that describes the left gaze origin in mm.
right_gaze_direction_unit_vector_x	The x component of unit vector that describes the right gaze direction in 3D.
right_gaze_direction_unit_vector_y	The y component of unit vector that describes the right gaze direction in 3D.
right_gaze_direction_unit_vector_z	The z component of unit vector that describes the right gaze direction in 3D.
left_gaze_direction_unit_vector_x	The x component of unit vector that describes the left gaze direction in 3D.
left_gaze_direction_unit_vector_y	The y component of unit vector that describes the left gaze direction in 3D.
left_gaze_direction_unit_vector_z	The z component of unit vector that describes the left gaze direction in 3D.
right_pupil_position_in_tracking_area_x	The x component of normalized 2D coordinates that describes the right pupil's position in the HMD's tracking area.
right_pupil_position_in_tracking_area_y	The y component of normalized 2D coordinates that describes the right pupil's position in the HMD's tracking area.
left_pupil_position_in_tracking_area_x	The x component of normalized 2D coordinates that describes the left pupil's position in the HMD's tracking area.
left_pupil_position_in_tracking_area_y	The y component of normalized 2D coordinates that describes the left pupil's position in the HMD's tracking area.

b. The file with 'bitalino' contains data from the Bitalino PPG sensor. The raw data columns and brief descriptions are given below.

Column	Descriptions
t	System aligned timestamp when the data point was recorded
A1	Data channel representing raw pulse information
A2	Unused channel, noise
A3	Unused channel, noise
A4	Unused channel, noise
A5	Unused channel, noise
A6	Unused channel, noise

c. The file with 'labels' contains ground truth information and time intervals of tasks. Brief descriptions are given below:

Column	Description
Participant ID	ID of the participant during data collection
time_start	Time boundary marking the beginning of task
time_end	Time boundary marking the end of task
task_difficulty	Difficulty of the task inducing cognitive load in accordance
TLX_score	Participant's self-reported difficulty score on a scale of 0 - 21
CRR	Correct Response Rate – agreement of CLX rating with task difficulty
reweighted_TLX_score	Combined individual TLX rating with population wide TLX task averages using a weighted sum

d. License

HP Omnicept Cognitive Load Dataset]

NonCommercial License

HP Inc. ("HP") is providing the HP Omnicept Cognitive Load Dataset and any corresponding documentation ("Dataset") for Your personal evaluation. By exercising the Licensed Rights (defined below), You accept and agree to be bound by the terms and conditions of this NonCommercial License ("License"). To the extent this License may be interpreted as a contract, HP grants You the Licensed Rights in consideration of Your acceptance of these terms and conditions, and HP grants You such rights in consideration of benefits that HP receives from making the Dataset available under these terms and conditions.

1. Definitions.

2. *Adapted Material* means material subject to Copyright and Similar Rights that is derived from or based upon the Dataset and in which the Dataset is translated, altered, arranged, transformed, or otherwise modified in a manner requiring permission under the Copyright and Similar Rights held by HP. For purposes of this License, where the Dataset is a musical work, performance, or sound recording, Adapted Material is always produced where the Dataset is synched in timed relation with a moving image.

3. *Adapter's License* means the license You apply to Your Copyright and Similar Rights in Your contributions to Adapted Material in accordance with the terms and conditions of this License.

4. *Copyright and Similar Rights* means copyright and/or similar rights closely related to copyright including, without limitation, performance, broadcast, sound recording, and Sui Generis Database Rights, without regard to how the rights are labeled or categorized. For purposes of this License, the rights specified in Section 2(b)(1)-(2) are not Copyright and Similar Rights.

5. *Licensed Rights* means the rights granted to You subject to the terms and conditions of this License, which are limited to all Copyright and Similar Rights that apply to Your use of the Dataset and that HP has authority to license.

6. *Licensor* means the individual(s) or entity(ies) granting rights under this License.

7. *NonCommercial* means not primarily intended for or directed towards commercial advantage or monetary compensation. For purposes of this License, the exchange of the Dataset for other material subject to Copyright and Similar Rights by digital file-sharing or similar means is NonCommercial provided there is no payment of monetary compensation in connection with the exchange.

8. *Sui Generis Database Rights* means rights other than copyright resulting from Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases, as amended and/or succeeded, as well as other essentially equivalent rights anywhere in the world.

9. *You* means the individual or entity exercising the Licensed Rights under this License. *Your* has a corresponding meaning.

10. Scope.

11. License grant.

12. Subject to the terms and conditions of this License, HP hereby grants You a worldwide, royalty-free, non-sublicensable, non-exclusive, revocable license to exercise the Licensed Rights in the Dataset to:

13. reproduce the Dataset, in whole or in part, for NonCommercial purposes only; and

14. produce and reproduce Adapted Material for NonCommercial purposes only.

