

Improving Upper-limb Prosthesis Usability: Cognitive Workload Measures Quantify Task Difficulty

1 Michael D. Paskett^{1*}, Jhorg K. Garcia¹, Sonny T. Jones¹, Mark R. Brinton^{1,2}, Tyler S.

2 Davis³, Christopher C. Duncan⁴, Joel M. Cooper⁵, David L. Strayer⁶, Gregory A. Clark¹

3 ¹Center for Neural Interfaces, Department of Biomedical Engineering, University of Utah, Salt Lake City,

4 Utah, USA

5 ²Engineering and Physics Department, Elizabethtown College, Elizabethtown, Pennsylvania, USA

6 ³Department of Neurosurgery, University of Utah, Salt Lake City, Utah, USA

7 ⁴Division of Physical Medicine and Rehabilitation, University of Utah, Salt Lake City, Utah, USA

8 ⁵Red Scientific Inc, Salt Lake City, Utah, USA

9 ⁶Cognition and Neural Science, Department of Psychology, University of Utah, Salt Lake City, Utah,

10 USA

11 * **Correspondence:**

12 Michael Paskett

13 michael.paskett@utah.edu

14

15 Abstract

16 Providing user-focused, objective, and quantified metrics for prosthesis usability may help reduce
17 the high (up to 50%) abandonment rates and accelerate the clinical adoption and cost reimbursement for
18 new and improved prosthetic systems. We comparatively evaluated several physiological, behavioral, and
19 subjective cognitive workload measures applied to upper-limb neuroprosthesis use.

20 Users controlled a virtual prosthetic arm via surface electromyography (sEMG) and completed a
21 virtual target control task at easy and hard levels of difficulty (with large and small targets, respectively).

22 As indices of cognitive workload, we took behavioral (Detection Response Task; DRT) and

23 electroencephalographic (EEG; parietal alpha and frontal theta power, and the P3 event-related potential)
24 measures for one group (n = 1 amputee participant, n = 10 non-amputee participants), and
25 electrocardiographic (ECG; low/high frequency heart-rate variability ratio) and pupillometric (task-
26 evoked pupillary response) measures for another group (n = 1 amputee participant, n = 10 non-amputee
27 participants), because all measures could not reasonably be recorded simultaneously. Participants of both
28 groups also completed the subjective NASA Task-Load Index (TLX) survey.

29 Ease of use, setup, piloting, and analysis complexity varied among measures. The DRT required
30 minimal piloting, was simple to set up, and used straightforward analyses. ECG measures required
31 moderate piloting, were simple to set up, and had somewhat complex analyses. Pupillometric measures
32 required extensive piloting but were simple to set up and relatively simple to analyze. EEG measures
33 required extensive piloting, extensive setup and equipment, careful monitoring, and moderately complex
34 analyses.

35 Across subjects, the DRT, low/high frequency heart-rate variability ratio, task-evoked pupillary
36 response, and NASA TLX significantly differentiated between the easy and hard tasks, whereas EEG
37 measures (alpha power, theta power, and P3 event-related potential) did not. Aside from the NASA TLX,
38 the DRT was the easiest to use and most sensitive to cognitive load across and within subjects. Among
39 physiological measures, we recommend ECG, pupillometry, and EEG/ERPs, in that order.

40 This study provides the first evaluation of multiple objective and quantified cognitive workload
41 measures during the same task with prosthesis use. User-focused cognitive workload assessments may
42 increase our understanding of human interactions with advanced upper-limb neuroprostheses and
43 facilitate their improvements and translation to real-world use.

44

45 **Significance Statement (194/250 words)**

46 The human arm is dexterous and able to sense objects it contacts. Restoring sensory and motor
47 function to a person with limb loss presents multiple challenges and requires improvements in robotics,
48 biological interfaces, decoding biological signals for prosthesis movement, and sensory restoration. The

49 scientific and engineering communities have made progress toward restoring arm function through
50 advanced neuroprostheses. However, most studies focus solely on task performance, and they typically
51 employ artificial experimental paradigms in which the user can devote full attention to the task, which is
52 often unrealistic for use in everyday activities. To develop neuroprostheses capable of restoring intuitive
53 arm function, engineers and scientists must also consider the difficulty of use, or cognitive burden, of
54 using the neuroprosthesis. Although many measures of cognitive workload have been developed, few
55 studies directly interrogate cognitive workload during neuroprosthesis use. An engineer or scientist
56 seeking to employ cognitive workload measures during neuroprosthesis use will likely wonder, as we did,
57 which measures are most suitable for their needs. To address this question, we empirically assess the
58 practical and functional merits and limitations of several physiological, behavioral, and subjective
59 techniques to measure cognitive workload during use of an advanced prosthesis. We anticipate that these
60 findings may influence other medical and consumer areas of human-computer interaction, such as virtual
61 reality or exoskeleton use.

62

63 **Keywords** (Min.3 - Max. 10):

64 cognitive workload, neuroprosthetics, rehabilitation, brain-computer interface, electromyography (EMG),
65 bionic arm, prosthesis, usability.

66 Introduction

67 Upper-limb prostheses generally rely on unintuitive controllers and do not restore sensation to the
68 user, often resulting in prosthesis abandonment (Biddiss and Chau, 2007a). These limitations are among
69 the major factors (Biddiss et al., 2007; Espinosa and Nathan-Roberts, 2019) in the high (30% to 50%)
70 prosthesis abandonment rate (Pons et al., 2005; Biddiss and Chau, 2007b). More recent innovations, such
71 as advanced, multi-articulating prostheses, have not yet produced substantial reductions in prosthesis
72 abandonment (Salminger et al., 2020). Sophisticated solutions for restoring sensation (Tan et al., 2014;
73 Graczyk et al., 2018; D'Anna et al., 2019; George et al., 2019; Schofield et al., 2019; Mastinu et al.,

74 2020), improving motor control (Ortiz-Catalan et al., 2014b; Hargrove et al., 2017; Ameri et al., 2019;
75 Salminger et al., 2019; Vu et al., 2020; Paskett et al., 2021), and improving prosthesis comfort through
76 interventions such as osseointegration (Ortiz-Catalan et al., 2014a; Mastinu et al., 2020) provide valuable
77 steps toward increasing user satisfaction and reducing abandonment.

78 One aspect of upper-limb prosthesis improvements that rarely is directly studied is the cognitive
79 workload or effort required to use a prosthesis. Previous work (Resnik et al., 2012) has conveyed the need
80 for direct cognitive workload measures for prosthesis use. Most studies with advanced prostheses
81 demonstrate some form of performance improvement; however, performance does not necessarily imply
82 ease-of-use and desirability. Ultimately, translating neuroprostheses from the laboratory to the clinic for
83 long-term use will require the technologies to be *desirable*. Desirability will very likely increase with
84 higher performance systems; it will certainly increase with high-performance systems that are easy to use.
85 We found strong subjective preferences for certain movement decoders even though the objective
86 performance was similar (Paskett et al., 2021), implying that user satisfaction and the desirability of the
87 decoder was influenced by more than performance alone. Humans move their endogenous hand with
88 dexterity and very little cognitive effort. That is, most movements are executed with a great deal of
89 automaticity, without occupying the mind with the low-level details of the action. The ideal prosthesis
90 should restore such automaticity to the user, enabling them to extend their focus beyond the prosthesis
91 when carrying out a task. Quantifying cognitive workload during prosthesis use may provide a clearer
92 path toward restoring automaticity.

93 Interrogating cognitive workload is possible through subjective, behavioral, and physiological
94 measures. There are benefits and limitations to each. In the upper-limb prosthesis domain, most attempts
95 at measuring cognitive workload (Markovic et al., 2018, 2020; Thomas et al., 2019) have been through
96 subjective measures, such as the NASA TLX survey (Hart and Staveland, 1988). Subjective measures are
97 quick and simple; however, they can suffer from large inter-individual variability, recall bias (Zahabi et
98 al., 2019), and task-order dependency (McKendricka and Cherry, 2018). A few studies have employed
99 behavioral measures (Witteveen et al., 2012; Raveh et al., 2018b; Valle et al., 2020) that generally use the

100 performance in a secondary task (e.g., a memory task) as an index of difficulty of the primary (prosthesis)
101 task. Behavioral measures are appealing because they measure cognitive workload contemporaneous with
102 the prosthesis task and do not suffer as directly from recall bias or task-order dependency. However, they
103 make the assumption that the paired secondary task will use mental capacity spared by the primary task
104 and that trade-off strategies are not employed during the tasks (Fisk et al., 1983). Some studies have used
105 physiological measures (Gonzalez et al., 2012; Deeny et al., 2014; White et al., 2017; Parr et al., 2019;
106 Thomas et al., 2021) to quantify cognitive workload. Physiological measures are valuable because they
107 rely on subconscious mechanisms to quantify cognitive workload and are relatively unaffected by
108 experimenters' or subjects' biases or expectations. However, capturing these phenomena generally
109 requires sophisticated equipment and well-prepared, and oftentimes constrained, conditions.

110 The question therefore arises: Which approach(es) should one use to measure cognitive
111 workload? To answer this question, we used an ordinary prosthesis control task – matching a virtual hand
112 to a target on a screen – for which we could easily manipulate task difficulty in order to compare
113 subjective, behavioral, and physiological measures of cognitive workload. By collecting multiple
114 cognitive workload measures during the same prosthesis task, our results facilitate direct comparisons of
115 the measures' effectiveness and utility. The results presented herein may thus aid researchers in selecting
116 quantified cognitive workload measures for their own studies with advanced prostheses. Additionally,
117 they may facilitate development, implementation, and clinical translation of easy-to-use prostheses.

