Supplementary Information

1 Methods

1.1 Participants' sample size calculation

Previous studies reported correlations between IAcc and HEP amplitude during the HTT, HDT, and HCT as .28, -.75, and .45, respectively. Based on these values, the required sample sizes were 97, 11, and 36 participants, respectively, to detect the given effect size with the specified alpha 0.05 and 80% of power [1–3] .Correlations ranging from .6 to .36 between IAcc in the HCT and HDT required sample sizes between 20 and 58 participants [4,5]. Finally, to detect a correlation of .8 between IAcc in HCT and HTT required a sample size of nine participants [6].

1.2 Exclusion criteria for participants' recruitment and selection

The exclusion criteria were as follows: presence of arterial hypertension (systolic blood pressure ≥ 140 mmHg and/or diastolic blood pressure ≥ 90 mmHg) based on office, home, or ambulatory blood pressure monitoring, presence of arrhythmias documented by ECG or Holter monitoring (no more than 200 ventricular or supraventricular extrasystoles), organic cardiac pathology (previous myocardial infarction, cardiomyopathies of various etiologies, scarring of unknown etiology, congenital heart defects, etc.), obstructive sleep apnea syndrome, significant atherosclerosis of central or peripheral arteries (arterial stenosis of 50% or more), mental disorders and pathology of the nervous system, thyroid dysfunction, systemic and/or autoimmune diseases, significant pathology of the liver, kidneys, or lungs, epilepsy, head trauma within the past year, administration of drugs passing through the blood-brain barrier, complications resulting from previous viral/infectious diseases, endocrine disorders (diabetes, obesity (>30 kg/m2), etc.).

Before the final inclusion participants had to (1) complete the Hospital Anxiety and Depression Scale (HADS)[7] those with the score >11 were not included in the further study and (2) undergo 24-h Holter and blood pressure monitoring to confirm the absence of arrhythmias and hypertension. Participants received no monetary compensation, but were provided with the results of the screening medical examination.

1.3 Technical details of heartbeat discrimination test implementation and its quality criteria

The implementation of the HDT was carried out using Python 3 programming language, the Pylsl package to capture the amplifier signal by the recording device. The algorithm accumulated resting-state ECG data and used the neurokit2 library to find the 75% of the average R peak amplitude, which served as the threshold for further real-time detection of R peaks. The Multiprocessing and Multithreading packages were used to enable parallel accumulation of physiological data, timestamps, and events from both the presentation computer and recording device. This setup allowed for the isolation of the ECG signal from the main data stream and real-time R peak detection. Each streamed sample was compared to the threshold, and when the threshold was exceeded, a sound with a specified delay was delivered. The Pywave package was used to generate the sound signal. The other two tasks were also implemented in Python, utilizing the Psychopy, tkinter, and numpy packages.

We assessed the quality of the biofeedback by examining two criteria: (1) whether the difference between the R peak detected by the presentation script and the actual R peak was within a 50 ms range, and (2) whether there were any pauses between tones in the sequence. Trials that did not meet these criteria were excluded from the set of 40 trials. Only participants with more than 10 synchronous and non-synchronous trials remaining were included in further analysis [8].

1.4 EMG data analysis

To account for the time required to execute the press, we corrected the button press time recorded by the keyboard using EMG data. EMG data analysis was performed using the neurokit2 package [9], applying the Teager-Kaiser Energy operator to detect movement onset thresholds, as implemented by the biosppy method. First, we found EMG onsets within an average reaction time of 250 ms before button presses [10] and determined the participant's mean time difference between the onset and the actual press. If no EMG onsets were detected near the button presses (14% of presses across all participants), the individual mean time difference was subtracted from the button press time. Mean time difference of presses across all participants was 140 ± 6 ms.

1.5 ICA components selection

We ensured the exclusion of ICA components by controlling their power spectral densities, signal correlation with the ECG and EOG signals, and their topographies. For ECG components we also looked at their averaging time-locked to R peaks. The topography of the EOG components were left-to-right oriented with positive values on one side and negative values on the other and up-down with either positive or negative values only. As a referent scalp topography we chose a topography of a clear in time ECG component from one of the participants. We selected two EOG and one ECG components that had the highest correlation with EOG and ECG signals, had similar scalp topographies to the referent topography, and in the case of the ECG component had a prominent R peak and QRS complex morphology in R-locked components averaging.

