Supplementary Material
A Comprehensive Analysis of the Influence of Cognitive Load on Physiological Signals in Virtual Reality

Category: Research

The supplementary material of this paper is composed of the following items:

- A video summarizing the design of our experiment and the data we captured.
- This PDF document.

This PDF contains additional information on the following topics:

- (S.1) Data Processing.
- (S.2) Physiological Markers Exploration.
- (S.3) Temporal Evolution of Signals.
- (S.4) Temporal Window Analysis.
- (S.5) Statistical Analysis.
- (S.6) Individual NASA-TXL Questionnaires.

S.1 Data Processing

To ensure replicability of our results, in this section we offer a detailed explanation of the data processing pipeline that was used to extract the physiological markers then analyzed. This includes the processing of the ECG, EDA, and respiration signals. Raw data from those signals is shown in Fig 1 for visualization purposes. Synchronization between the Shimmer ECG and Shimmer GSR+ units was achieved using their respective Unix timestamps. Although, as one of the limitations of our work, we did not extract markers from the gaze data, pupil data, or the PPG signal, we recorded them and included them in our dataset for completeness. The dataset offers eye data of the whole experiment recorded at 100Hz, comprising the eye forward vectors, head position and rotation, pupil diameter and openness, and validity of the measurement. Invalid measurements should be avoided as they mostly relate to blinks, and only constitute about 5% of the data. Previous studies have shown that pupillometry and the study of gaze behavior can produce valid indicators of CL variations [7, 9]. PPG can be used in combination with its synchronized ECG signal to estimate blood pressure, which has also proven to be related to activations of the Sympathetic System [3].

S.1.1 ECG

ECG signals were first resampled to 1000 Hz. R-wave detection was performed using a wavelet-based algorithm [8]. Using the R-wave occurrence time, the RR interval series were calculated, after ectopic removal and outlier correction through the Integral Pulse Frequency Modulation (IPFM) model [1].

S.1.2 EDA

EDA signals were analyzed using the convex optimization approach cvxEDA [4], robustly decomposing the signals into tonic and phasic components even under noisy conditions. Sudden skin conductance responses (sudomotor nerve activity, SMNA) were extracted from the phasic component, and specific biomarkers were computed: the normalized number of significant SMNA peaks (SMNA >1) and the spectral power density within the frequency range associated with sympathetic nervous activity (EDASymp, PSD range: 0.045–0.25 Hz) [5, 10].

S.1.3 Respiration

For estimating respiratory frequency, ECG-derived respiration (EDR) was obtained using a QRS slope-based method, which has been shown to be more effective than impedance pneumography for estimating respiratory rate in ambulatory conditions. Different features were extracted to derive the EDR signal. The R-wave amplitude was directly defined by the amplitude of the R-wave, while the R-wave angle was estimated based on the QRS slopes. Additionally, the Q-R slope was used by considering the upward slope of the R-wave as the EDR signal. A similar approach was applied to the R-S slope, which instead used a window centered around the sample with the steepest downward slope [12]. Respiratory frequency was then computed using a peakedness-based spectral averaging method and time-frequency analysis, providing instantaneous measures indicative of respiratory stability [6].

S.2 Physiological Markers Exploration

After processing the data recorded during our experiment, we extracted a broad number of physiological markers from the ECG and EDA signals. Although HR, RF, $EDA_{tonic}$and EDA_{phasic} were ultimately chosen for analysis in Sec. 5 of the main paper, in this section we present a visual exploration of all the different markers we extracted, which could be studied in depth as future work. Figure 3 shows markers related to heart rate, extracted from ECG, Figure 4 shows markers from EDA, and Figure 5 shows respiration-related markers. All of them are compared pairwise in scatter plots, together with the performance of each user, and subdivided in the different conditions by colors.

S.3 Temporal Evolution of Signals

In this section, we present the complete temporal evolution of the four markers selected for analysis: HR, RF, EDA_{tonic} and EDA_{phasic}. Figure 2 shows the mean of each of the metrics across all the participants for the whole four-minute duration of each segment.

S.4 Temporal Window Analysis

In Section 4.5, we analyzed the temporal evolution of the physiological markers by comparing their values at the beginning and end of each segment. We discussed the difference between

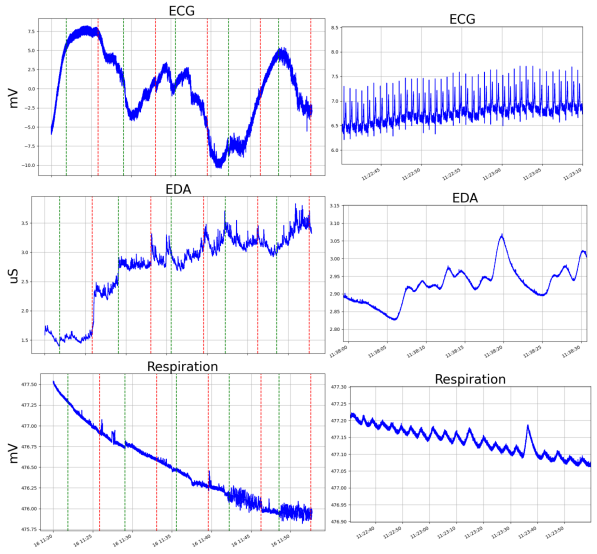

Figure 1: Examples of some of the raw signals provided in our dataset for ECG, EDA and respiration. Full signals for the whole experiment are present on the left, while close-ups of 30 seconds periods are shown on the right for better visualization. Timestamps for the start (green dotted lines) and end (red dotted lines) of each segments are also provided.

the initial (w_i) and final (w_f) temporal windows to assess how each marker evolved over time. To determine the appropriate length of the temporal window, we tested different durations ranging from 20 to 120 seconds, with 120 seconds corresponding to half of the segment. For each marker, we computed the relative change R as the difference between the mean value in the final temporal window w_f, minus the mean value in the initial temporal window w_i.

Figures 6 to 9 show the relative change for each metric and condition across different window durations. The relative change decreases as the window length increases, approaching zero. Approximately all markers stabilize at 40 seconds, showing a consistent difference between experimental conditions. Although most metrics show the largest differences at 20 seconds, we selected 40 seconds for analysis, as it offers a more stable estimate, reducing the impact of short-term variability.

S.5 Statistical Analysis

S.5.1 Performance

Given the within-subject design, we used a generalized linear mixed-effects model (GLMM) with CL and search area (S90 and S360) as fixed effects to analyze participants' performance, including two-way interactions, and participant and trial order as random effects. To stabilize variance, we applied a logarithmic transformation to the response variable. Post hoc tests were performed on estimated marginal means [11], with p-values corrected using the False Discovery Rate (FDR) method, which is appropriate for multifactorial designs involving complex contrasts and multiple comparisons, where maintaining statistical power is important [2]. The results can be seen in Tab. 3 and Tab. 4.

S.5.2 Global Physiological Markers Behavior

We used a generalized linear mixed-effects model (GLMM) with CL and search area (S90 and S360) as fixed effects, including two-way interactions, and participant and trial order as random effects, to account for within-subject variability and the repeated measures structure of our study. Post hoc tests were performed on estimated marginal means [11], with p-values corrected using the False Discovery Rate (FDR) method, which is appropriate for multifactorial designs involving complex contrasts and multiple comparisons, where maintaining statistical power is important [2]. The results presented in the main paper are summarized for easier visualization in Tab. 1 and 2.We perform a distinct model fitting for each physiological marker, showing the full results in Tab. 5 – 12.

S.5.3 Temporal Physiological Markers Behavior

We analyzed the temporal dependency by fitting for each physiologic marker a GLMM with CL, search area (S90 and S360), and period (w_i and w_f) as fixed effects, including again all two-way and three-way interactions, and participant and trial order as random effects. Post hoc tests were also performed on estimated marginal means [11], with p-values corrected using the False Discovery Rate (FDR) method [2]. The results can be found in Tab. 13 – 20.

S.6 Individual NASA-TXL Questionnaires

Figs. 12–14 show individual participant responses to the NASA-TLX questionnaire for each task: visual (finding as many objects as possible), auditory (detecting odd numbers in an audio recording), and audiovisual (combining both tasks). We normalized each score as explained in Sec. 4.2. For each task, participants rated six aspects: mental demand, frustration, physical demand, effort, and perceived performance. The figures illustrate the significant variability across participants' responses. We choose mental demand as our main aspect to analyze as it shows high correlation with the remaining rated aspects, as can be seen in Fig. 11, where we show the relationship between aspects and their distributions.

S.6.1 NASA-TLX Raw Scores

Fig. 10 shows the distribution of scores given in the mental demand category of the NASA-TLX questionnaire by users prior to normalization. Both distributions are clearly separated and present significant difference according to the statistical analysis presented in Sec. 4.2.