118 Methods

119 Participant Recruitment

120 The present study was completed with two groups. In one group, we recorded behavioral and
121 EEG measures of cognitive workload. One amputee participant, male, in his 40s, had a congenital left
122 amputation approximately 10 cm below the elbow. Ten non-amputee participants completed the study:
123 three female, seven male, 24.6 ± 3.4 years old, one left-handed, nine right-handed.

124 In the other group, we recorded cardiac and pupillometric measures of cognitive workload. One
125 amputee participant, male, in his 40s, had bilateral traumatic amputations 10 years prior, about 8 cm
126 below the elbow, and is right hand dominant. Ten non-amputee participants completed the study: five
127 female, five male, 27.3 ± 12.2 years old, all right-handed. No participant from the first group was
128 included in the second group, so that both groups had equally naïve participants.

129 Cognitive Workload Measure Overview

130 We first briefly introduce the measures we employed in this study. For more in-depth reviews, we
131 recommend (Charles and Nixon, 2019; Lohani et al., 2019).

132 *DRT*

133 The DRT is a secondary task in which a visual, auditory, or tactile stimulus prompts the user to
134 respond by pressing a button. As the primary task increases in difficulty, the response time typically
135 increases, and stimulus detection rate typically decreases (i.e., the user does not respond). An ISO
136 standard of the DRT (ISO 17488:2016, 2016) has been used extensively in driving contexts (Ranney, T.
137 A., Baldwin, G. H. S., Smith, L. A., Mazzae, E. N., & Pierce, 2014; Chang et al., 2017; Stojmenova et al.,
138 2017; Strayer et al., 2017; Stojmenova and Sodnik, 2018), in which the stimulus is presented at random
139 intervals of 3-5 s. More broadly, secondary tasks have been applied to prosthesis tasks with promising
140 results, such as auditory discrimination tasks (Witteveen et al., 2012), memory tasks (Valle et al., 2020),
141 and games (Raveh et al., 2018b, 2018a). In contrast with the referenced uses of secondary tasks, we find
142 the DRT attractive because trials are collected rapidly (every few seconds) and the response times are
143 nearly continuous. The DRT has not been applied previously to prosthesis research.

144 *EEG and Event-Related Potentials*

145 EEG is the measure of electrical potentials produced by the brain at the scalp surface. Alpha
146 waves (8-12 Hz) in parietal regions indicate cortical idling and alpha power decreases with increased task
147 demands (Keil et al., 2006). Theta waves (4 - 7 Hz) in frontal midline regions arise when cognitive
148 control is required for a task (i.e., the task cannot be completed through an automatic strategy). Two

149 studies have measured alpha waves (Gonzalez et al., 2012; Parr et al., 2019), but no study has analyzed
150 theta waves during prosthesis use.

151 Event-related potentials (ERPs) are the brain's electrophysiological response to a particular
152 sensory, cognitive, or motor event (Luck, 2005). ERPs contain several components that represent various
153 stages in neural processing of an event. When used to measure cognitive workload, ERPs are usually
154 elicited through a secondary task, such as a DRT (Strayer et al., 2014). The P3 component, a positive
155 potential arising roughly 300-ms post-stimulus, decreases in amplitude as the primary (in our case,
156 prosthesis) task increases in difficulty and requires more resource allocation (Luck, 2005). Only one
157 previous study has used ERPs as a cognitive workload measure during prosthesis use (Deeny et al., 2014).

158 *Pupillometry*

159 The eyes have been described as the “visible part of the brain” (Hess and Janisse, 1978).
160 Pupillometry is the continuous measure of pupil size over the course of a task. Pupil size increases with
161 cognitive demands, demonstrated as early as the 1960s (Kahneman and Beatty, 1966). Because pupil size
162 changes due to several environmental, neurological, and psychological factors, trial averaging is often
163 used to produce a “task-evoked pupillary response” (Beatty, 1982). Measuring the percentage of pupil
164 dilation provides a measure that is robust to inter-individual and inter-trial baseline pupil size differences
165 (Payne et al., 1968). Pupillometry has been used only rarely in the prosthesis domain (White et al., 2017;
166 Zahabi et al., 2019).

167 *Electrocardiography*

168 Electrocardiography (ECG) is the measure of electrical potentials produced from the heart.
169 Several time-domain and frequency-domain measures are sensitive to cognitive workload (Charles and
170 Nixon, 2019). For our study, we used the low frequency (0.02 – 0.06 Hz) to high frequency (0.15 – 0.5
171 Hz) ratio (LF/HF ratio) because it showed the greatest sensitivity in our pilot experiments. One other
172 prosthesis study has used ECG to measure cognitive workload (Gonzalez et al., 2012).

173 *NASA TLX Survey*

174 The NASA TLX is a subjective survey designed to measure perceived workload (Hart and
175 Staveland, 1988) through six different categories on a 100-point scale: mental demand, physical demand,
176 temporal demand, performance, effort, and frustration. Participants compare categories pairwise based on
177 perceived importance in the task, and individual weightings from these comparisons are used to produce a
178 composite score. The TLX has been used widely across many domains, including prostheses (Gonzalez et
179 al., 2012; Markovic et al., 2018, 2020; Shaw et al., 2019; Thomas et al., 2019).

180 *Experiment Overview*

181 Participants controlled a virtual prosthetic hand using sEMG signals to complete a virtual target
182 task at easy and hard difficulties (Fig. 1). During the virtual task, we recorded subjective (NASA TLX),
183 physiological (ECG, EEG, and pupillometry) and behavioral (DRT) data to be used as measures of
184 cognitive workload. Because all the measures could not reasonably be collected at the same time, we
185 recorded ECG and pupillometry together in one set of experiments, and EEG and the DRT together in
186 another set of experiments.

187 *Prosthesis Control*

188 The prosthesis control methodology used in this study has been described previously (George et
189 al., 2020a). In brief, sEMG was collected from an sEMG sleeve (George et al., 2020b) with the Grapevine
190 System (Ripple Neuro LLC, Salt Lake City, UT). Thirty-two single-ended channels were acquired at 1
191 kHz and band-pass filtered between 15 Hz and 375 Hz with 4th-order Butterworth filters, and 60, 120, and
192 180 Hz 2nd-order Butterworth notch filters. After the sEMG sleeve was connected to the acquisition
193 device, channels were manually inspected and removed if broken channels were detected (generally less
194 than two channels). The differential pairs of all monopolar channels were calculated, and features (single-
195 ended and differential) were created at 30 Hz using the mean absolute value of a 300-ms buffer (i.e., 528
196 features from an overlapping 300-ms boxcar filter). At 30 Hz, the buffer is updated every 33 ms. This
197 update rate and buffer length has been used by our group extensively (George et al., 2018, 2020a).

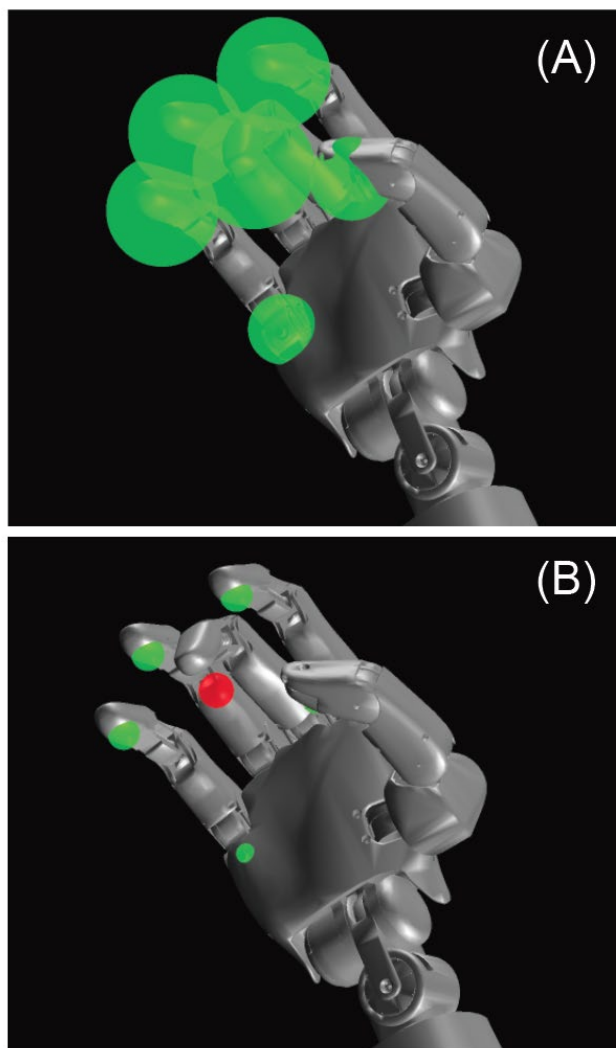


Figure 1. Virtual Target Task. Participants control the virtual hand and attempt to keep all targets green as random targets move to a target position for a specific time (5-15 s). (A) Large target with the middle finger active. Because the middle finger is within the target window, the target is green. (B) Small target with the middle finger active. Because the finger is outside the target window, the target is red.