1.6 Behavioral tasks analysis

1.6.1 HTT

The *delay-based* IAcc was calculated using equation (1), where the Total of Correct Answers represented the number of presses that fell within a specified time window after the nearest preceding R peak, and Recorded Heartbeats represented the total number of R peaks recorded during the task. We compared the interval between each R peak and the nearest subsequent press. If the interval was shorter than the specified time window, the press was labeled as correct; otherwise, the press was labeled as incorrect. The time windows were set based on the participant's average heart rate during the experiment: 750 ms for heart rates below 69.76 bpm, 600 ms for heart rates between 69.75 and 94.25 bpm, and 400 ms for heart rates above 94.25 bpm.

$$delay\_based = 1 - \frac{Recorded\ Heartbeats - Total\ of\ Correct\ Answers}{Recorded\ Heartbeats} \quad (1)$$

The *mSI* IAcc was assessed using the equation (1) where Total of the Correct Answers was replaced by the total of all the presses.

To calculate *md* IAcc, the ECG signal was divided into overlapping 10-s windows, starting from each R peak. Within each window, inter-response intervals (IRI) and inter-beat intervals (IBI) were estimated in seconds. If the variation (calculated as the ratio of the standard deviation to the mean) of the IBI in a window was less than or equal to 0.5, the difference between the response frequency (1/IRI) and cardiac frequency (1/IBI) was computed. The *md* IAcc was then derived as 1 minus the mean of these differences.

The *d_mod* IAcc with logarithmic correction [11] was calculated using equation (2), where z() represents the inverse of the cumulative distribution function of the standard normal distribution. According to signal detection theory (SDT) [12], the HTT test can be represented as a two-alternative forced choice task between "signal" and "noise". The "signal" was defined as a specific time window following the R peak, during which the participant should feel a heartbeat, while the "noise" was the rest of the interval until the next peak. The total number of correct presses that fell within this window, as determined by the *delay_based IAcc* (see above), were classified as hits, while presses that did not fall within the window were classified as false alarms.

$$d\_mod = z\left(\frac{hits+0.5}{Recorded\ Heartbeats+1}\right) - z\left(\frac{false\ alarms+0.5}{Recorded\ Heartbeats+1}\right) (2)$$

We calculated the time delay between each button press and the nearest preceding R peak to assess *resVec* IAcc. This time delay was then normalized by the corresponding inter-beat interval (IBI) for that R peak. The IBI was considered as a full circle (2π radians), and the ratio of the time delay to the IBI was multiplied by 2π to determine the position on the circular distribution. Next, we used the circvar() function from the astropy.stats Python library to calculate the circular variation. The result was subtracted from 1 to obtain the *resVec* value, where values closer to 0 indicated high variation of pressing time, and values closer to 1 indicated low variation of pressing time.

CAcmotor IAcc was calculated with equation (1) where Total of Correct Answers was replaced by the number of presses done within the 350-650 ms window after the R peak.

1.6.2 HDT

In our analysis, we applied the ratio of correct answers (*ncorrect*) and the d-prime index with log-linear correction (*d*) using equation (3). Additionally, we used the proportion correct (*Pc2IFC*) to convert the range of *d* to 0-1, following the formula $\Phi(d/\sqrt{2})$, where the cumulative distribution function Φ was applied [13].

$$d = z\left(\frac{hit+0.5}{hit+miss+1}\right) - z\left(\frac{false\ alarm+0.5}{false\ alarm+correct\ rejection+1}\right) (3)$$

where z() - inverse of the cumulative distribution function of the standard normal distribution; hit, miss - number of correct and incorrect responses in S250 respectively; correct rejection, false alarm - number of correct

and incorrect responses in S550 respectively. The participant's response was considered correct if it matched the type of the condition.

Criterion (*c*) was determined using equation (4).

$$c = -0.5\left(z\left(\frac{hit+0.5}{hit+miss+1}\right) + z\left(\frac{false\ alarm+0.5}{false\ alarm+correct\ rejection+1}\right)\right) \quad (4)$$

1.6.3 HCT

For the HCT test Shandry index (*SI*) was calculated using the equation (5), where Reported Heartbeats were participants' answers about the number of heartbeats felt per interval.

$$SI = \frac{1}{6}\sum\left(1 - \frac{|Recorded\ Heartbeats - Reported\ Heartbeats|}{Recorded\ Heartbeats}\right)$$

Also we applied correction (*corSI*) to take into account the number of counted heartbeats was much higher than the actual value by including (Recorded Heartbeats+Reported Heartbeats)/2 in the denominator [14].