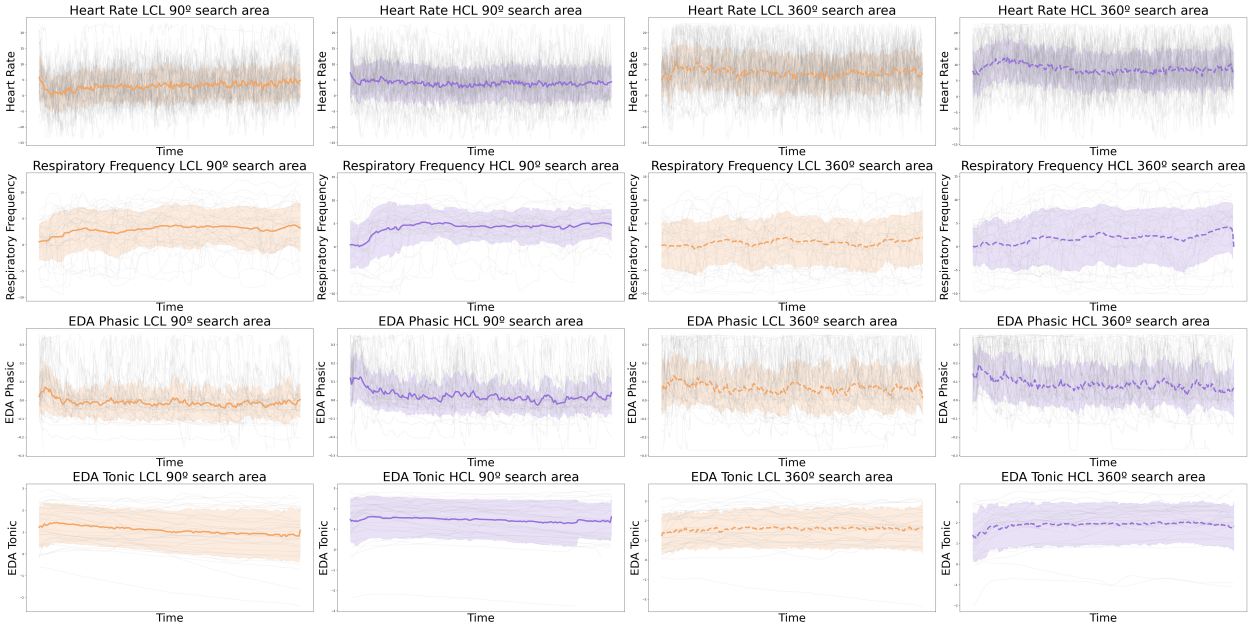

Figure 2: Temporal evolution for all four markers analyzed for each segment in the experiment. The x-axis shows the mean of the marker across all participants.

Table 1: GLMM results for physiological markers under varying cognitive load and search area conditions..

Marker	**Effect of CL**	**Effect of Area**
HR	$t = -2.58$, $\boldsymbol{p = 0.012}$, $\eta_p^2 = 0.07$	$t = -9.94$, $\boldsymbol{p < 0.0001}$, $\eta_p^2 = 0.51$
RF	$t = -2.60$, $\boldsymbol{p = 0.011}$, $\eta_p^2 = 0.07$	$t = +5.22$, $\boldsymbol{p < 0.0001}$, $\eta_p^2 = 0.22$
EDAtonic	$t = -4.52$, $\boldsymbol{p < 0.0001}$, $\eta_p^2 = 0.22$	$t = -6.01$, $\boldsymbol{p < 0.0001}$, $\eta_p^2 = 0.33$
EDAphasic	$t = -0.82$, $p = 0.416$, $\eta_p^2 = 0.01$	$t = -7.29$, $\boldsymbol{p < 0.0001}$, $\eta_p^2 = 0.39$

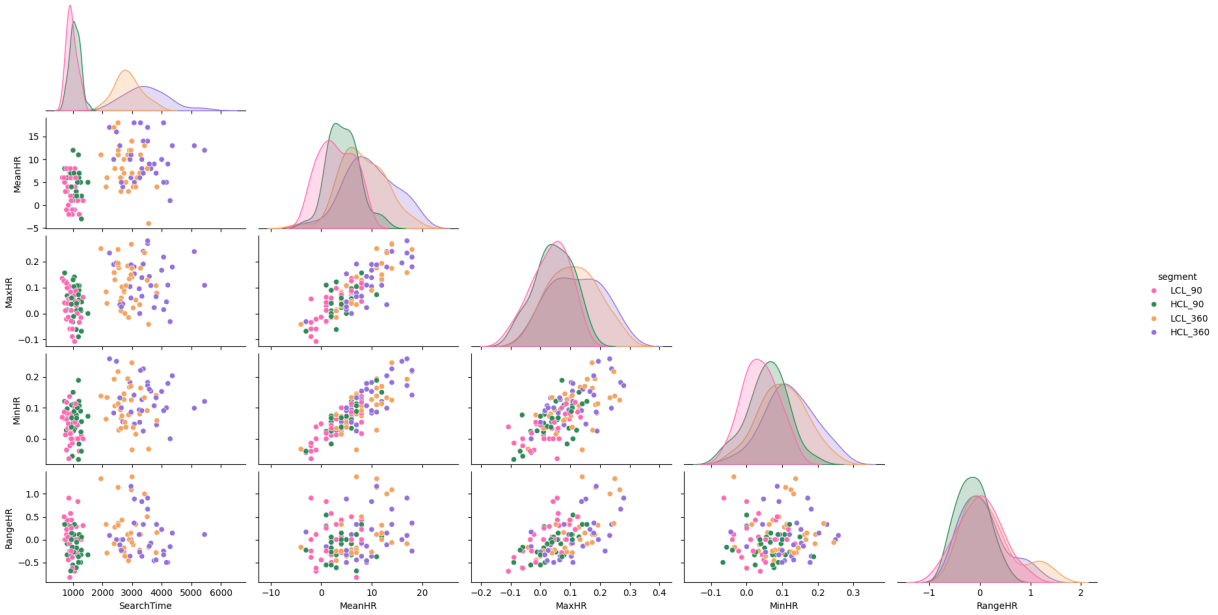

Figure 3: **Heart Rate Physiological Markers:** We present the distributions of Mean HR, Maximum HR, Minimum HR, and the HR range for each of the segments (pink for LCL 90º, green HCL 90º, orange LCL 360º and purple HCL 360º), as well as the visual search performance of each user for comparison. Scatter plots compare the data pairwise.

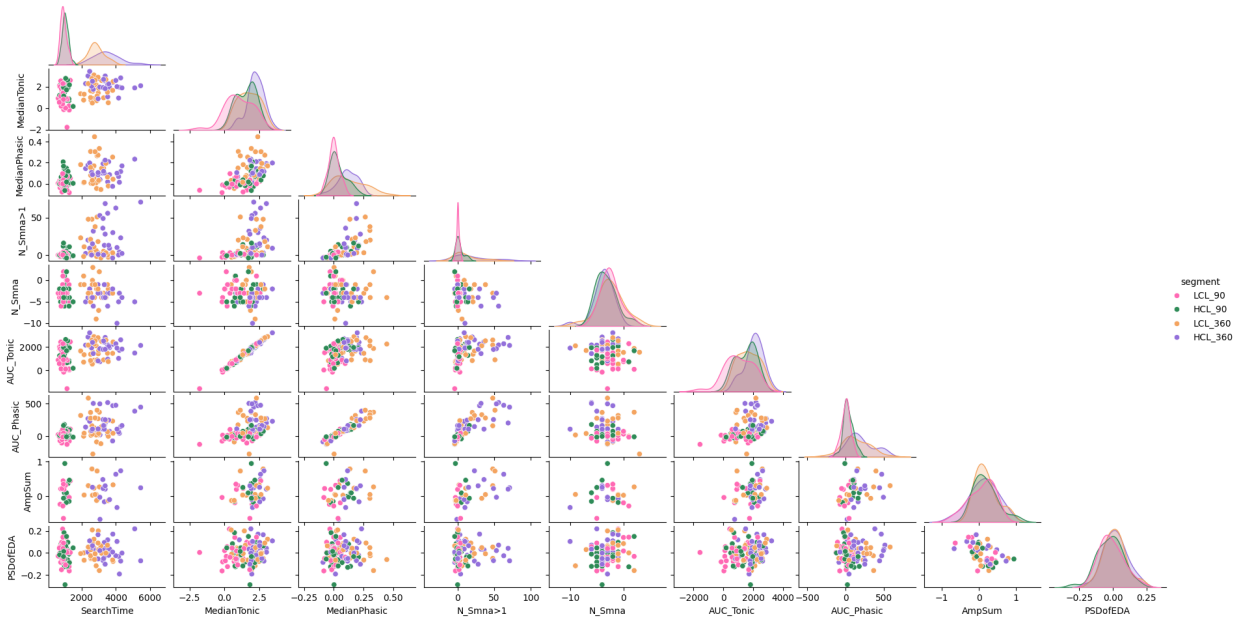

Figure 4: **Electrodermal Activity Physiological Markers:** We present the distributions of Median Tonic, Median Phasic, number of Sudomotor Nerve Activations (SMNA), number of significant SMNA (where amplitude of the peak > 1), AUC Tonic, AUC Phasic, accumulated amplitude of the SMNA>1, and the Power Spectral Density (PSD) of the EDA signal for each of the segments (pink for LCL 90º, green HCL 90º, orange LCL 360º and purple HCL 360º), as well as the visual search performance of each user for comparison. Scatter plots compare the data pairwise.