198 sEMG was collected as participants mimicked preprogrammed movements of the virtual MSMS
199 hand (Davoodi and Loeb, 2011). The preprogrammed movements consisted of index, middle, and ring
200 finger flexions. Each movement consisted of a 0.7-s transition to flexed position, 4-s hold, and 0.7-s
201 return to rest position. Participants completed two trials of each flexion as practice to gain familiarity with
202 the virtual environment. After familiarization, participants completed five trials of each movement. Using

203 a Gram-Schmidt forward selection algorithm (Nieveen et al., 2017), 48 sEMG features were selected as
204 inputs to the decoder, a modified Kalman filter (George et al., 2020a). The 48 features and virtual hand
205 kinematics were used to fit the parameters of the modified Kalman filter. After fitting the modified
206 Kalman filter, users were given control over the virtual hand. We let the participants spend a few minutes
207 exploring the control; in cases where the participants struggled to fully flex the fingers, participants
208 repeated the five-trial mimicry and the modified Kalman filter was refitted.

209 *DRT*

210 We made a custom DRT system that interfaced with the Ripple Grapevine Digital I/O board. This
211 system turned the tactile buzzer, a 10 mm x 2 mm vibration motor on a 4.5 V power supply, on or off
212 when an output of the Digital I/O was set to high or low, respectively. The response button, when
213 depressed, was recorded by the Ripple Grapevine system. The DRT vibrations were set to 1 s and turned
214 off if the user pressed the response button before the 1 s had ended. Timestamps for the Digital I/O board
215 are recorded at 30-kHz resolution. The DRT system was placed on the table near the participant and two
216 separate cables for the vibration motor and response button were routed to the participant. We attached
217 the DRT vibration motor to the collarbone with medical tape, opposite the hand used for the prosthesis
218 task. We attached the response button to the index finger using a hook and loop fastener.

219 *EEG & ERP Recordings*

220 EEG was recorded based on the standard 10-20 system using a 34-electrode cap (Ripple Neuro
221 LLC, Salt Lake City, UT). Electrode locations were: FP1, FP2, F7, F3, Fz, F4, F8, AFz, FT7, FT8, FC3,
222 FCz, FC4, T3, C3, Cz, C4, T4, CP3, CPz, CP4, T5, P3, Pz, P4, T6, O1, Oz, O2, A1, A2, VEOL, HEOR,
223 HEOL. The online reference was on electrode CPz, and the ground was AFz. We used Electro-Gel™ to
224 bridge the connection between the electrodes and the scalp. Impedances of the electrodes were brought
225 below 10 kOhm, typically close to 5 kOhm, using gentle scalp abrasion. We recorded the scalp EEG at 1
226 kHz and band-pass filtered between 1 Hz and 125 Hz with 4th-order Butterworth filters.

227 *Pupillometry Recordings*

228 Pupil diameter was recorded using the Pupil Labs' Pupil Core head-mounted pupillometry device.
229 We recorded pupil diameter with both pupil cameras at 120 Hz and 400x400 resolution. The 2D diameter
230 output from the Pupil Labs' software was used for analysis, which contains a measured diameter and the
231 measurement confidence, ranging from zero to one. Room lighting was kept constant at approximately
232 100 lux, as measured by an Urceri MT-912 light meter.

233 *ECG Recordings*

234 ECG was recorded with the five-wire, four-lead Shimmer3 ECG unit. The unit recorded at 512
235 Hz. The Vx electrode was placed at V5, as suggested in the Shimmer3 ECG user manual, and the
236 remaining electrodes were placed on the chest in the direction of the right arm, left arm, right leg, and left
237 leg. ECG recordings were programmatically started and stopped when a target set was started or finished,
238 respectively. The Shimmer3 logs the data onto an internal SD card, which was later extracted using
239 Shimmer3 Consensus software.

240 *Virtual Target Task*

241 Participants completed a virtual target task in the MSMS virtual environment (Davoodi and Loeb,
242 2011) for easy and hard difficulties. In the virtual target task, a spherical target indicates the desired
243 position of each degree-of-freedom. When a degree-of-freedom is within a specified radius of the target,
244 the target is green; outside the allowable radius, the target is red. For our target tasks, the target was
245 placed halfway through the movement window, with a target size (i.e., allowable radius) of 35% and 15%
246 of the movement window for the easy and hard difficulties, respectively.

247 Participants were instructed to focus most of their attention on the active target, which was only
248 one degree-of-freedom at a time. Participants were instructed that their objective was to keep the target
249 green, not to keep the active degree-of-freedom in the middle of the target. Participants were encouraged
250 to stay focused on the task and to avoid talking during the task in order to reduce cognitive demands
251 beyond the task itself.

252 The subsequent sections describe the target task paradigm for each group. Because the different
253 cognitive load measures have differing recording requirements, the experimental paradigms were slightly
254 different for the two groups. The difficulty of the task (i.e., target size) was identical for both groups.

255 EEG & DRT Experiments

256 In the EEG and DRT virtual target task, the targets were active for 15 s. The participants first
257 completed one practice set without the DRT that included one trial of each degree-of-freedom for each
258 target size in a random order, for a total of six trials. For the next practice set, vibrotactile stimuli from the
259 DRT system were presented randomly 3-5 s apart (uniformly distributed), according to ISO 17488 (ISO
260 17488:2016, 2016), resulting in, on average, 3 vibrotactile stimuli per active target. After the two practice
261 sets, the participants completed eight rounds of the target task with the DRT. After the final target set was
262 completed, users completed the NASA TLX for each target size in a random order.

263 ECG & Pupillometry Experiments

264 In the ECG and pupillometry virtual target task, the targets were active for 5 s with a random 3-5
265 s interval between targets (uniformly distributed). The participants first completed one practice target set
266 for each target size. In the practice sets, each degree-of-freedom was tested twice, in a random order.
267 After practicing the task, participants moved onto the full-length target sets. In one target set, each
268 degree-of-freedom (index, middle, and ring finger) was tested six times, in a random order, for a total of
269 18 target trials per set. A set included only one target size. To calculate difference waves with the
270 pupillary responses, participants also completed a “mimicry” set of targets. In the “mimicry” set, the
271 computer perfectly completed the target task while the participants watched and mimicked the
272 movements. Before the “mimicry” target set, participants were informed that the computer would be in
273 control of the virtual hand, and they were instructed to watch the task and mimic the computer’s
274 movements. The difference waves are discussed in greater detail in the analysis section. Participants
275 completed one target set and one “mimicry” set for a single target size, then completed one target set and
276 one “mimicry” set for the other target size. The initial target size was randomized. For the full

277 experiment, participants completed four active and four “mimicry” target sets for each target size,
278 resulting in 72 individual target trials per participant. After the final target set was completed for each
279 target size, users completed the NASA TLX survey.

280 Analysis

281 *DRT*

282 We analyzed the DRT according to the ISO standard (ISO 17488:2016, 2016). Responses (button
283 presses after vibrotactile stimuli) less than 100 ms or greater than 2500 ms were counted as a miss.
284 Response times more than three scaled median absolute deviations from the median were excluded from
285 the analysis. We measured the hit rates and response times during each target size.

286 *EEG & ERP*

287 EEG was analyzed using EEGLAB v2021.0. The data were first resampled to 250 Hz. We re-
288 referenced the electrodes to electrodes A1 and A2. The data were filtered from 0.1 Hz to 30 Hz using a
289 second-order Butterworth filter. Artificial blink and horizontal eye movement channels were created by
290 subtracting VEOL from FP1, and HEOL from HEOR, respectively.

291 For the frequency analysis, 15-s bins were created for the duration of the active target and
292 separated by target size. Artifacts were detected and removed if the blink or horizontal movement
293 channels exceeded a 100- μ V threshold within a 200-ms sliding window. The 200-ms window passed
294 across the 15-s bin in 50 ms increments. Individual bins were Hann-windowed prior to calculating the
295 power-spectral density of each trial to avoid edge effects. Power-spectral densities of each trial were
296 averaged together. The power for the alpha band (8-12 Hz) on electrode Pz and theta (4-7 Hz) band on
297 electrode Fz were calculated. The percentages of power in the alpha and theta bands were calculated by
298 dividing the power in the selected bands by the total power.

299 For the ERP analysis, bins were created from 200 ms prior to the buzzer onset to 1100 ms after
300 the onset and separated by target size. Artifacts were detected and removed if the blink or horizontal
301 movement channels exceeded a roughly 60- μ V threshold within a 200-ms sliding window. The threshold

302 was slightly adjusted when blinks or horizontal eye movements were not detected by the initial 60- μ V
303 threshold. The 200-ms window passed across the 1300 ms bin in 50 ms increments. Non-artifact trials
304 were averaged together to produce an averaged ERP for each participant. The signed area was calculated
305 from 200 ms to 650 ms (Strayer et al., 2014) to calculate the P3 ERP size. Averaged ERPs for each
306 participant were averaged across participants to produce grand-averaged ERPs.

307 *Pupillometry*

308 Pupil recordings were aggregated by target size. We removed outliers defined as measurements
309 greater than three scaled mean absolute deviations from the median of 60 samples (a 0.5-s window). We
310 removed measurements with measurement confidence less than 0.8. Removed measurements were
311 replaced with linearly interpolated values. Target trials with more than 20% low confidence
312 measurements were removed from the aggregated set. The pre-trial baseline diameter, 1 s before the
313 target became active, was subtracted from each trial. The percentage change in pupil size was calculated
314 by dividing the response by the average size of the pupil during the 1-s pre-trial baseline. The baseline-
315 subtracted pupillary responses of both eyes were combined and averaged to find the average pupillary
316 response to the target task. The averaged pupillary response from the mimicry target (where the computer
317 controlled the virtual hand) was subtracted from the averaged pupillary response to the active target
318 (where the user controlled the virtual hand) to create a difference wave that would mitigate target-size
319 dependent luminance effects in the response. We calculated the average value of the difference wave
320 during the 5 s the target was active.