1.7 Spatio-temporal permutation test on HEP comparison

Test solved the problem of multiple comparisons in point-to-point comparisons within evoked data between conditions/groups. First, the data was randomly assigned to one of the conditions/groups, then data were compared and t values were calculated. Second, from the points with a t-value above the threshold, a cluster with the highest sum of t-values was selected. Then steps 1-2 were repeated 1000 times to generate a distribution of t values to test the null hypothesis. The null hypothesis was rejected if the t value of the observed cluster (formed based on the adjacency matrix) was within 5% of the most extreme values of the distribution. For within group comparison we used mne.stats.spatio_temporal_cluster_1samp_testt function with paired t-test (ttest_1samp_no_p) as a stat_fun parameter and threshold calculated as t.ppf(1 - alpha / 2, df) where alpha was .05, df = N-1, N was a number of participants. For between group comparison we used one-way ANOVA (f_oneway) from scipy.stats Python library as a stat_fun parameter and threshold calculated as f.ppf(1 - alpha / 2, dfn = dfn, dfd = dfd) where alpha was .05, dfn = Nc-1, dfd = N-Nc, Nc was a number of conditions, N was a number of participants.

1.8 Spatio-temporal permutation test on correlation between HEP amplitude and IAcc

Spearman's correlation convetration to t-statistics was performed with rho*(sqrt(N-2))./sqrt((1-rho.^2)) derived from the formula t^2 = DF*R^2 / (1-R^2), where DF was equal to N-2 with N was a number of participants. Channel adjacency matrix was prepared by 'distance' method with a 8 minimum neighborhood distance. The test was performed with 1000 permutations with two-tailed 'ft_statfun_correlationT' as statistic mode, cluster alpha was .05, 'maxsum' method to correct for multiple comparisons and 'nonparametric_individual' method for single-sample threshold was chosen, p-value threshold was set to .05.

2 Discussion on montage

We should note that during the recordings, we were constrained by the specific montage configuration provided by the amplifier's manufacturer, which we could not alter at the time. Despite this limitation, our participant recruitment necessitated proceeding with the recordings. We believe that, although this setup is somewhat unconventional, our data remains valuable and worth reporting for the following reasons:

1. We explicitly describe our reference scheme, allowing interested researchers to replicate our findings by offline re-referencing their EEG setups.

2. Significant differences cannot arise solely from specific montages unless genuine differences exist initially.

3. The uniqueness of our data offers new information to the community investigating heart-brain interactions, even when obtained with an unusual montage, which we clearly describe.

4. While mastoids provide a more neutral reference point, researchers in the EEG community often use even bipolar montages, which inherently contain differences between two active points.

We discussed the current montage and outlined its associated limitations through two approaches: first, by qualitatively comparing our HEP data with findings from the literature, and second, by empirically examining the effect of the reference on HEP amplitude in a newly recruited subgroup of participants.

A methodological review by Coll et al. (2021) of a large number of studies on HEP revealed variability in the methods used for HEP processing and recording[15]. Researchers selected approaches based on the specific requirements of their study and their prior experience. In our study, we applied high- and low-pass filtering, baseline correction, and CFA artifact removal, following the most commonly used procedures reported in the literature. The choice of reference scheme complicated the comparison of HEP waveform and amplitude. Yoris et al. (2017) suggested that methodological heterogeneity, particularly in the number of electrodes used and the approach to removing the CFA artifact, might have influenced the characteristics of HEP deflections[16]. However, several studies[17–21] reported negative deflections in HEP amplitude within frontal regions, which were consistent with our findings (Fig. 3 in the submitted article).

Previous research noted that it is hard to find entirely neutral reference location in the body[22]. This complicated the selection of a reference and led to the use of various referencing montages. Lei X and Liao K (2017) demonstrated that specific reference schemes were preferable for interpreting EEG components generated by the visual, somatomotor, and other brain networks[23]. In our study, we adopted a montage that was more commonly used in clinical research. Given this methodological constraint, in our fundamental study, we refrained from making inferences regarding the brain networks involved in the investigated HEP paradigms.

To further assess the impact of referencing schemes, we conducted an additional study involving n=20 volunteers, comprising both healthy participants and individuals with cardiological conditions (for details on

inclusion criteria, see Limonova et al. (2024)[24]). Resting-state EEG was recorded for five minutes using the same procedure as in the submitted article but with a monopolar montage. We performed the same processing steps as in the submitted manuscript however, noisy epochs were not removed. To evaluate the effect of referencing schemes we performed two analyses which differed in the references: (1) T3 and T4 references as in the submitted manuscript, (2) a common average reference, which is often used in basic research[16]. Comparison of mean HEP amplitudes within the 200–600 ms time window in the six regions of interest revealed no significant differences (Fig. S5). The data and code generated for this additional study are available upon request.

3 References

1. Fittipaldi, S. *et al.* A multidimensional and multi-feature framework for cardiac interoception. *Neuroimage* **212**, (2020).

2. Katkin, E. S., Cestaro, V. L. & Weitkunat, R. Individual differences in cortical evoked potentials as a function of heartbeat detection ability. *International Journal of Neuroscience* **61**, 269–276 (1991).

3. Mai, S., Wong, C. K., Georgiou, E. & Pollatos, O. Interoception is associated with heartbeat-evoked brain potentials (HEPs) in adolescents. *Biol Psychol* **137**, 24–33 (2018).