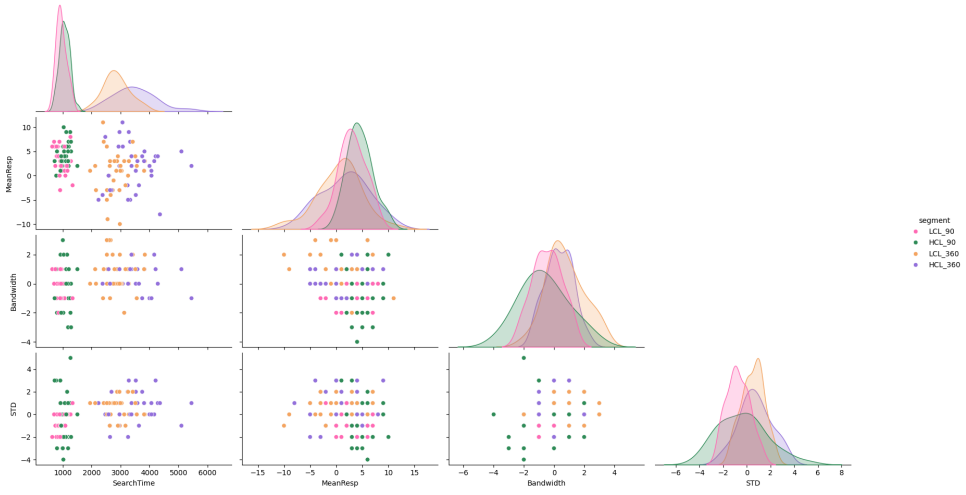

Figure 5: **Respiration Physiological Markers:** We present the distributions of Mean RF, Respiration Bandwidth, and the STD of the RF for each of the segments (pink for LCL 90º, green HCL 90º, orange LCL 360º and purple HCL 360º), as well as the visual search performance of each user for comparison. Scatter plots compare the data pairwise.

Table 2: Post-hoc comparisons between low (LCL) and high (HCL) cognitive load for each physiological marker within the two search area conditions. All comparisons are from GLMM estimated marginal means with FDR correction.

Marker	**Search Area**	**t-value (LCL vs. HCL)**	**p-value**	**Significant**
HR	90°	-1.568	0.120	No
	360°	-2.081	**0.0481**	Yes
RF	90°	-2.095	0.058	No
	360°	-1.570	0.120	No
EDAtonic	90°	-3.370	**0.002**	Yes
	360°	-3.009	**0.004**	Yes
EDAphasic	90°	-1.723	0.107	No
	360°	0.533	0.596	No

Table 3: **Performance GLMM**: We used a GLMM model with cognitive load (CL) and search area as fixed effects, including two-way interactions, and participant and order as random effects. We report the estimates, standard errors, degrees of freedom (df), t-values, and p-values, as well as variance and standard deviation for random effects, participant and order. Significant values are highlighted in bold.

	Estimate	Std. Error	df	t value	Pr($>\|t\|$)
(Intercept)	7.490	0.027	34	275.158	**<.0001**
CL	0.087	0.018	102	7.416	**<.0001**
Search Area	0.566	0.018	102	48.132	**<.0001**
CL:Search Area	0.007	0.018	102	0.621	0.536

Groups	Variance	Std. Dev.
Participant	0.022	0.149
Order	0.004	0.065
Residual	0.015	0.123

Number of obs: 140, groups: participant, 35; order, 4

Table 4: **Post-hoc Performance**: Post-hoc test for pair-wise comparison performed on estimated means using FDR for p-value adjustment. We report the estimates, standard errors (SE), degrees of freedom (df), t-ratio, and p-values. Significant values are highlighted in bold. We present all combinations of the two CL levels (HCL and LCL), and two search area ranges (S90 and S360). We additionally include the effect sizes (η_p^2) for the CL comparison in both search areas.

contrast	estimate	SE	df	t.ratio	p.value	η_p^2
HCL S360 - LCL S360	0.189	0.0293	99.0	6.437	**<.0001**	0.968
HCL S360 - HCL S90	1.144	0.0293	99.0	38.968	**<.0001**	—
HCL S360 - LCL S90	1.304	0.0293	99.0	44.481	**<.0001**	—
LCL S360 - HCL S90	0.955	0.0294	99.1	32.503	**<.0001**	—
LCL S360 - LCL S90	1.115	0.0294	99.1	37.942	**<.0001**	—
HCL S90 - LCL S90	0.161	0.0293	99.0	5.475	**<.0001**	0.418

Degrees-of-freedom method: kenward-roger

P value adjustment: fdr method for 6 tests

Table 5: **Mean HR GLMM**: We used a GLMM model with cognitive load (CL) and search area as fixed effects, including two-way interactions, and participant and order as random effects. We report the estimates, standard errors, degrees of freedom (df), t-values, p-values, and effect size (η_p^2) with 95% confidence intervals for fixed effects, as well as variance and standard deviation for random effects, participant and order. Significant values are highlighted in bold.

	Estimate	Std. Error	df	t value	Pr($>\|t\|$)	η_p^2 [95% CI]
(Intercept)	6.302	0.548	19.701	11.495	**<.0001**	—
CL	-0.670	0.260	94.816	-2.577	**0.0115**	0.070 [0.01, 1.00]
Search Area	-2.599	0.261	96.439	-9.943	**<.0001**	0.510 [0.39, 1.00]
CL:Search Area	0.084	0.260	95.017	0.322	0.748	0.001 [0.00, 1.00]

Groups	Variance	Std. Dev.
Participant	7.774	2.788
Order	0.041	0.201
Residual	9.158	3.026

Number of obs: 136, groups: participant, 35; order, 4

Table 6: **Post-hoc Mean HR**: Post-hoc test for pair-wise comparison performed on estimated means using FDR for p-value adjustment. We report the estimates, standard errors (SE), degrees of freedom (df), t-ratio, and p-values. Significant values are highlighted in bold. We present all combinations of the two CL levels (HCL and LCL), and two search area ranges (S90 and S360)

contrast	estimate	SE	df	t.ratio	p.value
S90 LCL - S360 HCL	-5.0308	0.7338	96.51	-6.856	**<.0001**
S90 LCL - S90 HCL	-1.1732	0.7484	95.64	-1.568	0.1203
S90 LCL - S360 HCL	-6.5390	0.7312	95.71	-8.943	**<.0001**
S360 LCL - S90 HCL	3.8576	0.7475	96.84	5.161	**<.0001**
S360 LCL - S360 HCL	-1.5082	0.7246	95.46	-2.081	**0.0481**
S90 HCL - S360 HCL	-5.3658	0.7456	96.26	-7.197	**<.0001**

Degrees-of-freedom method: kenward-roger

P value adjustment: fdr method for 6 tests

Table 7: **Mean RF GLMM**: We used a GLMM model with cognitive load (CL) and search area as fixed effects, including two-way interactions, and participant and order as random effects. We report the estimates, standard errors, degrees of freedom (df), t-values, p-values, and effect size (η_p^2) with 95% confidence intervals for fixed effects, as well as variance and standard deviation for random effects, participant and order. Significant values are highlighted in bold.

	Estimate	Std. Error	df	t value	Pr($>|t|$)	η_p^2 [95% CI]
(Intercept)	2.473	0.518	8.877	4.776	**0.0011**	—
CL	-0.648	0.249	97.092	-2.602	**0.0107**	0.070 [0.01, 1.00]
Search Area	1.302	0.250	97.570	5.220	**<.0001**	0.220 [0.11, 1.00]
CL:Search Area	-0.102	0.250	97.547	-0.408	0.684	0.002 [0.00, 1.00]

Groups	Variance	Std. Dev.
Participant	4.332	2.081
Order	0.329	0.574
Residual	8.451	2.907

Number of obs: 117, groups: participant, 33; order, 4

Table 8: **Post-hoc Mean RF**: Post-hoc test for pair-wise comparison performed on estimated means using FDR for p-value adjustment. We report the estimates, standard errors (SE), degrees of freedom (df), t-ratio, and p-values. Significant values are highlighted in bold. We present all combinations of the two CL levels (HCL and LCL), and two search area ranges (S90 and S360)

contrast	estimate	SE	df	t.ratio	p.value
S90 LCL - S360 LCL	2.4012	0.7047	97.09	3.408	**0.0019**
S90 LCL - S90 HCL	-1.4995	0.7156	97.43	-2.095	0.0581
S90 LCL - S360 HCL	1.3085	0.7021	96.60	1.864	0.0785
S360 LCL - S90 HCL	-3.9007	0.7097	97.13	-5.496	**<.0001**
S360 LCL - S360 HCL	-1.0927	0.6960	96.24	-1.570	0.1197
S90 HCL - S360 HCL	2.8080	0.7095	97.09	3.958	**0.0004**

Degrees-of-freedom method: kenward-roger

P value adjustment: fdr method for 6 tests

Table 9: **Mean EDA$_{tonic}$ GLMM**: We used a GLMM model with cognitive load (CL) and search area as fixed effects, including two-way interactions, and participant and order as random effects. We report the estimates, standard errors, degrees of freedom (df), t-values, p-values, and effect size (η_p^2) with 95% confidence intervals for fixed effects, as well as variance and standard deviation for random effects, participant and order. Significant values are highlighted in bold.

	Estimate	Std. Error	df	t value	Pr($>|t|$)	η_p^2 [95% CI]
(Intercept)	1.5742	0.1775	5.1758	8.870	**<.0001**	—
CL	-0.2389	0.0529	74.3820	-4.516	**<.0001**	0.22 [0.09, 1.00]
Search Area	-0.3176	0.0529	74.4317	-6.005	**<.0001**	0.33 [0.19, 1.00]
CL:Search Area	-0.0101	0.0530	74.4785	-0.191	0.8489	0.0000 [0.00, 1.00]

Groups	Variance	Std. Dev.
Participant	0.2563	0.5062
Order	0.0825	0.2872
Residual	0.3404	0.5834

Number of obs: 123, groups: participant, 32; order, 4

Table 10: **Post-hoc Mean EDA$_{tonic}$**: Post-hoc test for pair-wise comparison performed on estimated means using FDR for p-value adjustment. We report the estimates, standard errors (SE), degrees of freedom (df), t-ratio, and p-values. Significant values are highlighted in bold. We present all combinations of the two CL levels (HCL and LCL), and two search area ranges (S90 and S360)

contrast	estimate	SE	df	t.ratio	p.value
S90 LCL - S360 LCL	-0.6555	0.1478	86.17	-4.436	**0.0001**
S90 LCL - S90 HCL	-0.4982	0.1478	86.13	-3.370	**0.0017**
S90 LCL - S360 HCL	-1.1131	0.1511	86.72	-7.366	**<.0001**
S360 LCL - S90 HCL	0.1574	0.1483	85.22	1.061	0.2915
S360 LCL - S360 HCL	-0.4576	0.1521	85.90	-3.009	**0.0041**
S90 HCL - S360 HCL	-0.6150	0.1521	85.90	-4.044	**0.0002**

Degrees-of-freedom method: kenward-roger

P value adjustment: fdr method for 6 tests

Table 11: **Mean EDA$_{phasic}$ GLMM**: We used a GLMM model with cognitive load (CL) and search area as fixed effects, including two-way interactions, and participant and order as random effects. We report the estimates, standard errors, degrees of freedom (df), t-values, p-values, and effect size (η_p^2) with 95% confidence intervals for fixed effects, as well as variance and standard deviation for random effects, participant and order. Significant values are highlighted in bold.