321 *ECG*

322 We obtained the LF/HF heart-rate variability ratio using the standard settings of PhysioZoo
323 version 1.2.0 (Behar et al., 2018). The ECG was band-pass filtered from 3 Hz to 100 Hz with second and
324 fifth-order Butterworth filters, respectively. Peaks in the ECG were detected using an energy-based QRS
325 detector (Behar et al., 2014). The heart rate variability (intervals between normal heart beats) was
326 calculated after removing outliers in the R-R peak intervals. Outliers were defined as intervals above or

327 below 20% of the average of the moving window, which was 21 intervals. Power spectral density of the
328 heart rate variability was calculated with Welch's method. Finally, the LF/HF ratio was calculated by
329 dividing the power in the low frequency region (0.04 Hz to 0.15 Hz) by the power in the high frequency
330 region (0.15 Hz to 0.4 Hz).

331 *Target Task Performance*

332 We calculated the average percentage of time spent within the target window for each target size
333 for each participant.

334 *Statistical Procedures: Across-Subject*

335 We tested the paired values derived from each measure for normality using the Shapiro-Wilk test.
336 If the paired values were normally distributed, we used a paired t-test to show differences between the
337 responses to the large and small targets. If the paired values were nonparametric, we used Wilcoxon's
338 signed-rank test. Because only one amputee participant completed each experiment, we did not include
339 amputee participant results in our across-subject statistical measures and instead overlay results from
340 amputee participants with the results of non-amputee participants.

341 *Statistical Procedures: Within-Subject*

342 Different measures may work well for some individuals, but not others. Additionally, due to the
343 costs associated with implanting neural and electromyographic interfaces, it is common for studies to be
344 completed with only a few subjects. We therefore were interested in the within-subject reliability of the
345 cognitive workload measures. We completed within-subject analyses for each subject for each measure as
346 appropriate for the measure and experimental paradigm. For the DRT, we conducted a two-sample t-test
347 for all the DRT trials in an experiment. For the EEG & ERP measures, we conducted paired-sample t-
348 tests with the average response for each of the eight rounds of the target task. For the pupil & ECG
349 measures, we conducted paired-sample t-tests with the average response for each of the four rounds of the
350 target task. We calculated the p-value from the statistical test and the absolute effect size using Cohen's D
351 for each participant. We report the median and first and third quartiles of p and D across non-amputee

352 participants and the individual outcomes for the amputee participant. We report the number of
353 participants for whom $p < 0.05$ by measure.

354 Results

355 In brief, several but not all measures of cognitive load differentiated between the easy and hard
356 tasks reliably in the aggregate intact subject pool. Significant differences ($p < 0.05$ or less) occurred for
357 DRT, pupil dilation, LF/HF ratio, and TLX scores. Averaged ERPs, alpha and theta EEG powers, task-
358 evoked pupillary responses, and heart-rate variability powers for the easy and hard tasks are shown in
359 Fig. 2. Outcomes from individual participants are shown in Fig. 3. Each measure is discussed in detail in
360 the following subsections. Parametric statistics are reported as mean \pm standard error of the mean, and
361 nonparametric statistics are reported as median [inter-quartile range].

362 Target Task

363 Confirming empirical differences in task difficulty, non-amputee participants performed
364 significantly worse on the hard task (i.e., small target) compared with the easy task. For the DRT and
365 EEG paradigm, non-amputee participants spent, on average, $33\% \pm 3\%$ less time within the target
366 window on the hard task ($p < 0.001$, paired t-test). The amputee participant had similar performance to
367 the non-amputee participants, spending 41% less time within the target window for the hard task (small
368 target: 48% [33%]; large target: 89% [9%]; $p < 0.001$, Wilcoxon's rank sum test).

369 For the ECG & pupillometry paradigm, non-amputee participants spent $36\% \pm 2\%$ less time
370 within the target window for the hard task ($p < 0.001$, paired t-test). The amputee participant had similar
371 performance to the non-amputee participants, spending 33% less time within the target window for the
372 hard task (small target: 47% [21%]; large target: 80% [9%]; $p < 0.001$, Wilcoxon's rank sum test).

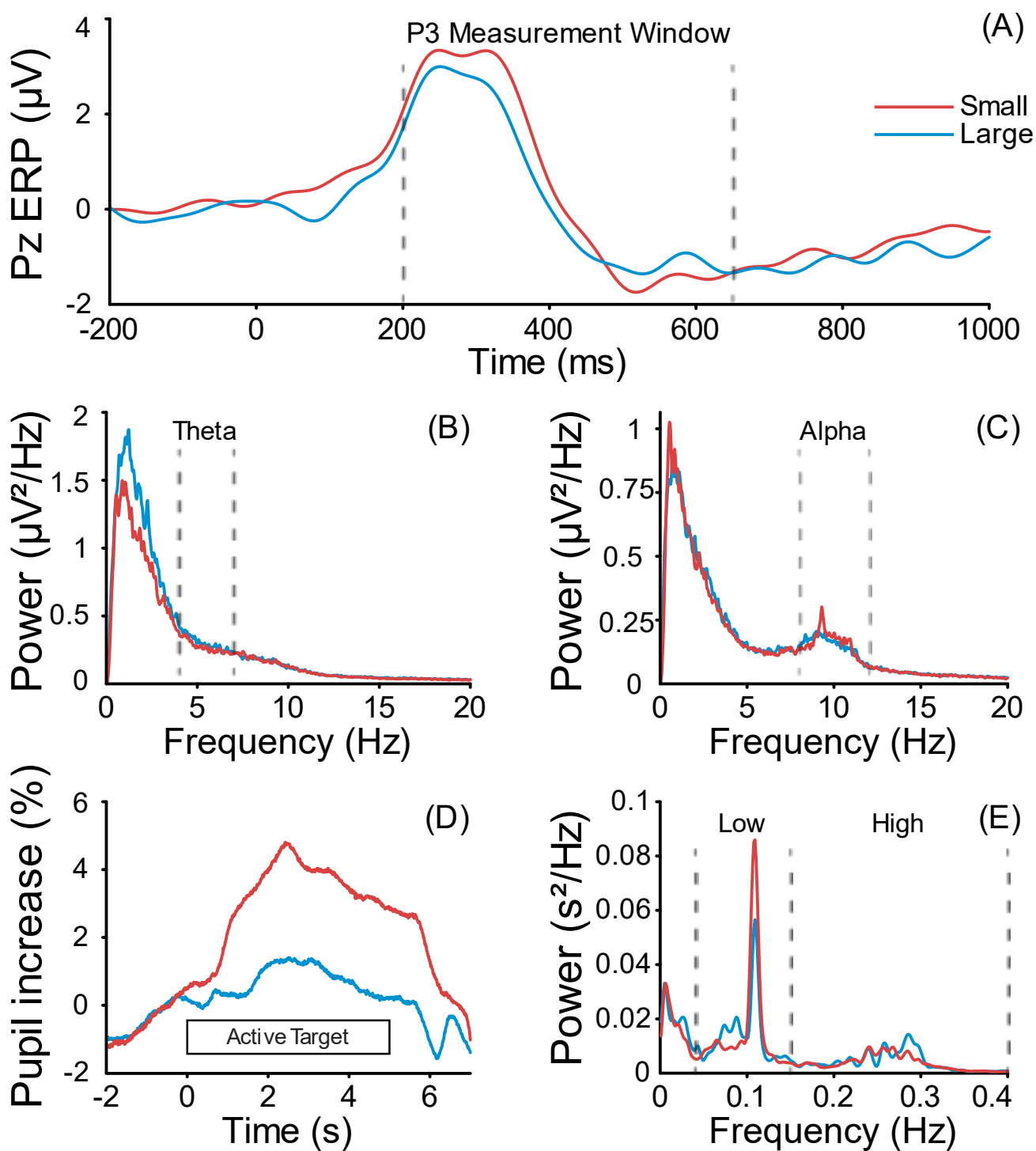


Figure 2. Raw physiological measures of cognitive load acquired during virtual target task at easy (large) and hard (small) difficulties for non-amputee participants. (A) Event-related potential (ERP) at electrode Pz arising from vibrotactile DRT stimulus. (B) Theta EEG power (4-7 Hz) at electrode Fz. (C) Alpha EEG power (8-12 Hz) at electrode Pz. (D) Luminance-corrected task-evoked pupil response. (E) Heart-rate variability power.