4. Knoll, J. F. & Hodapp, V. A Comparison between Two Methods for Assessing Heartbeat Perception. *Psychophysiology* **29**, 218–222 (1992).

5. Garfinkel, S. N. *et al.* Discrepancies between dimensions of interoception in autism: Implications for emotion and anxiety. *Biol Psychol* **114**, 117–126 (2016).

6. Körmendi, J., Ferentzi, E. & Köteles, F. A heartbeat away from a valid tracking task. An empirical comparison of the mental and the motor tracking task. *Biol Psychol* **171**, (2022).

7. Zigmond, A. S. & Snaith, R. P. The Hospital Anxiety and Depression Scale. *Acta Psychiatr Scand* **67**, 361–370 (1983).

8. Palser, E. R., Fotopoulou, A., Pellicano, E. & Kilner, J. M. The link between interoceptive processing and anxiety in children diagnosed with autism spectrum disorder: Extending adult findings into a developmental sample. *Biol Psychol* **136**, 13–21 (2018).

9. Makowski, D. *et al.* NeuroKit2: A Python toolbox for neurophysiological signal processing. *Behav Res Methods* **53**, 1689–1696 (2021).

10. Jain, A., Bansal, R., Kumar, A. & Singh, K. A comparative study of visual and auditory reaction times on the basis of gender and physical activity levels of medical first year students. *Int J Appl Basic Med Res* **5**, 124 (2015).

11. Hautus, M. J. Corrections for extreme proportions and their biasing effects on estimated values ofd′. *Behavior Research Methods, Instruments, & Computers* **27**, 46–51 (1995).

12. Macmillan, N. A. & Creelman, C. D. *Detection Theory*. (Psychology Press, 2004). doi:10.4324/9781410611147.

13. Dai, H. & Micheyl, C. A general formula for computing maximum proportion correct scores in various psychophysical paradigms with arbitrary probability distributions of stimulus observations. *Atten Percept Psychophys* **77**, 1448–1460 (2015).

14. Garfinkel, S. N., Seth, A. K., Barrett, A. B., Suzuki, K. & Critchley, H. D. Knowing your own heart: Distinguishing interoceptive accuracy from interoceptive awareness. *Biol Psychol* **104**, 65–74 (2015).

15. Coll, M. P., Hobson, H., Bird, G. & Murphy, J. Systematic review and meta-analysis of the relationship between the heartbeat-evoked potential and interoception. *Neuroscience and Biobehavioral Reviews* vol. 122 190–200 Preprint at https://doi.org/10.1016/j.neubiorev.2020.12.012 (2021).

16. Yoris, A. *et al.* The inner world of overactive monitoring: neural markers of interoception in obsessive–compulsive disorder. *Psychol Med* **47**, 1957–1970 (2017).

17. Canales-Johnson, A. *et al.* Auditory Feedback Differentially Modulates Behavioral and Neural Markers of Objective and Subjective Performance When Tapping to Your Heartbeat. *Cerebral Cortex (New York, NY)* **25**, 4490 (2015).

18. García-Cordero, I. *et al.* Feeling, learning from and being aware of inner states: Interoceptive dimensions in neurodegeneration and stroke. *Philosophical Transactions of the Royal Society B: Biological Sciences* **371**, (2016).

19. Pollatos, O., Kirsch, W. & Schandry, R. Brain structures involved in interoceptive awareness and cardioafferent signal processing: A dipole source localization study. *Hum Brain Mapp* **26**, 54–64 (2005).

20. Gray, M. A. *et al.* A cortical potential reflecting cardiac function. *Proc Natl Acad Sci U S A* **104**, 6818–6823 (2007).