	Estimate	Std. Error	df	t value	Pr($>|t|$)	η_p^2 [95% CI]
(Intercept)	0.0702	0.0111	28.6486	6.334	**<.0001**	—
CL	-0.0056	0.0069	81.3552	-0.817	0.416	0.008 [0.00, 1.00]
Search Area	-0.0503	0.0069	82.6143	-7.288	**<.0001**	0.390 [0.26, 1.00]
CL:Search Area	-0.0109	0.0069	82.2618	-1.583	0.117	0.030 [0.00, 1.00]

Groups	Variance	Std. Dev.
Participant	0.0024	0.0490
Order	0.0000	0.0000
Residual	0.0054	0.0737

Number of obs: 117, groups: participant, 32; order, 4

Table 12: **Post-hoc Mean EDA$_{phasic}$**: Post-hoc test for pair-wise comparison performed on estimated means using FDR for p-value adjustment. We report the estimates, standard errors (SE), degrees of freedom (df), t-ratio, and p-values. Significant values are highlighted in bold. We present all combinations of the two CL levels (HCL and LCL), and two search area ranges (S90 and S360).

contrast	estimate	SE	df	t.ratio	p.value
S90 LCL - S360 LCL	-0.1224	0.0193	81.93	-6.356	**<.0001**
S90 LCL - S90 HCL	-0.0330	0.0192	80.95	-1.723	0.1065
S90 LCL - S360 HCL	-0.1118	0.0202	82.68	-5.528	**<.0001**
S360 LCL - S90 HCL	0.0893	0.0188	80.73	4.744	**<.0001**
S360 LCL - S360 HCL	0.0106	0.0199	82.47	0.533	0.5956
S90 HCL - S360 HCL	-0.0787	0.0199	83.01	-3.950	**0.0002**

Degrees-of-freedom method: kenward-roger

P value adjustment: fdr method for 6 tests

Table 13: **Mean HR GLMM with $\mathbf{w}_{\pm 40}$**: We used a GLMM model with cognitive load (CL), search area, and period (beginning and final 40 seconds) as fixed effects, including two-way and three-way interactions, and participant and order as random effects. We report the estimates, standard errors, degrees of freedom (df), t-values, p-values, and effect size (η_p^2) with 95% confidence intervals for fixed effects, as well as variance and standard deviation for random effects, participant and order. Significant values are highlighted in bold.

	Estimate	Std. Error	df	t value	Pr($>|t|$)	η_p^2 [95% CI]
(Intercept)	4.5393	0.4452	25.17	10.20	**<.0001**	—
CL	-0.1497	0.1027	226.63	-1.46	0.1462	0.01 [0.00, 1.00]
Search Area	-0.5306	0.1028	227.05	-5.16	**<.0001**	0.10 [0.05, 1.00]
Period	0.7951	0.1027	226.61	7.74	**<.0001**	0.21 [0.14, 1.00]
CL:Search Area	-0.0462	0.1027	226.69	-0.45	0.6531	0.00 [0.00, 1.00]
CL:Period	-0.0073	0.1025	226.33	-0.07	0.9431	0.00 [0.00, 1.00]
Search Area:Period	-0.2576	0.1024	226.29	-2.51	**0.0126**	0.03 [0.00, 1.00]
CL:Search Area:Period	-0.0597	0.1027	226.62	-0.58	0.5621	0.00 [0.00, 1.00]

Groups	Variance	Std. Dev.
Participant	5.2499	2.2913
Order	0.1503	0.3877
Residual	2.8489	1.6879

Number of observations: 272, groups: Participant, 35; Order, 4

Table 14: **Post-hoc Mean HR with $\mathbf{w}_{\pm 40}$**: Post-hoc test for pair-wise comparison performed on estimated means using FDR for p-value adjustment. We report the estimates, standard errors (SE), degrees of freedom (df), t-ratio, and p-values. Significant values are highlighted in bold. We present all combinations of the two CL levels (HCL and LCL), two search area ranges (S90 and S360), and two periods (w_i and w_f).

contrast	estimate	SE	df	t.ratio	p.value
S90 LCL w_i - S360 LCL w_i	-1.7881	0.4177	227.62	-4.281	**0.0001**
S90 LCL w_i - S90 HCL w_i	-0.5258	0.4131	227.15	-1.273	0.2601
S90 LCL w_i - S360 HCL w_i	-1.8904	0.4105	227.30	-4.605	**<.0001**
S90 LCL w_i - S90 LCL w_f	0.9412	0.4094	227.00	2.299	**0.0392**
S90 LCL w_i - S360 LCL w_f	0.4220	0.4078	227.48	1.035	0.3675
S90 LCL w_i - S90 HCL w_f	0.6833	0.4073	227.29	1.678	0.1475
S90 LCL w_i - S360 HCL w_f	0.1103	0.4072	227.25	0.271	0.8071
S360 LCL w_i - S90 HCL w_i	1.2624	0.4203	227.55	3.004	**0.0064**
S360 LCL w_i - S360 HCL w_i	-0.1023	0.4184	227.69	-0.245	0.8071
S360 LCL w_i - S90 LCL w_f	2.7293	0.4177	227.62	6.534	**<.0001**
S360 LCL w_i - S360 LCL w_f	2.2101	0.4146	227.44	5.331	**<.0001**
S360 LCL w_i - S90 HCL w_f	2.4714	0.4155	227.84	5.948	**<.0001**
S360 LCL w_i - S360 HCL w_f	1.8985	0.4149	227.57	4.576	**<.0001**
S90 HCL w_i - S360 HCL w_i	-1.3647	0.4142	227.42	-3.295	**0.0027**
S90 HCL w_i - S90 LCL w_f	1.4669	0.4131	227.15	3.551	**0.0012**
S90 HCL w_i - S360 LCL w_f	0.9477	0.4112	227.52	2.305	**0.0392**
S90 HCL w_i - S90 HCL w_f	1.2090	0.4108	227.33	2.943	**0.0072**
S90 HCL w_i - S360 HCL w_f	0.6361	0.4110	227.41	1.548	0.1813
S360 HCL w_i - S90 LCL w_f	2.8316	0.4105	227.30	6.897	**<.0001**
S360 HCL w_i - S360 LCL w_f	2.3124	0.4071	227.22	5.679	**<.0001**
S360 HCL w_i - S90 HCL w_f	2.5737	0.4070	227.17	6.323	**<.0001**
S360 HCL w_i - S360 HCL w_f	2.0008	0.4069	227.09	4.917	**<.0001**
S90 LCL w_f - S360 LCL w_f	-0.5192	0.4078	227.48	-1.273	0.2601
S90 LCL w_f - S90 HCL w_f	-0.2579	0.4073	227.29	-0.633	0.5679
S90 LCL w_f - S360 HCL w_f	-0.8308	0.4072	227.25	-2.040	0.0700
S360 LCL w_f - S90 HCL w_f	0.2613	0.4039	227.21	0.647	0.5679
S360 LCL w_f - S360 HCL w_f	-0.3116	0.4037	227.13	-0.772	0.5145
S90 HCL w_f - S360 HCL w_f	-0.5730	0.4038	227.13	-1.419	0.2202

Degrees-of-freedom method: kenward-roger

P value adjustment: fdr method for 28 tests

Table 15: **Mean RF GLMM with $\mathbf{w}_{\pm 40}$**: We used a GLMM model with cognitive load (CL), search area, and period (beginning and final 40 seconds) as fixed effects, including two-way and three-way interactions, and participant and order as random effects. We report the estimates, standard errors, degrees of freedom (df), t-values, p-values, and effect size (η_p^2) with 95% confidence intervals for fixed effects, as well as variance and standard deviation for random effects, participant and order. Significant values are highlighted in bold.