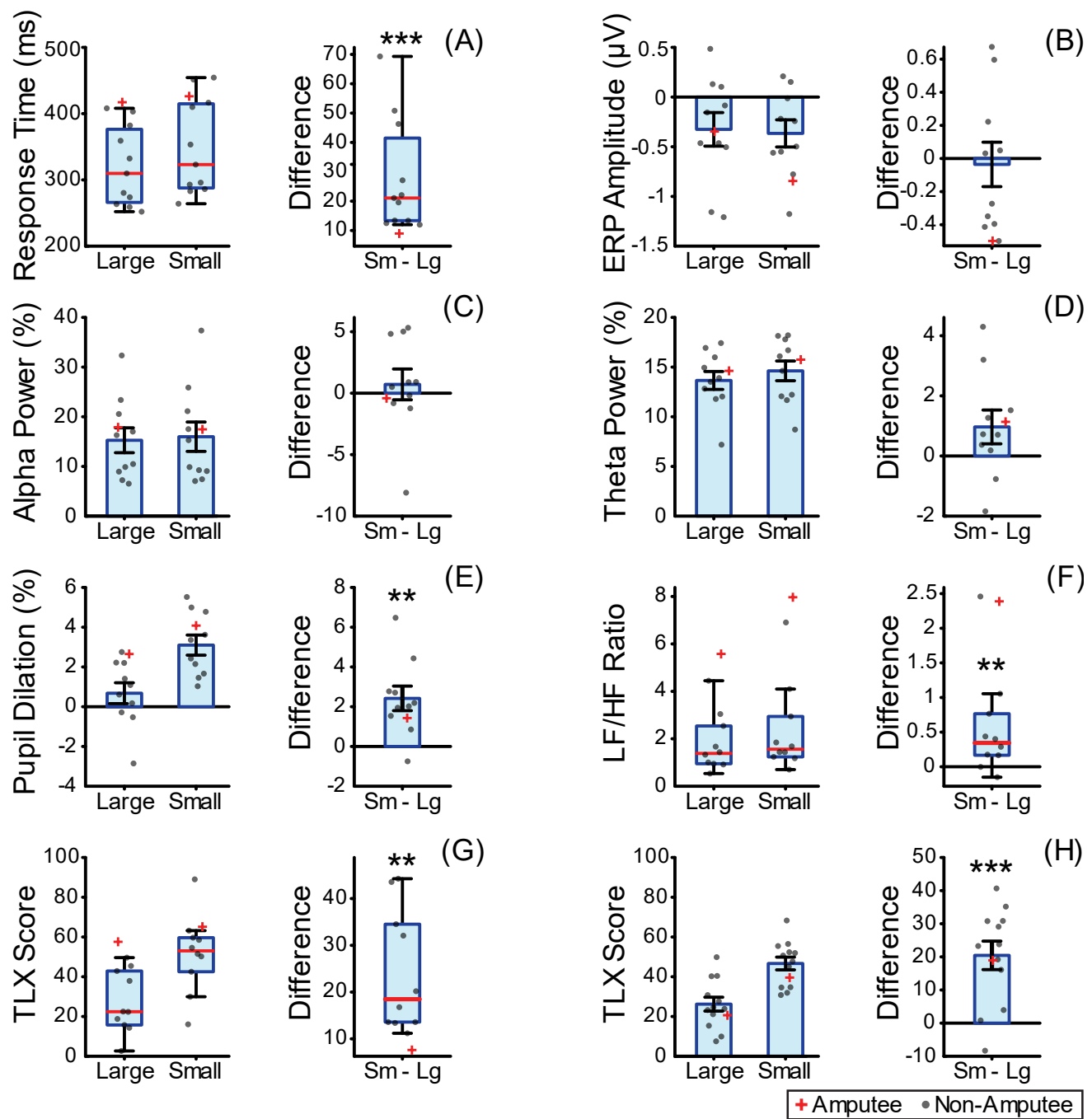


Figure 3. Several but not all measures of cognitive load changed with task difficulty. Shown are cognitive load measures from the (A) DRT, (B) P3 event-related potential, (C) alpha EEG power, (D) theta EEG power, (E) pupil dilation, (F) heart-rate variability low/high frequency ratio, and NASA TLX scores from the (G) EEG & DRT set and the (H) ECG & pupillometry set. Group descriptive and inferential statistics are depicted for the non-amputee participants only, without data from the amputee subject. For boxplots, red lines represent the median, the box represents Q1 and Q3, and the whiskers represent the outermost non-outlying points, as defined by the 1.5 * interquartile range extending from Q1 and Q3. For bar graphs, the top of the bar represents the mean, and the error bars represent the standard error of the mean. Paired comparisons were made (right subfigures) using parametric or nonparametric statistical tests, as applicable, for non-amputee participants only. *, **, and *** represent $p < 0.05$, $p < 0.01$, $p < 0.001$, respectively.

375 DRT

376 Non-amputee participants' response times to the vibrotactile stimulus significantly increased by
377 21 ms [28 ms] when participants were completing the hard task ($p < 0.001$, Wilcoxon's signed-rank test;
378 Fig. 3a). The amputee participant's response times increased by 9 ms for the hard task (small target: 419
379 ms [90 ms]; large target: 410 ms [87 ms]), but the difference was not significant ($p = 0.42$; Wilcoxon's
380 rank sum test). Hit rates (i.e., responses between 100 ms and 2500 ms) for both conditions were above
381 98% for all participants with no significant differences.

382 EEG & ERP

383 EEG power spectra and grand-averaged ERPs for non-amputee participants are shown in Fig. 2a-
384 c. No EEG or ERP measures were found to differ significantly between the easy and hard tasks (Fig. 3b-
385 d). Theta power was not significantly different between easy and hard tasks for non-amputee participants
386 (mean difference, hard task - easy task, $1.0\% \pm 0.6\%$; $p = 0.12$, paired t-test) or for the amputee
387 participant (mean difference, hard task - easy task, $1.1\% \pm 0.4\%$; $p = 0.22$, paired t-test). Alpha power
388 was not significantly changed for non-amputee participants (mean difference, hard task - easy task, 0.7%
389 $\pm 1.3\%$; $p = 0.58$, paired t-test) or amputee participant (mean difference, hard task - easy task, $0.4\% \pm$
390 0.2% ; $p = 0.70$, paired t-test) for the amputee participant. The ERP size significantly decreased by $0.5 \mu\text{V}$
391 $\pm 0.1 \mu\text{V}$ for the hard task for the amputee participant ($p < 0.001$; paired t-test), but there was no
392 significant difference for the non-amputee participants (mean difference, hard task - easy task, $0.0 \mu\text{V} \pm$
393 $0.1 \mu\text{V}$; $p = 0.8$, paired t-test).

394 Pupillometry

395 The task-evoked pupillary responses for the non-amputee participants are shown in Fig. 2d. The
396 task-evoked pupillary response significantly increased by $2.4\% \pm 0.6\%$ for the hard task for non-amputee
397 participants ($p < 0.01$, paired t-test; Fig. 3e). The amputee participant's pupil response was not
398 significantly different (mean difference, hard task - easy task, $1.4\% \pm 1.4\%$; paired t-test).

399 ECG

400 The heart-rate variability power spectrum is shown in Fig. 2e. The LF/HF ratio significantly
401 increased by a median value of 0.34 ($p < 0.01$, Wilcoxon's signed-rank test; Fig. 3f) for non-amputee
402 participants. The LF/HF ratio for the amputee participant did not significantly differ (mean difference,
403 hard task - easy task, $2.4 \pm 2.0\%$; $p = 0.32$, paired t-test) .

404 NASA TLX

405 For the DRT and EEG paradigm, the TLX score significantly increased by a median value of 18
406 for non-amputee participants ($p < 0.01$, Wilcoxon's signed-rank test; Fig. 3g), and increased by 7 for the
407 amputee participant. For the ECG & pupillometry paradigm, the TLX score significantly increased by an
408 average value of 20 for non-amputee participants ($p < 0.001$, paired t-test; Fig. 3h), and increased by 19
409 for the amputee participant.

410 Within-Subject Analysis

411 The p-values and effect sizes for the different cognitive load measures are shown in Table 1 for
412 within-subject analyses for non-amputee participants and the amputee participant. Consistent with
413 statistically significant results for across-subjects analyses, the DRT was the most reliable for the within-
414 subject analysis, being significantly different for eight of ten non-amputee participants, with a median p-
415 value of 0.001. Although EEG alpha power and theta power were not significantly different in the across-
416 subjects analyses, these measures were significantly different for 5 and 3 individual non-amputee
417 participants, respectively. Pupil dilation and LF/HF ration were both significant for the across-subjects
418 analyses, but showed significant differences for only 2 and 1 individual non-amputee participants,
419 respectively. The ERP was not significant for any individual non-amputee participant, consistent with the
420 lack of significantly different results in across-subjects analyses. In contrast, for the amputee participant,
421 the ERP was the only measure to be significantly different between the easy and hard tasks. The NASA
422 TLX was administered only a single time for each participant, so inferential statistical analyses were not

Table 1. Within-subject reliability of the cognitive workload measures

| Measure | Probability (p) | | | Absolute Effect Size (Cohen's D) | |
|----------------------------|------------------------------|----------------------------------|---------------------|----------------------------------|---------------------|
| | Median (Q1, Q3) [†] | Number of Participants, p < 0.05 | Amputee Participant | Median (Q1, Q3) | Amputee Participant |
| DRT | 0.001 (<0.001, 0.023) | 8 | 0.371 | 0.382 (0.266, 0.578) | 0.103 |
| EEG: Alpha Power | 0.138 (0.005, 0.429) | 5 | 0.700 | 0.436 (0.211, 0.993) | 0.156 |
| EEG: Theta Power | 0.248 (0.044, 0.410) | 3 | 0.220 | 0.365 (0.280, 0.563) | 0.527 |
| ERP | 0.368 (0.221, 0.724) | 0 | 0.001 | 0.300 (0.155, 0.514) | 1.722 |
| Task-Evoked Pupil Response | 0.096 (0.043, 0.207) | 2 ^{††} | 0.460 | 0.498 (0.358, 0.744) | 0.276 |
| ECG: LF/HF Ratio | 0.319 (0.149, 0.585) | 1 | 0.320 | 0.298 (0.189, 0.556) | 0.270 |

[†] N = 10 non-amputee participants, except as otherwise indicated

^{††} out of 7 non-amputee participants (3 non-amputee participants had insufficient data for within-subject analysis)

423 possible on a per-subject basis. However, all amputee and non-amputee participants rated the small target
 424 task as harder.