15. No endorsement. Nothing in this License constitutes or may be construed as permission to assert or imply that You are, or that Your use of the Dataset is, connected with, or sponsored, endorsed, or granted official status by, HP or others designated to receive attribution as provided in Section 3(a)(1)(A)(i).

16. Usage Limitations.

17. You may not share or otherwise redistribute the Dataset

18. HP has carefully produced the Dataset to deidentify any individual whose information was used to create the Dataset. YOU MUST NOT ATTTEMPT TO REIDENTIFY THOSE INDIVIDUALS. DOING SO RESULTS IN AUTOMATIC TERMINATION OF THE LICENSE

19. You may not disclose Your evaluation of the Datasets to any third party, unless You include the following citation in such disclosure:

> Siegel, E.H., Wei, J., Gomes, A., Oliviera, M., Sundaramoorthy, P., Smathers, K., Vankipuram, M., Ghosh, S., Bailenson, J., & Ballagas, R. "HP Omnicept Cognitive Load Database (HPO-CLD) – Developing a Multimodal Inference Engine for Detecting Real-time Mental Workload in VR" Technical Report, HP Labs, Palo Alto, CA.

20. Other rights.

21. Moral rights, such as the right of integrity, are not licensed under this License, nor are publicity, privacy, and/or other similar personality rights; however, to the extent possible, HP waives and/or agrees not to assert any such rights held by HP to the limited extent necessary to allow You to exercise the Licensed Rights, but not otherwise.

22. Patent and trademark rights are not licensed under this License.

23. To the extent possible, HP waives any right to collect royalties from You for the exercise of the Licensed Rights, whether directly or through a collecting society under any voluntary or waivable statutory or compulsory licensing scheme. In all other cases HP expressly reserves any right to collect such royalties, including when the Dataset is used other than for NonCommercial purposes.

24. Disclaimer of Warranties and Limitation of Liability.

25. Unless otherwise separately undertaken by HP, to the extent possible, HP offers the Dataset as-is and as-available, and makes no representations or warranties of any kind concerning the Dataset, whether express, implied, statutory, or other. This includes, without limitation, warranties of title, merchantability, fitness for a particular purpose, non-infringement, absence of latent or other defects, accuracy, or the presence or absence of errors, whether or not known or discoverable. Where disclaimers of warranties are not allowed in full or in part, this disclaimer may not apply to You.

26. To the extent possible, in no event will HP be liable to You on any legal theory (including, without limitation, negligence) or otherwise for any direct, special, indirect, incidental, consequential, punitive, exemplary, or other losses, costs, expenses, or damages arising out of this License or use of the Dataset, even if HP has been advised of the possibility of such losses, costs, expenses, or damages. Where a limitation of liability is not allowed in full or in part, this limitation may not apply to You.

27. The disclaimer of warranties and limitation of liability provided above shall be interpreted in a manner that, to the extent possible, most closely approximates an absolute disclaimer and waiver of all liability.

28. Term and Termination.

29. Unless revoked earlier by HP, this License applies for the term of the Copyright and Similar Rights licensed here. However, if You fail to comply with this License, then Your rights under this License terminate automatically.

30. For the avoidance of doubt, this Section 6(b) does not affect any right HP may have to seek remedies for Your violations of this License.

31. For the avoidance of doubt, HP may also offer the Dataset under separate terms or conditions or stop distributing the Dataset at any time; however, doing so will not terminate this License.

32. Sections 1, 3, 4, 5, and 6 survive any revocation or termination of this License.

33. Other Terms and Conditions.

34. HP shall not be bound by any additional or different terms or conditions communicated by You unless expressly agreed.

35. Any arrangements, understandings, or agreements regarding the Dataset not stated herein are separate from and independent of the terms and conditions of this License.

36. Interpretation.

37. For the avoidance of doubt, this License does not, and shall not be interpreted to, reduce, limit, restrict, or impose conditions on any use of the Dataset that could lawfully be made without permission under this License.

38. To the extent possible, if any provision of this License is deemed unenforceable, it shall be automatically reformed to the minimum extent necessary to make it enforceable. If the provision

cannot be reformed, it shall be severed from this License without affecting the enforceability of the remaining terms and conditions.

39. No term or condition of this License will be waived and no failure to comply consented to unless expressly agreed to by HP.

40. Nothing in this License constitutes or may be interpreted as a limitation upon, or waiver of, any privileges and immunities that apply to HP or You, including from the legal processes of any jurisdiction or authority.