	Estimate	Std. Error	df	t value	Pr($>|t|$)	η_p^2 [95% CI]
(Intercept)	2.3470	0.4914	7.98	4.78	**0.0014**	—
CL	-0.6919	0.2064	224.28	-3.35	**0.0009**	0.05 [0.01, 1.00]
Search Area	0.8031	0.2068	225.03	3.88	**0.0001**	0.06 [0.02, 1.00]
Period	-1.0898	0.2065	224.49	-5.28	**<.0001**	0.11 [0.05, 1.00]
CL:Search Area	-0.0751	0.2068	224.83	-0.36	0.7168	0.00 [0.00, 1.00]
CL:Period	0.3224	0.2064	224.13	1.56	0.1196	0.01 [0.00, 1.00]
Search Area:Period	0.2756	0.2068	225.11	1.33	0.1839	0.01 [0.00, 1.00]
CL:Search Area:Period	-0.2083	0.2063	224.04	-1.01	0.3138	0.00 [0.00, 1.00]

Groups	Variance	Std. Dev.
Participant	3.4595	1.8600
Order	0.3993	0.6319
Residual	11.3035	3.3621

Number of observations: 268, groups: Participant, 35; Order, 4

Table 16: **Post-hoc Mean RF with $\mathbf{w}_{\pm 40}$**: Post-hoc test for pair-wise comparison performed on estimated means using FDR for p-value adjustment. We report the estimates, standard errors (SE), degrees of freedom (df), t-ratio, and p-values. Significant values are highlighted in bold. We present all combinations of the two CL levels (HCL and LCL), two search area ranges (S90 and S360), and two periods (w_i and w_f).

contrast	estimate	SE	df	t.ratio	p.value
S90 LCL w_i - S360 LCL w_i	1.5908	0.8050	223.44	1.976	0.0864
S90 LCL w_i - S90 HCL w_i	-1.3056	0.8580	225.50	-1.522	0.1727
S90 LCL w_i - S360 HCL w_i	1.4186	0.8039	223.10	1.765	0.1222
S90 LCL w_i - S90 LCL w_f	-1.4000	0.8037	223.04	-1.742	0.1222
S90 LCL w_i - S360 LCL w_f	-0.0786	0.8204	224.74	-0.096	0.9237
S90 LCL w_i - S90 HCL w_f	-3.1622	0.8043	223.20	-3.932	**0.0006**
S90 LCL w_i - S360 HCL w_f	-2.3736	0.8245	224.16	-2.879	**0.0111**
S360 LCL w_i - S90 HCL w_i	-2.8963	0.8584	225.61	-3.374	**0.0027**
S360 LCL w_i - S360 HCL w_i	-0.1722	0.8043	223.21	-0.214	0.8946
S360 LCL w_i - S90 LCL w_f	-2.9908	0.8050	223.44	-3.715	**0.0010**
S360 LCL w_i - S360 LCL w_f	-1.6694	0.8176	223.97	-2.042	0.0791
S360 LCL w_i - S90 HCL w_f	-4.7530	0.8047	223.32	-5.907	**<.0001**
S360 LCL w_i - S360 HCL w_f	-3.9643	0.8257	224.49	-4.801	**<.0001**
S90 HCL w_i - S360 HCL w_i	2.7242	0.8581	225.51	3.175	**0.0048**
S90 HCL w_i - S90 LCL w_f	-0.0944	0.8580	225.50	-0.110	0.9237
S90 HCL w_i - S360 LCL w_f	1.2269	0.8712	225.96	1.408	0.2041
S90 HCL w_i - S90 HCL w_f	-1.8566	0.8576	225.38	-2.165	0.0629
S90 HCL w_i - S360 HCL w_f	-1.0680	0.8764	225.82	-1.219	0.2730
S360 HCL w_i - S90 LCL w_f	-2.8186	0.8039	223.10	-3.506	**0.0019**
S360 HCL w_i - S360 LCL w_f	-1.4972	0.8191	224.41	-1.828	0.1135
S360 HCL w_i - S90 HCL w_f	-4.5808	0.8043	223.20	-5.696	**<.0001**
S360 HCL w_i - S360 HCL w_f	-3.7922	0.8247	224.22	-4.598	**<.0001**
S90 LCL w_f - S360 LCL w_f	1.3214	0.8204	224.74	1.611	0.1521
S90 LCL w_f - S90 HCL w_f	-1.7622	0.8043	223.20	-2.191	0.0629
S90 LCL w_f - S360 HCL w_f	-0.9736	0.8245	224.16	-1.181	0.2788
S360 LCL w_f - S90 HCL w_f	-3.0836	0.8195	224.52	-3.763	**0.0010**
S360 LCL w_f - S360 HCL w_f	-2.2949	0.8377	224.16	-2.740	**0.0155**
S90 HCL w_f - S360 HCL w_f	0.7886	0.8247	224.21	0.956	0.3808

Degrees-of-freedom method: kenward-roger

P value adjustment: fdr method for 28 tests

Table 17: **Mean EDA$_{tonic}$ GLMM with w$_{\pm 40}$**: We used a GLMM model with cognitive load (CL), search area, and period (beginning and final 40 seconds) as fixed effects, including two-way and three-way interactions, and participant and order as random effects. We report the estimates, standard errors, degrees of freedom (df), t-values, p-values, and effect size (η_p^2) with 95% confidence intervals for fixed effects, as well as variance and standard deviation for random effects, participant and order. Significant values are highlighted in bold.

	Estimate	Std. Error	df	t value	Pr($>|t|$)	η_p^2 [95% CI]
(Intercept)	1.5608	0.1954	6.06	7.99	**0.0002**	—
CL	-0.2171	0.0382	203.35	-5.69	**<.0001**	0.14 [0.07, 1.00]
Search Area	-0.2813	0.0380	202.33	-7.40	**<.0001**	0.21 [0.14, 1.00]
period	0.0564	0.0379	202.07	1.49	0.139	0.01 [0.00, 1.00]
CL:Search Area	0.0092	0.0381	202.41	0.24	0.809	0.000 [0.00, 1.00]
CL:period	0.0307	0.0379	202.07	0.81	0.419	0.003 [0.00, 1.00]
Search Area:period	0.1346	0.0379	202.07	3.55	**0.0005**	0.06 [0.02, 1.00]
CL:Search Area:period	0.0092	0.0379	202.07	0.24	0.8080	0.000 [0.00, 1.00]

Groups	Variance	Std. Dev.
Participant	0.3727	0.6105
Order	0.1002	0.3166
Residual	0.3556	0.5963

Number of observations: 248, groups: Participant, 32; Order, 4

Table 18: **Post-hoc Mean EDA$_{tonic}$ with w$_{\pm 40}$**: Post-hoc test for pair-wise comparison performed on estimated means using FDR for p-value adjustment. We report the estimates, standard errors (SE), degrees of freedom (df), t-ratio, and p-values. Significant values are highlighted in bold. We present all combinations of the two CL levels (HCL and LCL), two search area ranges (S90 and S360), and two periods (w_i and w_f).

contrast	estimate	SE	df	t.ratio	p.value
S90 LCL w_i - S360 LCL w_i	-0.2566	0.1492	206.06	-1.720	0.1107
S90 LCL w_i - S90 HCL w_i	-0.3359	0.1507	206.52	-2.228	**0.0428**
S90 LCL w_i - S360 HCL w_i	-0.6662	0.1538	206.83	-4.332	**0.0001**
S90 LCL w_i - S90 LCL w_f	0.4618	0.1491	206.02	3.097	**0.0044**
S90 LCL w_i - S360 LCL w_f	-0.3701	0.1507	206.52	-2.456	**0.0260**
S90 LCL w_i - S90 HCL w_f	-0.0338	0.1507	206.52	-0.224	0.8228
S90 LCL w_i - S360 HCL w_f	-0.8656	0.1523	206.76	-5.682	**<.0001**
S360 LCL w_i - S90 HCL w_i	-0.0793	0.1506	206.47	-0.527	0.6452
S360 LCL w_i - S360 HCL w_i	-0.4096	0.1539	206.85	-2.662	**0.0157**
S360 LCL w_i - S90 LCL w_f	0.7183	0.1492	206.06	4.815	**<.0001**
S360 LCL w_i - S360 LCL w_f	-0.1136	0.1506	206.48	-0.754	0.5059
S360 LCL w_i - S90 HCL w_f	0.2228	0.1506	206.47	1.479	0.1711
S360 LCL w_i - S360 HCL w_f	-0.6091	0.1524	206.77	-3.996	**0.0003**
S90 HCL w_i - S360 HCL w_i	-0.3303	0.1548	206.38	-2.134	0.0501
S90 HCL w_i - S90 LCL w_f	0.7976	0.1507	206.52	5.292	**<.0001**
S90 HCL w_i - S360 LCL w_f	-0.0343	0.1515	206.04	-0.226	0.8228
S90 HCL w_i - S90 HCL w_f	0.3021	0.1515	206.02	1.994	0.0664
S90 HCL w_i - S360 HCL w_f	-0.5298	0.1533	206.28	-3.456	**0.0017**
S360 HCL w_i - S90 LCL w_f	1.1279	0.1538	206.83	7.336	**<.0001**
S360 HCL w_i - S360 LCL w_f	0.2960	0.1548	206.39	1.913	0.0762
S360 HCL w_i - S90 HCL w_f	0.6324	0.1548	206.38	4.086	**0.0002**
S360 HCL w_i - S360 HCL w_f	-0.1995	0.1555	206.15	-1.283	0.2344
S90 LCL w_f - S360 LCL w_f	-0.8319	0.1507	206.52	-5.521	**<.0001**
S90 LCL w_f - S90 HCL w_f	-0.4956	0.1507	206.52	-3.288	**0.0028**
S90 LCL w_f - S360 HCL w_f	-1.3274	0.1523	206.76	-8.713	**<.0001**
S360 LCL w_f - S90 HCL w_f	0.3363	0.1515	206.04	2.220	**0.0428**
S360 LCL w_f - S360 HCL w_f	-0.4955	0.1533	206.30	-3.231	**0.0031**
S90 HCL w_f - S360 HCL w_f	-0.8318	0.1533	206.28	-5.426	**<.0001**

Degrees-of-freedom method: kenward-roger

P value adjustment: fdr method for 28 tests

Table 19: **Mean EDA$_{phasic}$ GLMM with w$_{\pm 40}$**: We used a GLMM model with cognitive load (CL), search area, and period (beginning and final 40 seconds) as fixed effects, including two-way and three-way interactions, and participant and order as random effects. We report the estimates, standard errors, degrees of freedom (df), t-values, p-values, and effect size (η_p^2) with 95% confidence intervals for fixed effects, as well as variance and standard deviation for random effects, participant and order. Significant values are highlighted in bold.