425 Discussion

426 This is the first prosthesis study that directly compares the efficacy and utility of several different
 427 objective, quantified cognitive workload measures that span physiological, behavioral, and subjective
 428 domains. Our objective was to determine the best technologies for user-focused prosthesis evaluations
 429 that will push laboratory developments toward clinical realities. We found the DRT to be the easiest to
 430 use and most sensitive to cognitive load across and within subjects. On the basis of their utility and their
 431 ability to differentiate among task difficulties, we next recommend ECG, pupillometry, and EEG/ERPs,
 432 in that order.

433 The comparative evaluations herein can inform the field’s use of cognitive workload measures in
434 subsequent studies. Such studies could explore users’ responses to aspects of motor control, such as
435 comparing decoders, finding a desirable number of degrees-of-freedom, or showing potential benefits of
436 an active wrist. On the sensory side, one could explore the cognitive implications of sensorized and non-
437 sensorized prostheses, compare feedback modalities (electrical vs. vibrotactile) or compare stimulation
438 algorithms.

439 Designing experiments that could accommodate the recording requirements of the various
440 measures used in this study was challenging because design choices could preferentially benefit a
441 particular measure. We strived to provide suitable environments for all the measures and an experimental
442 design that would enable effective collection of all the cognitive workload measures used. In the end,
443 however, we were seeking for measures that are robust to environmental and experimental changes. We
444 discuss the results, strengths, and limitations of the individual measures in the following subsections.

445 DRT

446 We found that the DRT resulted in the most significant differentiation between the easy and hard
447 tasks and was the most reliable for within-subject analysis. Overall, we recommend the DRT as a very
448 reliable measure of cognitive workload that requires minimal setup and technical expertise. The DRT
449 required minimal piloting and experimental manipulations before moving forward with recorded
450 experiments. The DRT has several desirable characteristics as a cognitive workload measure: it is
451 portable, requires minimal setup and its results are easily interpreted. This study demonstrates the first
452 application of the DRT to a prosthesis task.

453 The DRT is limited by requiring physical button presses; however, many tasks for quantifying
454 prosthesis performance are completed with one hand. Additionally, the response button could be modified
455 for two-handed tasks (e.g., placed at the foot). The strengths and limitations of the DRT are discussed
456 further in (Stojmenova and Sodnik, 2018). There are some aspects of behavioral measures that are not as

457 attractive as physiological measures; however, the sensitivity and robustness of the DRT overcame our
458 bias for physiological measures.

459 ECG

460 The LF/HF ratio reliably detected differences in task difficulty. We recommend ECG, specifically
461 the LF/HF ratio, as a viable physiological measure of cognitive workload that works for short-duration
462 tasks. Although ECG worked well across subjects, for within-subject reliability, a greater number of trials
463 is likely required. ECG is a relatively simple signal to obtain. The vast number of heart rate and heart-rate
464 variability metrics (see (Charles and Nixon, 2019) for a review containing several ECG measures of
465 cognitive workload) created a large parameter space to explore. Deciding on an ECG measure required a
466 fair amount of piloting before use in experiments for the present study. Once selected, the LF/HF ratio
467 remained robust. ECG measures of cognitive workload require relatively long recordings (>3-4 minutes),
468 longer than many standardized prosthesis tasks, which can make task selection difficult.

469 In a study measuring cognitive load with a sensorized prosthesis, heart rate was found to decrease
470 when participants had audiovisual feedback vs. visual feedback alone (Gonzalez et al., 2012). However,
471 in the same study, heart-rate variability had no significant effect for any of the three conditions tested.

472 Pupillometry

473 The task-evoked pupillary response successfully differentiated between the easy and hard tasks.
474 With some reservation, we recommend pupillometry as a viable method of measuring cognitive workload
475 during prosthesis use if the task can be modified for trial-averaged pupil responses. With no widely
476 accepted continuous measure of cognitive workload, the task had to be time-locked to perform trial
477 averaging. For the virtual target task, time-locking is straightforward; however, this is not the case with
478 many physical prosthesis tasks. Additionally, pupillometry requires controlled luminance, which adds
479 more complexity to experiment setup. Although pupillometry provided a robust response in the end, we
480 had to pilot the experiments extensively and carefully design our analyses to uncover the effect.
481 Pupillometry actually resulted in the largest effect size on an individual basis, but was not as consistent

482 across subjects. Setting up pupillometry was relatively simple; the head-mounted pupillometry system
483 used was nonintrusive, and robust to head movements.

484 Two other studies have used pupillometry as a measure of cognitive workload during a prosthesis
485 control comparison. One study showed that the number of pupillary increases was significantly different
486 for direct and classifier prosthesis control (White et al., 2017). The other study showed that average pupil
487 size was significantly different for a similar comparison (Zahabi et al., 2019).

488 EEG & ERP

489 EEG and ERPs were lacking in sensitivity and diagnostic ability. Although we find the measures
490 attractive, these barriers make it difficult to recommend using EEG & ERPs as reliable, easy-to-use
491 measures of cognitive workload. Frontal theta power, a measure of cognitive control (i.e., when a task
492 cannot be completed with an automatic, subconscious strategy) was close to a statistical trend across
493 subjects, but parietal alpha power and the P3 ERP were far from any across-subject statistical trend.
494 Alpha power was significantly different within-subject for five of ten participants, but the shift in power
495 was inconsistent, resulting in no across-subject trend. The P3 response was surprisingly consistent for the
496 amputee participant, highlighting the concept that different measures may work well for some persons but
497 not others. EEG and ERPs are appealing as they are direct measures of neural activity; however,
498 recording EEG and ERPs requires specialized training, time-consuming setup, and relatively expensive
499 equipment.

500 In a similar virtual prosthesis task (Deeny et al., 2014), the P300 ERP differed between passively
501 viewing the task and a hard condition, but there was no statistical difference between actively completing
502 the task under easy or hard conditions. In a physical task evaluating a sensory feedback system, alpha
503 power significantly differed between different feedback modalities (Gonzalez et al., 2012).

504 NASA TLX

505 We found that the NASA TLX worked well with the target matching task, as can be reasonably
506 expected when the task difficulty is quite obviously manipulated (i.e., it is very obvious to expect a small

507 target to be more difficult than a large target). We recommend the TLX as a cognitive workload measure
508 because of its simplicity, short duration, and widespread use. Because it is completed after a task is over
509 and depends entirely on subjective self-report, the TLX suffers from recall bias (Zahabi et al., 2019), task-
510 order dependency (McKendricka and Cherry, 2018) and other possible subjective biases. These effects
511 can generally be mitigated through proper experimental design and participant instruction. However,
512 some argue the TLX measures task difficulty more than it measures perceived mental workload
513 (McKendricka and Cherry, 2018).

514 Many prosthesis studies have employed the NASA TLX for comparing movement decoders
515 (Deeny et al., 2014; White et al., 2017; Osborn et al., 2021; Paskett et al., 2021) and sensory feedback
516 (Gonzalez et al., 2012; Markovic et al., 2018, 2020; Thomas et al., 2021). The TLX generally provides a
517 reliable response to changes in task difficulty.

518 Conclusion

519 This study utilizes several physiological, behavioral, and subjective cognitive workload measures
520 during a prosthesis task with known difficulty manipulations. Through collecting multiple measures
521 during the same task, the study enables researchers to comparatively evaluate the effectiveness and utility
522 of the various measures. Directly comparing several cognitive workload measures will aid
523 neuroprosthesis researchers in applying cognitive workload to their own studies. Overall, we recommend
524 the DRT, ECG, pupillometry, and EEG/ERPs, in that order, along with the traditional NASA TLX.
525 EEG/ERP measures typically were not reliably informative across subjects, although some EEG measures
526 worked well for a subset of individuals. Incorporating cognitive workload measures, and general user
527 experience, to neuroprosthesis studies provides a path for better, more intuitive neuroprostheses which
528 can more readily be translated to clinical realities.

529 Declarations

530 Ethics approval and consent to participate

531 Participants completed the study after providing informed consent. All experiments were
532 conducted under the oversight of the University of Utah IRB.

533 Consent for publication

534 Not applicable.

535 Availability of data and materials

536 The data that support the findings of this study are available from the corresponding author,
537 MDP, upon reasonable request.

538 Competing interests

539 JMC owns Red Scientific, which develops human factors research tools, including the DRT.
540 TSD, CCD, and GAC are inventors on a patent for decoding EMG motor signals. The remaining authors
541 declare that the research was conducted in the absence of any commercial or financial relationships that
542 could be construed as a potential conflict of interest.

543 Funding

544 This work was sponsored by: the Hand Proprioception and Touch Interfaces (HAPTIX) program
545 administered by the Biological Technologies Office (BTO) of the Defense Advanced Research Projects
546 Agency (DARPA) through the Space and Naval Warfare Systems Center, Contract No. N66001-15-C-
547 4017; the National Center for Advancing Translational Sciences of the National Institutes of Health under
548 Award Number ULTR002538 and TL1TR002540; and the National Institute of Neurological Disorders
549 and Stroke of the National Institutes of Health under Ruth L. Kirchstein National Research Service Award
550 Number 1F31NS118938.