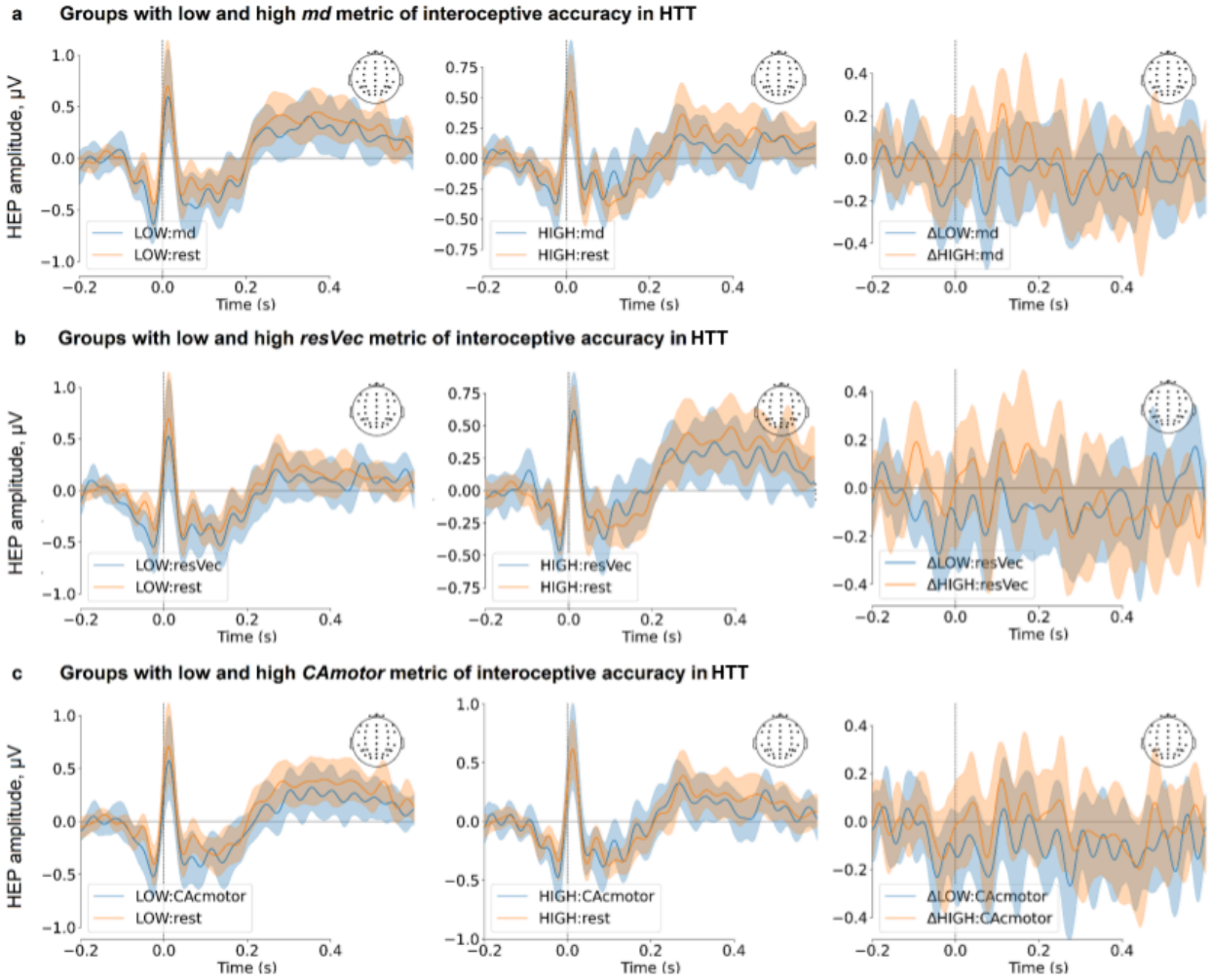
21. Kumral, D. *et al.* Attenuation of the Heartbeat-Evoked Potential in Patients With Atrial Fibrillation. *JACC Clin Electrophysiol* (2022) doi:10.1016/J.JACEP.2022.06.019.