	Estimate	Std. Error	df	t value	Pr($>|t|$)	η_p^2 [95% CI]
(Intercept)	1.023e-01	1.257e-02	8.95	8.14	**<.0001**	—
CL	-2.596e-02	6.792e-03	191.63	-3.82	**0.0002**	0.07 [0.02, 1.00]
Search Area	-6.715e-02	6.814e-03	193.90	-9.85	**<.0001**	0.33 [0.25, 1.00]
Period	4.699e-02	6.801e-03	192.61	6.91	**<.0001**	0.20 [0.12, 1.00]
CL:Search Area	4.979e-04	6.796e-03	192.10	0.07	0.9417	0.00 [0.00, 1.00]
CL:Period	-5.854e-03	6.801e-03	192.51	-0.86	0.3904	0.00 [0.00, 1.00]
Search Area:Period	-2.038e-02	6.788e-03	191.30	-3.00	**0.0030**	0.05 [0.01, 1.00]
CL:Search Area:Period	2.558e-03	6.788e-03	191.31	0.38	0.7067	0.00 [0.00, 1.00]

Groups	Variance	Std. Dev.
Participant	0.0024	0.0485
Order	0.0002	0.0123
Residual	0.0106	0.1028

Number of observations: 231, groups: Participant, 32; Order, 4

Table 20: **Post-hoc Mean EDA$_{phasic}$ with w$_{\pm 40}$**: Post-hoc test for pair-wise comparison performed on estimated means using FDR for p-value adjustment. We report the estimates, standard errors (SE), degrees of freedom (df), t-ratio, and p-values. Significant values are highlighted in bold. We present all combinations of the two CL levels (HCL and LCL), two search area ranges (S90 and S360), and two periods (w_i and w_f).

contrast	estimate	SE	df	t.ratio	p.value
S90 LCL w_i - S360 LCL w_i	-0.1690	0.0271	191.64	-6.226	**<.0001**
S90 LCL w_i - S90 HCL w_i	-0.0575	0.0276	192.14	-2.082	0.0542
S90 LCL w_i - S360 HCL w_i	-0.2387	0.0274	192.03	-8.712	**<.0001**
S90 LCL w_i - S90 LCL w_f	0.0466	0.0276	192.05	1.687	0.1134
S90 LCL w_i - S360 LCL w_f	-0.0510	0.0276	192.22	-1.846	0.0845
S90 LCL w_i - S90 HCL w_f	0.0023	0.0272	191.97	0.084	0.9331
S90 LCL w_i - S360 HCL w_f	-0.0871	0.0272	192.34	-3.204	**0.0028**
S360 LCL w_i - S90 HCL w_i	0.1114	0.0271	191.23	4.110	**0.0001**
S360 LCL w_i - S360 HCL w_i	-0.0697	0.0269	190.99	-2.596	**0.0158**
S360 LCL w_i - S90 LCL w_f	0.2156	0.0272	192.54	7.925	**<.0001**
S360 LCL w_i - S360 LCL w_f	0.1179	0.0271	191.13	4.350	**0.0001**
S360 LCL w_i - S90 HCL w_f	0.1712	0.0267	191.89	6.416	**<.0001**
S360 LCL w_i - S360 HCL w_f	0.0818	0.0267	191.57	3.068	**0.0041**
S90 HCL w_i - S360 HCL w_i	-0.1812	0.0273	190.95	-6.626	**<.0001**
S90 HCL w_i - S90 LCL w_f	0.1041	0.0277	192.54	3.766	**0.0005**
S90 HCL w_i - S360 LCL w_f	0.0065	0.0276	191.05	0.235	0.8442
S90 HCL w_i - S90 HCL w_f	0.0598	0.0271	191.68	2.204	**0.0423**
S90 HCL w_i - S360 HCL w_f	-0.0296	0.0272	192.73	-1.088	0.2993
S360 HCL w_i - S90 LCL w_f	0.2853	0.0275	193.19	10.391	**<.0001**
S360 HCL w_i - S360 LCL w_f	0.1877	0.0273	190.51	6.872	**<.0001**
S360 HCL w_i - S90 HCL w_f	0.2410	0.0270	192.63	8.936	**<.0001**
S360 HCL w_i - S360 HCL w_f	0.1516	0.0269	191.87	5.634	**<.0001**
S90 LCL w_f - S360 LCL w_f	-0.0977	0.0277	193.47	-3.524	**0.0011**
S90 LCL w_f - S90 HCL w_f	-0.0443	0.0272	191.79	-1.632	0.1216
S90 LCL w_f - S360 HCL w_f	-0.1338	0.0272	192.43	-4.915	**<.0001**
S360 LCL w_f - S90 HCL w_f	0.0533	0.0272	192.75	1.959	0.0687
S360 LCL w_f - S360 HCL w_f	-0.0361	0.0271	191.54	-1.330	0.2072
S90 HCL w_f - S360 HCL w_f	-0.0894	0.0267	191.56	-3.350	**0.0018**

Degrees-of-freedom method: kenward-roger

P value adjustment: fdr method for 28 tests

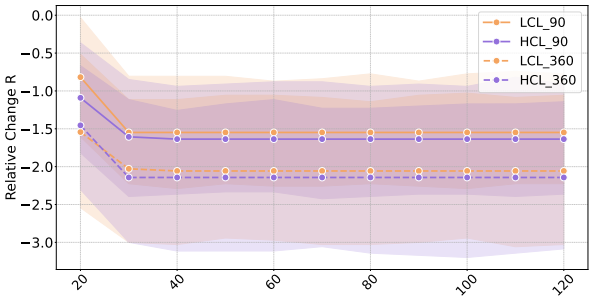

Figure 6: **Mean HR Relative Change:** We present the relative change in HR between the first and last temporal windows for each experimental condition. The x-axis indicates the duration of the window used, while the y-axis shows the relative change R. Conditions with high CL are represented in purple, whereas conditions with low CL are shown in orange. The search areas are distinguished by line styles: continuous lines for 90° and dashed lines for 360°.

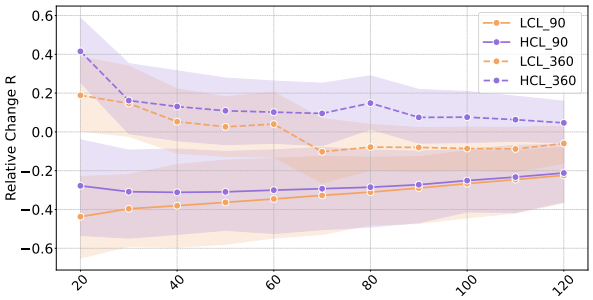

Figure 7: **Mean Median Tonic (EDA) Relative Change:** We present the relative change in median tonic between the first and last temporal windows for each experimental condition. The x-axis indicates the duration of the window used, while the y-axis shows the relative change R. Conditions with high CL are represented in purple, whereas conditions with low CL are shown in orange. The search areas are distinguished by line styles: continuous lines for 90° and dashed lines for 360°.

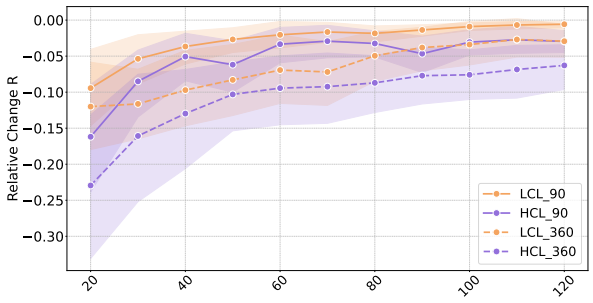

Figure 8: **Mean Median Phasic (EDA) Relative Change:** We present the relative change in median phasic between the first and last temporal windows for each experimental condition. The x-axis indicates the duration of the window used, while the y-axis shows the relative change R. Conditions with high CL are represented in purple, whereas conditions with low CL are shown in orange. The search areas are distinguished by line styles: continuous lines for 90° and dashed lines for 360°.

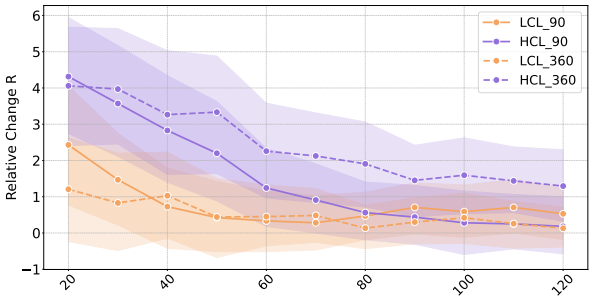

Figure 9: **Mean Respiratory Frequency Relative Change:** We present the relative change in respiratory frequency between the first and last temporal windows for each experimental condition. The x-axis indicates the duration of the window used, while the y-axis shows the relative change R. Conditions with high CL are represented in purple, whereas conditions with low CL are shown in orange. The search areas are distinguished by line styles: continuous lines for 90° and dashed lines for 360°.