551 Authors' Contributions

552 MDP performed background research, developed the software, built the custom DRT, developed
553 the experiments, conducted the experiments, analyzed the data, and wrote the manuscript. JKG performed
554 background research, conducted the experiments, and analyzed the data. STJ performed background
555 research, conducted the experiments, and analyzed the data. MRB developed the software, built the
556 custom DRT, and helped revise the manuscript. TSD developed the majority of the software and helped
557 revise the manuscript. CCD aided in participant recruitment, background research, and helped revise the
558 manuscript. JMC aided in experiment development and provided cognitive load measurement expertise.
559 DLS aided in experiment development and provided cognitive load measurement expertise. GAC
560 oversaw all aspects of the study.

561 Acknowledgments

562 We thank Dr. Brennan Payne, Dr. Trafton Drew, Sara LoTemplio, and Jack Silcox for their
563 advice and expertise with developing this study.

564 We thank Ripple Neuro, LLC for providing EEG caps compatible with their neural interface
565 processors, enabling the EEG recordings for this study.

566 References

- 567 Ameri, A., Akhaee, M. A., Scheme, E., and Englehart, K. (2019). Regression convolutional neural
568 network for improved simultaneous EMG control. *J. Neural Eng.* doi:10.1088/1741-2552/ab0e2e.
- 569 Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing
570 resources. *Psychol. Bull.* 91, 276–292. doi:10.1037/0033-2909.91.2.276.
- 571 Behar, J. A., Rosenberg, A. A., Weiser-Bitoun, I., Shemla, O., Alexandrovich, A., Konyukhov, E., et al.
572 (2018). PhysioZoo: A novel open access platform for heart rate variability analysis of mammalian
573 electrocardiographic data. *Front. Physiol.* 9, 1–14. doi:10.3389/fphys.2018.01390.

- 574 Behar, J., Johnson, A., Clifford, G. D., and Oster, J. (2014). A comparison of single channel fetal ecg
575 extraction methods. *Ann. Biomed. Eng.* 42, 1340–1353. doi:10.1007/s10439-014-0993-9.
- 576 Biddiss, E., Beaton, D., and Chau, T. (2007). Consumer design priorities for upper-limb prosthetics.
577 *Disabil. Rehabil. Assist. Technol.* 2, 346–357. Available at:
578 <http://search.ebscohost.com/login.aspx?direct=true&db=rzh&AN=105864881&site=ehost-live>.
- 579 Biddiss, E., and Chau, T. (2007a). Upper-limb prosthetics: Critical factors in device abandonment. *Am. J.*
580 *Phys. Med. Rehabil.* 86, 977–987. doi:10.1097/PHM.0b013e3181587f6c.
- 581 Biddiss, E., and Chau, T. (2007b). Upper limb prosthesis use and abandonment: A survey of the last 25
582 years. *Prosthet. Orthot. Int.* 31, 236–257. doi:10.1080/03093640600994581.
- 583 Chang, C. C., Boyle, L. N., Lee, J. D., and Jenness, J. (2017). Using tactile detection response tasks to
584 assess in-vehicle voice control interactions. *Transp. Res. Part F Traffic Psychol. Behav.* 51, 38–46.
585 doi:10.1016/j.trf.2017.06.008.
- 586 Charles, R. L., and Nixon, J. (2019). Measuring mental workload using physiological measures: A
587 systematic review. *Appl. Ergon.* 74, 221–232. doi:10.1016/j.apergo.2018.08.028.
- 588 D’Anna, E., Valle, G., Mazzoni, A., Strauss, I., Ibertie, F., Patton, J. J. J., et al. (2019). A closed-loop
589 hand prosthesis with simultaneous intraneural tactile and position feedback. *Sci. Robot.* 4.
590 doi:10.1126/scirobotics.aau8892.
- 591 Davoodi, R., and Loeb, G. E. (2011). MSMS software for VR simulations of neural prostheses and patient
592 training and rehabilitation. *Stud. Health Technol. Inform.* 163, 156–162. doi:10.3233/978-1-60750-706-2-
593 156.
- 594 Deeny, S., Chicoine, C., Hargrove, L., Parrish, T., and Jayaraman, A. (2014). A simple ERP method for
595 quantitative analysis of cognitive workload in myoelectric prosthesis control and human-machine
596 interaction. *PLoS One* 9. doi:10.1371/journal.pone.0112091.
- 597 Espinosa, M., and Nathan-Roberts, D. (2019). Understanding Prosthetic Abandonment. *Proc. Hum.*
598 *Factors Ergon. Soc. Annu. Meet.* 63, 1644–1648. doi:10.1177/1071181319631508.

599 Fisk, A. D., Derrick, W. L., and Schneider, W. (1983). Assessment of Workload: Dual Task
600 Methodology. in *Proceedings of the Human Factors and Ergonomics Society*, 229–233.

601 George, J. A., Brinton, M. R., Duncan, C. C., Hutchinson, D. T., and Clark, G. A. (2018). Improved
602 Training Paradigms and Motor-decode Algorithms: Results from Intact Individuals and a Recent
603 Transradial Amputee with Prior Complex Regional Pain Syndrome. in *40th International Engineering in
604 Medicine and Biology Conference* doi:10.1109/EMBC.2018.8513342.

605 George, J. A., Davis, T. S., Brinton, M. R., and Clark, G. A. (2020a). Intuitive neuromyoelectric control
606 of a dexterous bionic arm using a modified Kalman filter. *J. Neurosci. Methods* 330, 108462.
607 doi:10.1016/j.jneumeth.2019.108462.

608 George, J. A., Kluger, D. T., Davis, T. S., Wendelken, S. M., Okorokova, E. V., He, Q., et al. (2019).
609 Biomimetic sensory feedback through peripheral nerve stimulation improves dexterous use of a bionic
610 hand. *Sci. Robot.* 4, eaax2352. doi:10.1126/scirobotics.aax2352.

611 George, J. A., Neibling, A., Paskett, M. D., and Clark, G. A. (2020b). Inexpensive surface
612 electromyography sleeve with consistent electrode placement enables dexterous and stable prosthetic
613 control through deep learning. *41st Int. Eng. Med. Biol. Conf.* 2020. Available at:
614 <http://arxiv.org/abs/2003.00070>.

615 Gonzalez, J., Soma, H., Sekine, M., and Yu, W. (2012). Psycho-physiological assessment of a prosthetic
616 hand sensory feedback system based on an auditory display: A preliminary study. *J. Neuroeng. Rehabil.*
617 9, 1–14. doi:10.1186/1743-0003-9-33.

618 Graczyk, E. L., Resnik, L., Schiefer, M. A., Schmitt, M., and Tyler, D. J. (2018). Home use of a neural-
619 connected sensory prosthesis provides the functional and psychosocial experience of having a hand again.
620 *Sci. Rep.* In press, 1–17. doi:10.1038/s41598-018-26952-x.

621 Hargrove, L. J., Miller, L. A., Turner, K., and Kuiken, T. A. (2017). Myoelectric Pattern Recognition
622 Outperforms Direct Control for Transhumeral Amputees with Targeted Muscle Reinnervation: A
623 Randomized Clinical Trial. *Sci. Rep.* 7, 1–9. doi:10.1038/s41598-017-14386-w.

624 Hart, S. G., and Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of
625 Empirical and Theoretical Research. *Adv. Psychol.* doi:10.1016/S0166-4115(08)62386-9.

626 Hess, E. H., and Janisse, M. P. (1978). Pupillometry: The Psychology of the Pupillary Response. *Am. J.*
627 *Psychol.* 91. doi:10.2307/1421703.

628 ISO 17488:2016 (2016). Road vehicles — Transport information and control systems — Detection-
629 response task (DRT) for assessing attentional effects of cognitive load in driving. Geneva, Switzerland:
630 International Organization for Standardization.

631 Kahneman, D., and Beatty, J. (1966). Pupil diameter and load on memory. *Science (80-.)*. 154, 1583–
632 1585. doi:10.1126/science.154.3756.1583.

633 Keil, A., Mussweiler, T., and Epstude, K. (2006). Alpha-band activity reflects reduction of mental effort
634 in a comparison task: A source space analysis. *Brain Res.* 1121, 117–127.
635 doi:10.1016/j.brainres.2006.08.118.

636 Lohani, M., Payne, B. R., and Strayer, D. L. (2019). A Review of Psychophysiological Measures to
637 Assess Cognitive States in Real-World Driving. *Front. Hum. Neurosci.* 13, 1–27.
638 doi:10.3389/fnhum.2019.00057.

639 Luck, S. (2005). *An Introduction to the Event-Related Potential*. 1st ed. The MIT Press.

640 Markovic, M., Schweisfurth, M. A., Engels, L. F., Bentz, T., Wustefeld, D., Farina, D., et al. (2018). The
641 clinical relevance of advanced artificial feedback in the control of a multi- functional myoelectric
642 prosthesis. *J. Neuroeng. Rehabil.* 15. doi:10.1364/nlo.2007.we2.

643 Markovic, M., Varel, M., Schweisfurth, M. A., Schilling, A. F., and Dosen, S. (2020). Closed-Loop
644 Multi-Amplitude Control for Robust and Dexterous Performance of Myoelectric Prosthesis. *IEEE Trans.*
645 *Neural Syst. Rehabil. Eng.* 28, 498–507. doi:10.1109/TNSRE.2019.2959714.

646 Mastinu, E., Engels, L., Clemente, F., Dione, M., Sassu, P., Aszmann, O., et al. (2020). Neural feedback
647 strategies to improve grasping coordination in neuromusculoskeletal prostheses. *Sci. Rep.*, 1–14.
648 doi:10.1038/s41598-020-67985-5.