22. Nunez, P. L. REST: A good idea but not the gold standard. *Clinical Neurophysiology* **121**, 2177–2180 (2010).

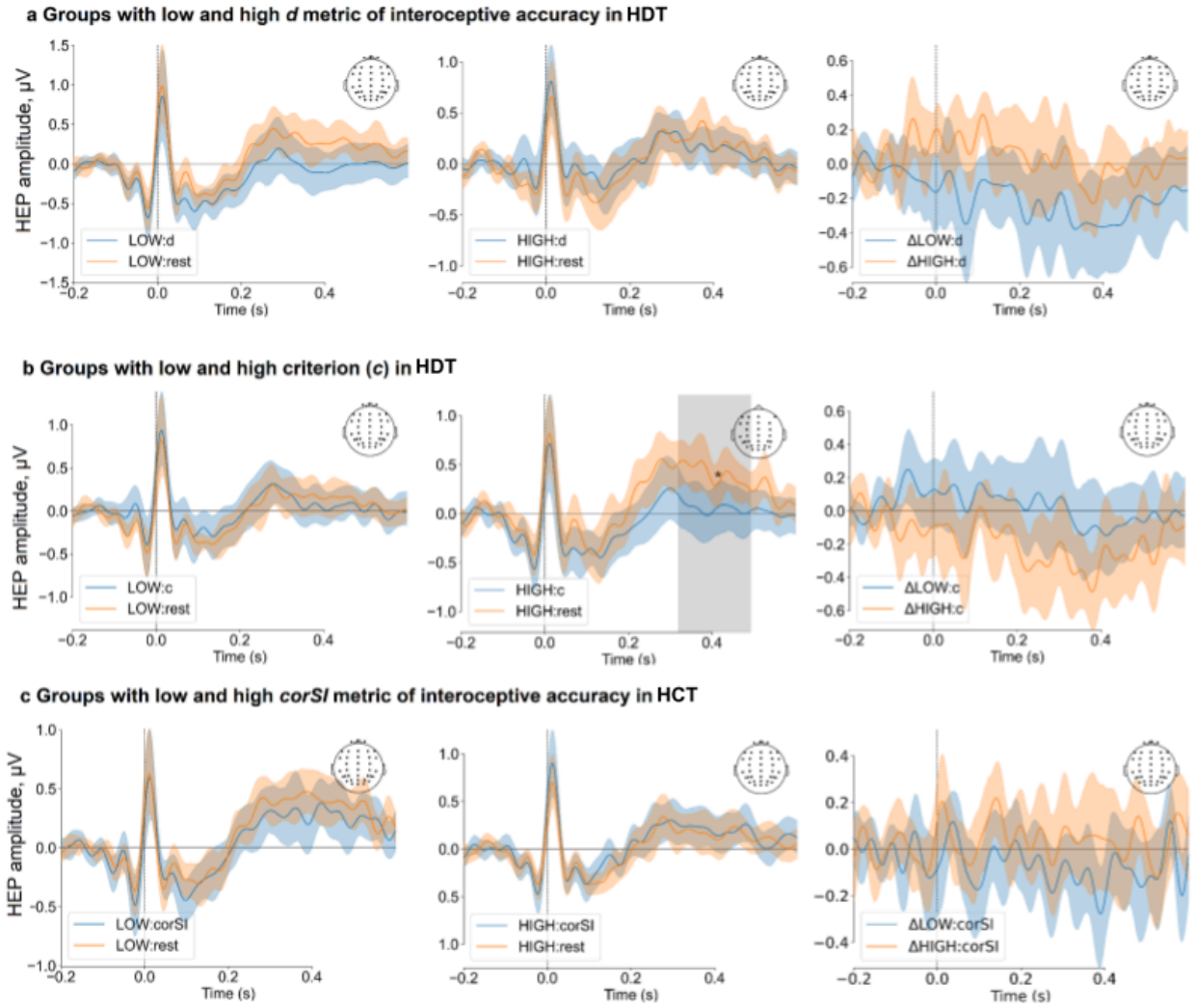
23. Lei, X. & Liao, K. Understanding the Influences of EEG Reference: A Large-Scale Brain Network Perspective. *Front Neurosci* **11**, (2017).

24. Limonova, A. S. *et al.* Exploring the Link Between Interoception and Symptom Severity in Premature Ventricular Contractions. *J Clin Med* **13**, 7756 (2024).