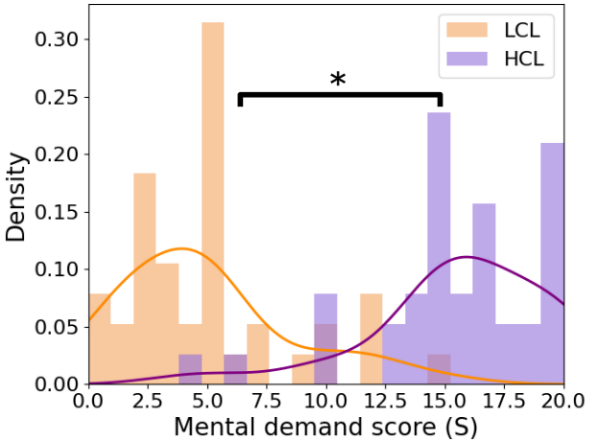

Figure 10: We present the distribution of the mental demand scores from the NASA-TLX questionnaire prior to normalization for the low (LCL, orange) and high cognitive load (HCL, purple) conditions.

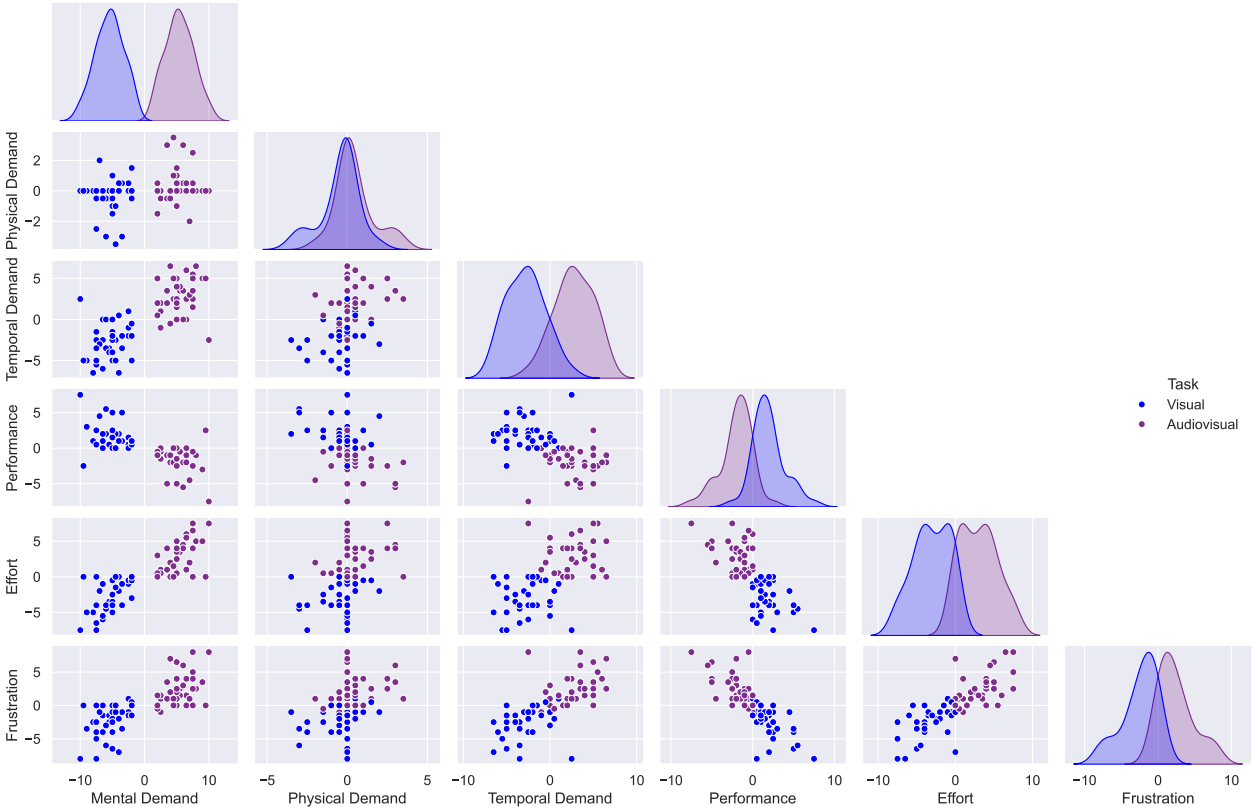

Figure 11: We represent the distributions of each NASA-TXL aspect (top of each column) and the scatter plot of individual responses comparing the different aspects pairwise.

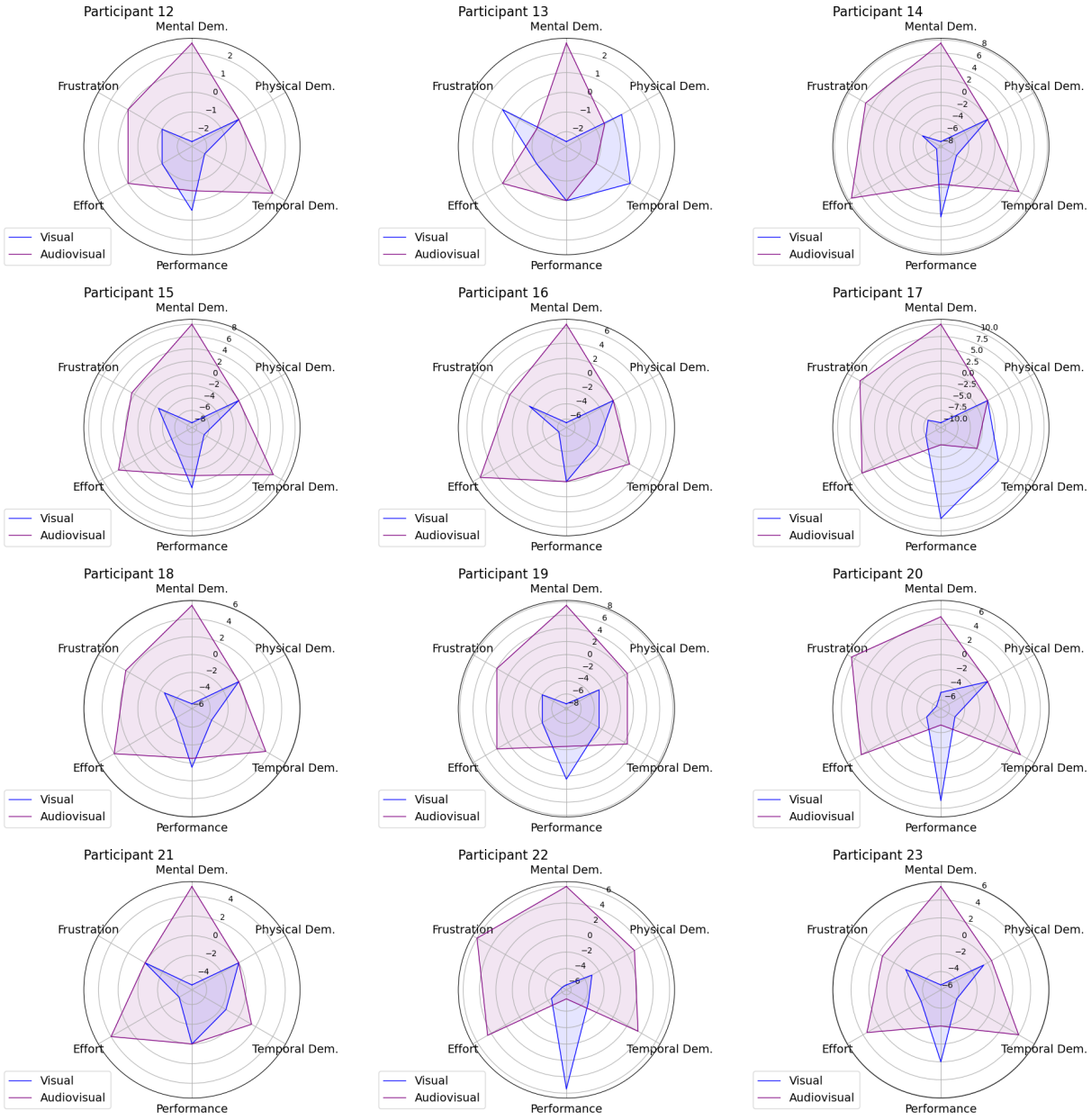

Figure 12: **Participants 12–23 NASA-TXL Questionnaires:** We show the normalized participants' answers to the NASA-TXL questionnaire for the visual (blue) and audiovisual tasks.