- 649 McKendricka, R. D., and Cherry, E. (2018). A deeper look at the NASA TLX and where it falls short.
650 *Proc. Hum. Factors Ergon. Soc.* 1, 44–48. doi:10.1177/1541931218621010.
- 651 Nieveen, J., Warren, D., Wendelken, S., Davis, T., Kluger, D., and Page, D. (2017). Channel Selection of
652 Neural And Electromyographic Signals for Decoding of Motor Intent. in *Myoelectric Controls*
653 *Conference*, 720.
- 654 Ortiz-Catalan, M., Håkansson, B., and Brånemark, R. (2014a). An osseointegrated human-machine
655 gateway for long term sensory feedback and control of artificial limbs. *Sci. Transl. Med.* 6, 1–9.
656 doi:10.1126/scitranslmed.3008933.
- 657 Ortiz-Catalan, M., Håkansson, B., and Brånemark, R. (2014b). Real-time and simultaneous control of
658 artificial limbs based on pattern recognition algorithms. *IEEE Trans. Neural Syst. Rehabil. Eng.* 22, 756–
659 764. doi:10.1109/TNSRE.2014.2305097.
- 660 Osborn, L. E., Moran, C., Johannes, M., Sutton, E., Wormley, J., Dohopolski, C., et al. (2021). Extended
661 home use of an advanced osseointegrated prosthetic arm improves function, performance, and control
662 efficiency. *J. Neural Eng.* doi:10.1088/1741-2552/abe20d.
- 663 Parr, J. V. V., Vine, S. J., Wilson, M. R., Harrison, N. R., and Wood, G. (2019). Visual attention, EEG
664 alpha power and T7-Fz connectivity are implicated in prosthetic hand control and can be optimized
665 through gaze training. *J. Neuroeng. Rehabil.* 16, 1–20. doi:10.1186/s12984-019-0524-x.
- 666 Paskett, M. D., Brinton, M. R., Hansen, T. C., George, J. A., Davis, T. S., Duncan, C. C., et al. (2021).
667 Activities of daily living with bionic arm improved by combination training and latching filter in
668 prosthesis control comparison. *J. Neuroeng. Rehabil.* 18. doi:10.1186/s12984-021-00839-x.
- 669 Payne, D. T., Parry, M. E., and Harasymiw, S. J. (1968). Percentage of pupillary dilation as a measure of
670 item difficulty. *Percept. Psychophys.* 4, 139–143. doi:10.3758/BF03210453.
- 671 Pons, J. L., Ceres, R., Rocon, E., Reynaerts, D., Saro, B., Levin, S., et al. (2005). Objectives and
672 technological approach to the development of the multifunctional MANUS upper limb prosthesis.
673 *Robotica* 23, 301–310. doi:10.1017/S0263574704001328.

674 Ranney, T. A., Baldwin, G. H. S., Smith, L. A., Mazzae, E. N., & Pierce, R. S. (2014). Detection
675 Response Task (DRT) Evaluation for Driver Distraction Measurement Application. *U.S. Dep. Transp.*
676 *Natl. Highw. Traffic Saf. Adm.*

677 Raveh, E., Friedman, J., and Portnoy, S. (2018a). Evaluation of the effects of adding vibrotactile feedback
678 to myoelectric prosthesis users on performance and visual attention in a dual-task paradigm. *Clin.*
679 *Rehabil.* 32, 1308–1316. doi:10.1177/0269215518774104.

680 Raveh, E., Friedman, J., and Portnoy, S. (2018b). Visuomotor behaviors and performance in a dual-task
681 paradigm with and without vibrotactile feedback when using a myoelectric controlled hand. *Assist.*
682 *Technol.* 30, 274–280. doi:10.1080/10400435.2017.1323809.

683 Resnik, L., Meucci, M. R., Lieberman-Klinger, S., Fantini, C., Kelty, D. L., Disla, R., et al. (2012).
684 Advanced upper limb prosthetic devices: Implications for upper limb prosthetic rehabilitation. *Arch.*
685 *Phys. Med. Rehabil.* 93, 710–717. doi:10.1016/j.apmr.2011.11.010.

686 Salminger, S., Stino, H., Pichler, L. H., Gstoettner, C., Sturma, A., Mayer, J. A., et al. (2020). Current
687 rates of prosthetic usage in upper-limb amputees—have innovations had an impact on device acceptance?
688 *Disabil. Rehabil.* doi:10.1080/09638288.2020.1866684.

689 Salminger, S., Sturma, A., Hofer, C., Evangelista, M., Perrin, M., Bergmeister, K. D., et al. (2019). Long-
690 term implant of intramuscular sensors and nerve transfers for wireless control of robotic arms in above-
691 elbow amputees. *Sci. Robot.* 4. doi:10.1126/scirobotics.aaw6306.

692 Schofield, J. S., Shell, C. E., Beckler, D. T., Thumser, Z. C., and Marasco, P. D. (2019). Long-term
693 home-use of sensory-motor-integrated bidirectional bionic prosthetic arms promotes functional,
694 perceptual, and cognitive changes. *Front. Neurosci.* In Review, 1–20. doi:10.3389/fnins.2020.00120.

695 Shaw, E. P., Rietschel, J. C., Hendershot, B. D., Pruziner, A. L., Wolf, E. J., Dearth, C. L., et al. (2019). A
696 Comparison of Mental Workload in Individuals with Transtibial and Transfemoral Lower Limb Loss
697 during Dual-Task Walking under Varying Demand. *J. Int. Neuropsychol. Soc.* 25, 985–997.
698 doi:10.1017/s1355617719000602.

- 699 Stojmenova, K., Jakus, G., and Sodnik, J. (2017). Sensitivity evaluation of the visual, tactile, and auditory
700 detection response task method while driving. *Traffic Inj. Prev.* 18, 431–436.
701 doi:10.1080/15389588.2016.1214868.
- 702 Stojmenova, K., and Sodnik, J. (2018). Detection-response task—uses and limitations. *Sensors* 18.
703 doi:10.3390/s18020594.
- 704 Strayer, D. L., Cooper, J. M., Turrill, J., Coleman, J. R., and Hopman, R. J. (2017). The smartphone and
705 the driver’s cognitive workload: A comparison of Apple, Google, and Microsoft’s intelligent personal
706 assistants. *Can. J. Exp. Psychol.* 71, 93–110. doi:10.1037/cep0000104.
- 707 Strayer, D. L., Turrill, J., Coleman, J. R., Ortiz, E. V, and Cooper, J. M. (2014). Measuring Cognitive
708 Distraction in the Automobile II: Assessing In-Vehicle Voice-Based Interactive Technologies. *AAA*
709 *Found. Traffic Saf.* Available at: www.aaafoundation.org.
- 710 Tan, D. W., Schiefer, M. A., Keith, M. W., Anderson, J. R., Tyler, J., and Tyler, D. J. (2014). A neural
711 interface provides long-term stable natural touch perception. *Sci. Transl. Med.* 6.
712 doi:10.1126/scitranslmed.3008669.
- 713 Thomas, N., Ung, G., Ayaz, H., and Brown, J. D. (2021). Neurophysiological Evaluation of Haptic
714 Feedback for Myoelectric Prostheses. *IEEE Trans. Human-Machine Syst.* 51, 253–264.
715 doi:10.1109/THMS.2021.3066856.
- 716 Thomas, N., Ung, G., McGarvey, C., and Brown, J. D. (2019). Comparison of vibrotactile and joint-
717 torque feedback in a myoelectric upper-limb prosthesis. *J. Neuroeng. Rehabil.* 16, 1–18.
718 doi:10.1186/s12984-019-0545-5.
- 719 Valle, G., D’Anna, E., Strauss, I., Clemente, F., Granata, G., Di Iorio, R., et al. (2020). Hand Control
720 With Invasive Feedback Is Not Impaired by Increased Cognitive Load. *Front. Bioeng. Biotechnol.* 8, 1–7.
721 doi:10.3389/fbioe.2020.00287.
- 722 Vu, P. P., Vaskov, A. K., Irwin, Z. T., Henning, P. T., Lueders, D. R., Laidlaw, A. T., et al. (2020). A
723 regenerative peripheral nerve interface allows real-time control of an artificial hand in upper limb
724 amputees. *Sci. Transl. Med.* 12, 1–12. doi:10.1126/scitranslmed.aay2857.

725 White, M. M., Zhang, W., Winslow, A. T., Zahabi, M., Zhang, F., Huang, H., et al. (2017). Usability
726 Comparison of Conventional Direct Control Versus Pattern Recognition Control of Transradial
727 Prostheses. *IEEE Trans. Human-Machine Syst.* 47, 1146–1157. doi:10.1109/THMS.2017.2759762.
728 Witteveen, H. J. B., de Rond, L., Rietman, J. S., and Veltink, P. H. (2012). Hand-opening feedback for
729 myoelectric forearm prostheses: Performance in virtual grasping tasks influenced by different levels of
730 distraction. *J. Rehabil. Res. Dev.* 49, 1517–1526. doi:10.1682/JRRD.2011.12.0243.
731 Zahabi, M., White, M. M., Zhang, W., Winslow, A. T., Zhang, F., Huang, H., et al. (2019). Application of
732 Cognitive Task Performance Modeling for Assessing Usability of Transradial Prostheses. *IEEE Trans.*
733 *Human-Machine Syst.* 49, 381–387. doi:10.1109/THMS.2019.2903188.
734
735