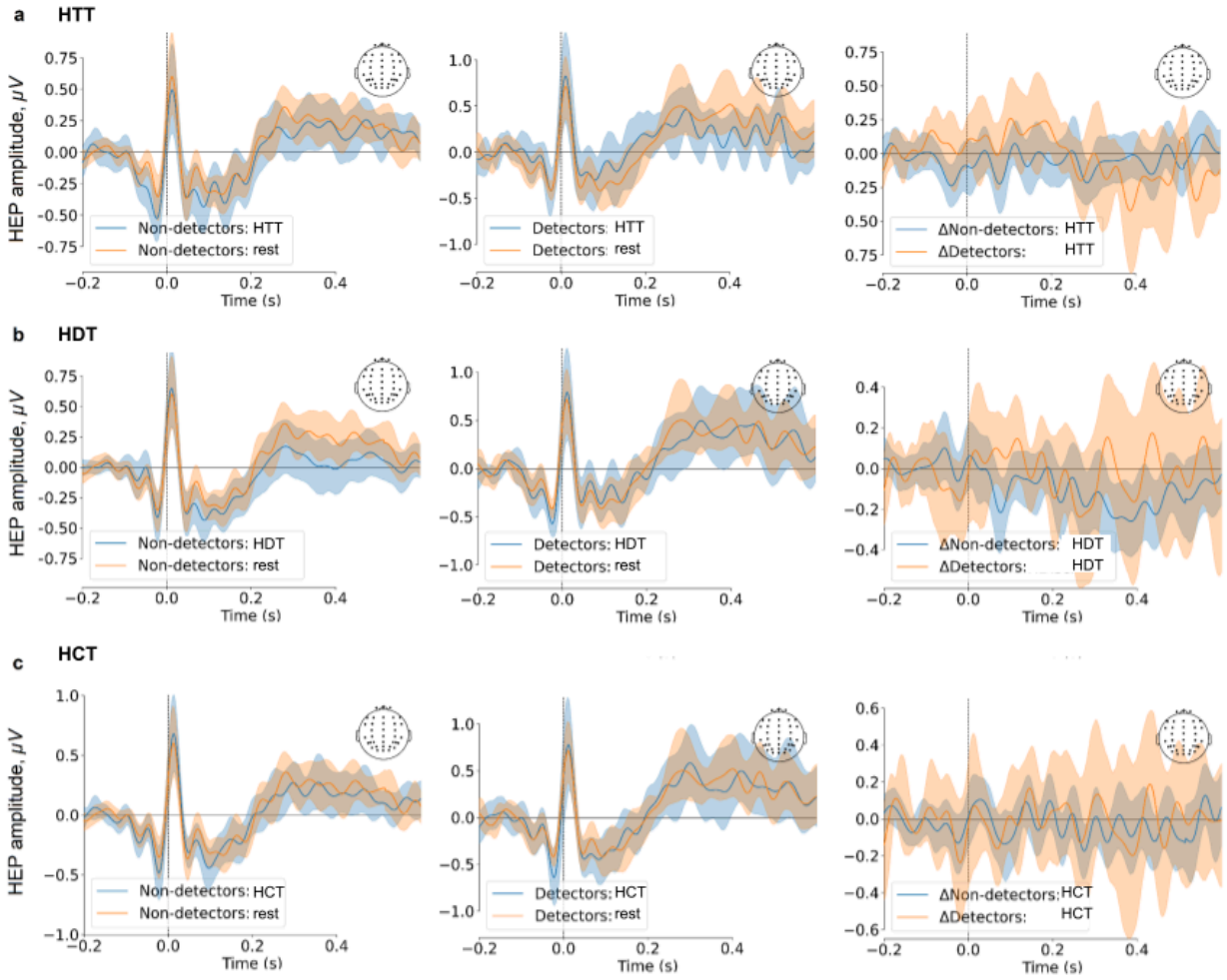
4 Tables and figures



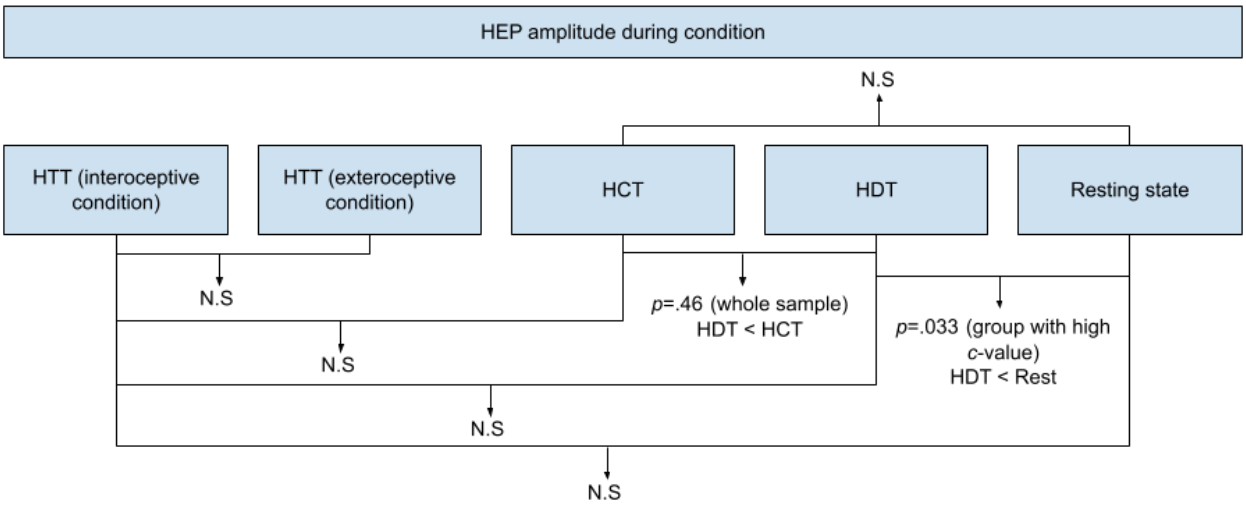
Supplementary Figure S1. Grand averages of HEP averaged over all channels for groups with high (HIGH) and low (LOW) interoceptive accuracy (IAcc) splitted based on the median of the following IAcc metrics - (a) *md*, (b) *resVec*, (c) *CAcmotor* in the HTT. Left figures show HEP amplitude during tasks and resting state within the group with low IAcc. Central - HEP amplitude within the group with high IAcc. Right - HEP amplitude modulation in groups with high (ΔHIGH) and low (ΔLOW) IAcc i.e HEP amplitude during tasks from which HEP amplitude during resting state was subtracted. The 95% confidence band is drawn.