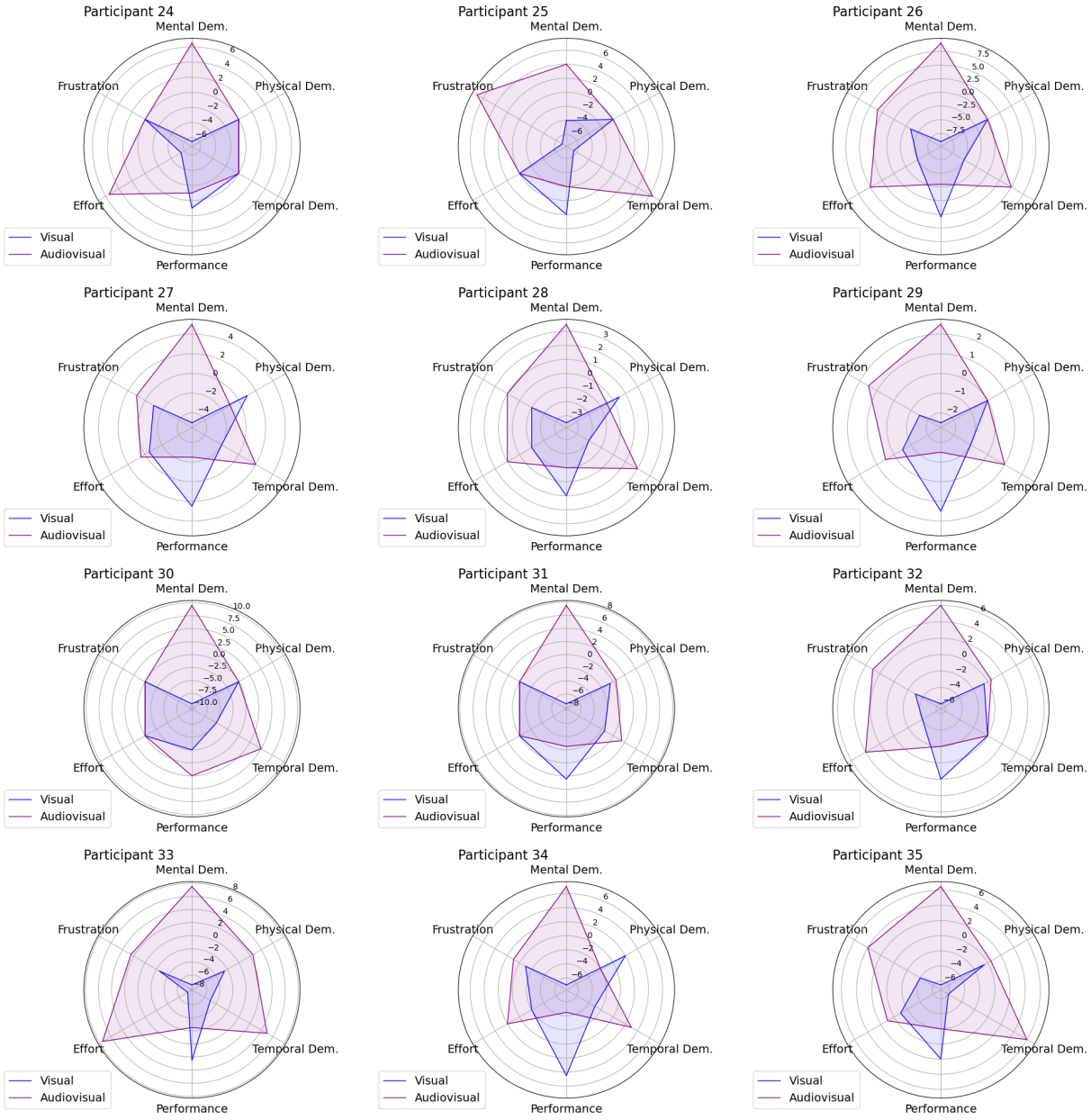

Figure 13: **Participants 24–35 NASA-TXL Questionnaires:** We show the normalized participants' answers to the NASA-TXL questionnaire for the visual (blue) and audiovisual tasks.

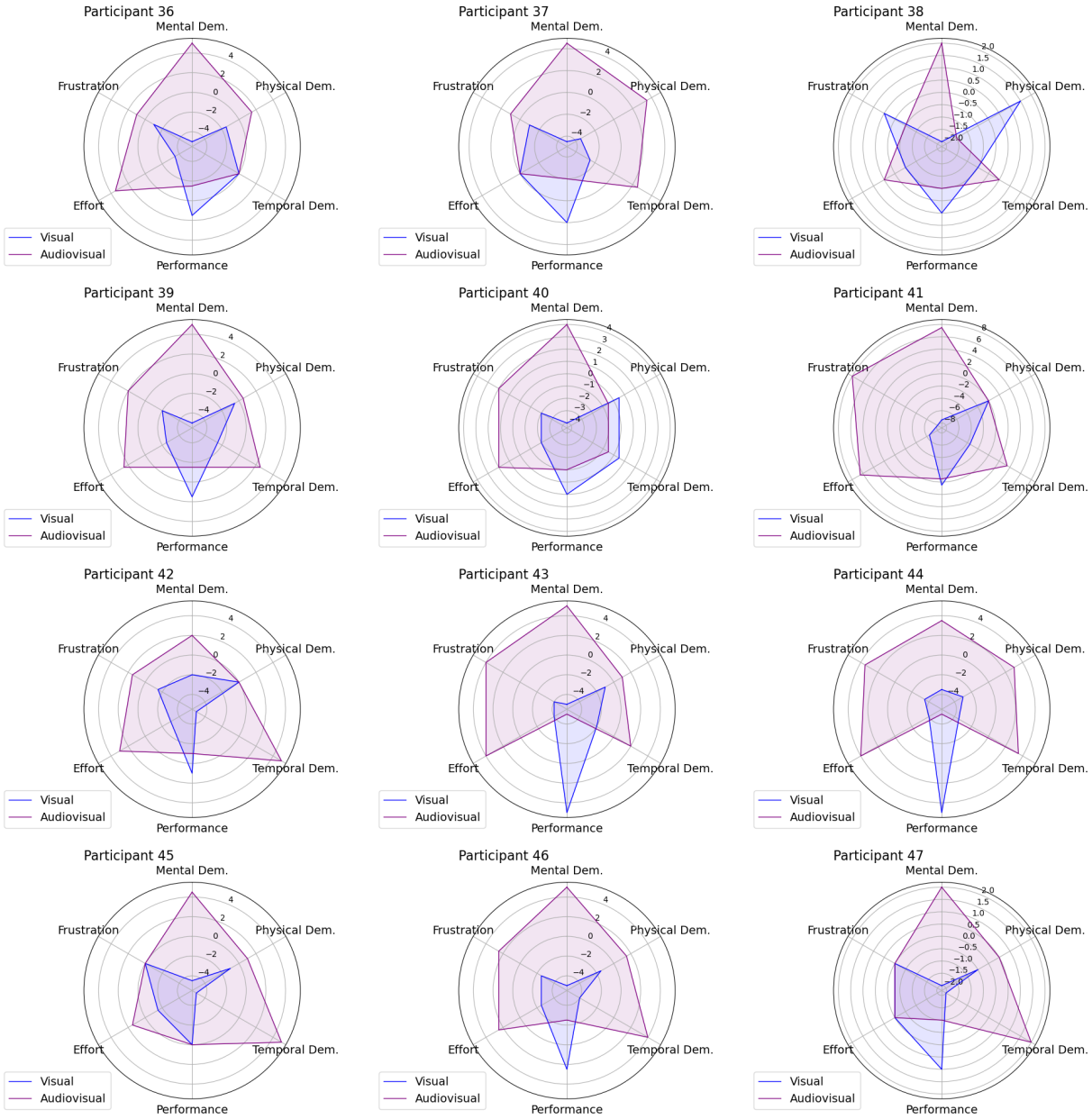

Figure 14: **Participants 36–47 NASA-TXL Questionnaires:** We show the normalized participants' answers to the NASA-TXL questionnaire for the visual (blue) and audiovisual tasks.

REFERENCES

[1] R. Bailón, G. Laouini, C. Grao, M. Orini, P. Laguna, and O. Meste. The integral pulse frequency modulation model with time-varying threshold: application to heart rate variability analysis during exercise stress testing. *IEEE transactions on biomedical engineering*, 58(3):642–652, 2010.

[2] Y. Benjamini and Y. Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)*, 57(1):289–300, 1995.

[3] P. H. Charlton, P. A. Kyriacou, J. Mant, V. Marozas, P. Chowienczyk, and J. Alastruey. Wearable photoplethysmography for cardiovascular monitoring. *Proceedings of the IEEE*, 110(3):355–381, 2022.

[4] A. Greco, G. Valenza, A. Lanata, E. P. Scilingo, and L. Citi. cvxeda: A convex optimization approach to electrodermal activity processing. *IEEE transactions on biomedical engineering*, 63(4):797–804, 2015.

[5] A. Greco, G. Valenza, J. Lazaro, J. M. Garzon-Rey, J. Aguilo, C. de la Cámara, R. Bailón, and E. P. Scilingo. Acute stress state classification based on electrodermal activity modeling. *IEEE Transactions on Affective Computing*, 14(1):788–799, 2021.

[6] S. Kontaxis, J. Lázaro, V. D. Corino, F. Sandberg, R. Bailón, P. Laguna, and L. Sörnmo. Ecg-derived respiratory rate in atrial fibrillation. *IEEE Transactions on Biomedical Engineering*, 67(3):905–914, 2019.

[7] J. Y. Lee, N. de Jong, J. Donkers, H. Jarodzka, and J. J. van Merriënboer. Measuring cognitive load in virtual reality training via pupillometry. *IEEE Transactions on Learning Technologies*, 17:704–710, 2023.

[8] J. P. Martínez, R. Almeida, S. Olmos, A. P. Rocha, and P. Laguna. A wavelet-based ecg delineator: evaluation on standard databases. *IEEE Transactions on biomedical engineering*, 51(4):570–581, 2004.

[9] H. Mitre-Hernandez, R. C. Carrillo, C. Lara-Alvarez, et al. Pupillary responses for cognitive load measurement to classify difficulty levels in an educational video game: Empirical study. *JMIR Serious Games*, 9(1):e21620, 2021.

[10] H. F. Posada-Quintero, J. P. Florian, A. D. Orjuela-Cañón, T. Aljama-Corrales, S. Charleston-Villalobos, and K. H. Chon. Power spectral density analysis of electrodermal activity for sympathetic function assessment. *Annals of Biomedical Engineering*, 44(10):3124–3135, Oct 2016. doi: 10.1007/s10439-016-1606-6

[11] S. Searle, F. Speed, and G. Milliken. Population marginal means in the linear model: an alternative to least squares means. *The American Statistician*, 34(4):216–221, 1980. doi: 10.1080/00031305.1980.10483031

[12] C. Varon, J. Morales, J. Lázaro, M. Orini, M. Deviaene, S. Kontaxis, D. Testelmans, B. Buyse, P. Borzée, L. Sörnmo, et al. A comparative study of ecg-derived respiration in ambulatory monitoring using the single-lead ecg. *Scientific reports*, 10(1):5704, 2020.