Supplementary Figure S2. Grand averages of HEP averaged over all channels for groups with high (HIGH) and low (LOW) IAcc splitted based on the median of the following IAcc metrics - (a) *d* in the HDT, *corSI* (c) in the HCT, and with high and low characteristics of answers in the HDT test splitted based on the median *c* (b). Left figures show HEP amplitude during tasks and resting state within the group with low values. Central - HEP amplitude within the group with high values. Right - HEP amplitude modulation in groups with high (ΔHIGH) and low (ΔLOW) values i.e HEP amplitude during tasks from which HEP amplitude during resting state was subtracted. The time at which a significant difference is observed is highlighted in gray. The 95% confidence band is drawn.



Supplementary Figure S3. Grand averages of HEP averaged over all channels for detectors and non-detectors groups in the (a) HTT, (b) HDT, and (c) HCT. Left figures show HEP amplitude during tasks and resting state within the non-detectors group. Central - HEP amplitude within the detectors group. Right - HEP amplitude modulation (Δ) in non-detectors and detectors groups i.e HEP amplitude during tasks from which HEP amplitude during resting state was subtracted. The 95% confidence band is drawn.



Supplementary Figure S4. Summary of HEP amplitude comparison within conditions within the whole sample and within groups with different levels of cardioception (detectors group, non-detectors group, six groups with low IAcc-metrics, six groups with high IAcc-metrics). p-values for significant effects are indicated with group in brackets; if no significance is reported, the effect was nonsignificant (N.S.) for all groups.

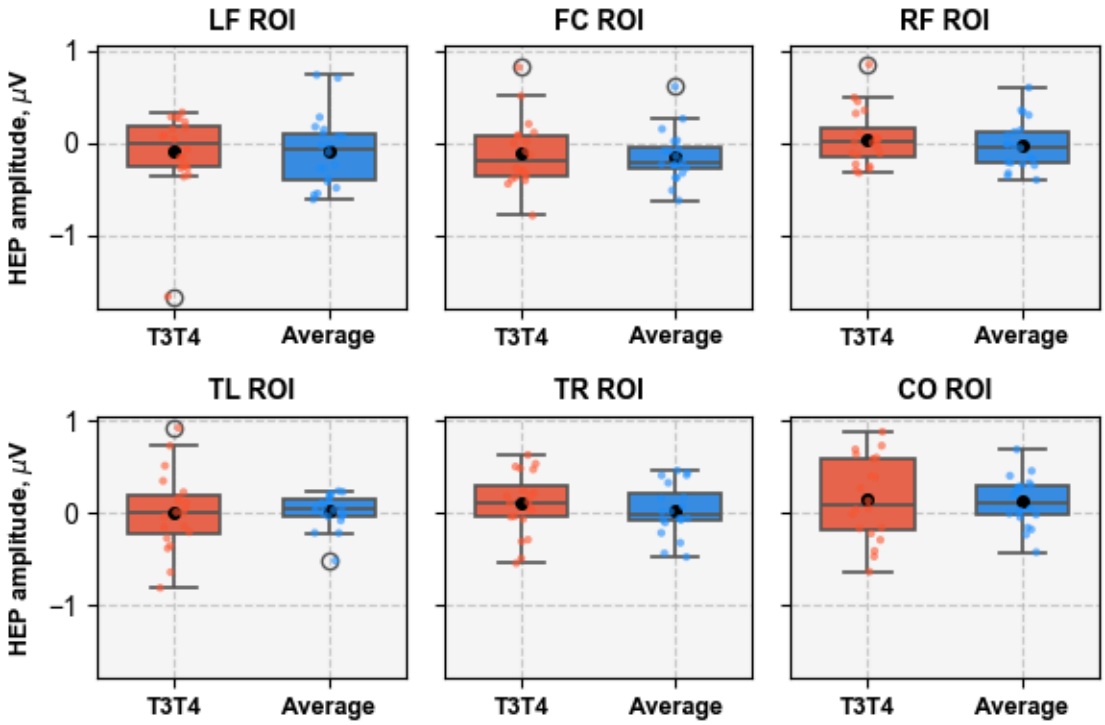


Figure S5. Mean HEP amplitude from 0,2 to 0,6 s averaged in LF (left frontal: Fp1, F3, FC3, C3, F7, FT7), FC (central frontal: Fpz, Fz, FCz, Cz), RF (right frontal: Fp2, F4, FC4, C4, F8, FT8), TL (left temporal: TP7, CP3, P3, T5, P5, PO7), CO (central occipital: CPz, Pz, POz, Oz, PO3, O1, PO4, O2), and TR (right temporal: TP8,

CP4, P4, T6, P6, PO8) regions of interest (ROI). T3T4 – montage with ipsilateral T3 and T4 reference, Average – common average reference.