Human-Centered Evaluation of EMG-based Upper-limb Prosthetic Control Modes

Yunmei Liu, Joseph Berman, Albert Dodson, Junho Park, Maryam Zahabi, *Member, IEEE*, He Huang, *Senior Member, IEEE*, Jaime Ruiz, and David Kaber, *Fellow, IEEE*

***Abstract*—The aim of this study was to experimentally test the effects of different electromyographic (EMG)-based prosthetic control modes on user task performance, cognitive workload, and perceived usability, to inform further human-centered design and application of these prosthetic control interfaces. We recruited 30 able-bodied participants for a between-subjects comparison of three control modes: direct control (DC), pattern recognition (PR), and continuous control (CC). Multiple human-centered evaluations were used, including task performance, cognitive workload, and usability assessments. To ensure that results were not task-dependent, this study used two different test tasks, including the Clothespin Relocation Task (CRT) and Southampton Hand Assessment Procedure-Door Handle Task (SHAP-DHT). Results revealed performance with each control mode to vary among tasks. When the task had high angle adjustment accuracy requirements, the PR control outperformed DC. For cognitive workload, the CC mode was superior to DC in reducing user load across tasks. Both CC and PR control appear to be effective alternatives to DC in terms of task performance and cognitive load. Furthermore, we observed that when comparing control modes, multi-task testing and multi-faceted evaluations are critical to avoid task-induced or method-induced evaluation bias. Hence, future studies with larger samples and different designs will be needed to expand the understanding of prosthetic device features and workload relationships.**

***Index Terms*—Electromyography, prosthetics, human-centered design, prosthetic control**

I. Introduction

Upper limb amputation causes a permanent disability. Basic activities of daily living, such as grasping, eating, and using zippers on clothing become difficult to perform for individuals with upper limb amputations [1]. To restore their motor function and improve the quality of life, advanced prosthesis technology is needed. In the past decade, significant technological advances have made powered, dexterous prosthetic hands and arms commercially available. The key challenge in making these modern devices functional for upper limb amputees is an intuitive human-machine interface for easy prosthesis operation. Since electromyographic (EMG) signals represent user's movement intent, EMG signals recorded from residual muscles have been widely used as neural sources in human-machine interfaces for powered prosthetic arm control [2].

A. Prosthetic Control Modes

There are several EMG-based control modes described in the literature that map EMG signals to prosthesis control commands. Direct control (DC) has been widely used in clinics for decades [3]. DC allows users to control one degree of freedom (DOF) from EMG signals recorded from a single pair of residual agonist-antagonist muscles. If more than one DOF needs to be operated, the user must generate a special muscle activation pattern, e.g., co-contraction, to switch the control DOF and then use the same muscle pair to control another joint, which is non-intuitive. Another control mode, EMG pattern recognition (PR) [4], has become clinically available in recent years. PR control recognizes a user's intent based on the activation patterns of residual muscles. At the core of PR control is a classifier that maps features of multichannel EMG signals to discrete motion classes (i.e., hand open/close, wrist pronation/supination). Compared to DC, PR control is more intuitive to use. However, this method still only allows for the control of one DOF at a time, leading to unnatural arm/hand motions. To address this limitation, recent research has focused on developing EMG-based control that can estimate multi-joint coordinated, continuous arm kinematics (i.e., position or velocity) for continuous control (CC) of multifunctional prostheses [5-12]. The capability of CC to drive multiple DOFs simultaneously may allow users to adopt more natural motion strategies to efficiently complete tasks. Since CC is relatively new, it has not been adopted by commercial prostheses yet. Theoretically, CC is more natural to use. However, there are gaps in objective and subjective comparison of CC with other widely used control modes, such as DC and PR control.

Despite tremendous engineering efforts in developing EMG-based control modes for prosthetics, human-centered research on how the different modes impact user cognition and physical task performance has been limited. Rehabilitation engineers often evaluate EMG-based control methods by the EMG decoding accuracy in estimating user intent [6-14]. Task performance of users wearing EMG-controlled prostheses is sometimes included in these evaluations just to show the feasibility of a new design. Cognitive function effects have rarely been quantified.

B. Human-Centered Evaluation Measures

For assistive technologies, human-centered evaluations are critical to understand how a device may be interpreted by a user and to ensure the intended benefits for users. Such evaluations quantify not only the performance of user tasks assisted by the technology, but also the demand on cognitive function (i.e., workload) and usability. Despite the importance of human-centered evaluations, related research that uses human-centered approaches to evaluate different EMG control modes for prostheses is quite limited.

Several clinical translational studies compared DC and PR control modes. These evaluations have included performance of prosthesis users in various tasks and cognitive load [15-17]. In these clinical trials, the evaluation of cognitive load is primarily addressed by dual-task paradigms, using specific tasks with overlapping or competing demands. In our previous research, we evaluated the physical performance and associated cognitive load of human participants who used powered upper-extremity prosthetics with either DC or PR control while performing the Clothespin Relocation Task (CRT) [18-20].

Regarding cognitive load assessment, a wide variety of metrics are documented in the literature, primarily due to the multidimensional nature of the mental workload construct [21]. Several studies classify measures of cognitive load into three categories: physiological state, subjective reports, and dual-task paradigms [21,22]. The drawback of dual-task paradigms is that they can distract/subtract from primary task performance, which is undesirable when studying the effectiveness of task interfaces [22]. Based on the literature, there are pros and cons for both subjective reports of workload and objective physiological indicators. Some subjective reporting methods have been demonstrated to be usable and effective for capturing perceptions of workload, but inevitably have a bias due to memory dependencies and the influence of perceived task performance [21,23]. Opposite to this, physiological state measures have shown advantages in providing objective indicators of cognitive load with strong correlations between pupillometry features and objective measures of task demands. However, physiological measures can also be influenced by the task environment and other cognitive demands, such as anxiety, stress, and sexual arousal [23]. Given the limitations of each technique, we elected to use a combination of subjective reports and physiological measures, including eye-tracking responses, which are non-invasive and highly sensitive to cognitive task demands. The percent change in pupil size (PCPS) has been used in previous research to assess the effect of prosthetic device control modes on cognitive load [18]. It was found that PCPS has a higher value in mentally complex tasks than in more manageable tasks [24]. Blink rate (BR) has also been frequently used as an indicator of cognitive load in other domains [25,26]. However, some studies suggest that this measure is only adequate to assess visual workload [27]. The number of eye blinks and blink duration decrease as visual workload increases [27]. In addition, to these eye-tracking measures, we also applied the NASA Task Load Index (TLX) as a subjective workload report method, as this index has been used extensively in prior research, including perceived workload in prosthesis device use [28-30].

The existing human-centered evaluation studies reveal several knowledge gaps. First, none of the previous studies considered CC as a comparative design for EMG-based prosthetic control. Second, when evaluating cognitive load, only one test task has been used, which may challenge the generalizability of study conclusions. In addition, the analysis of cognitive load has been based on one type of physiological measure, such as pupil size, electro-encephalography signals, heart rate, etc., or subjective measures. Each measure has its own challenges, and there is no "gold standard" among cognitive assessment methods. Finally, to our knowledge, none of the prior studies include usability assessment [18-20].

C. Objective and Hypotheses

Motivated by the need to address the identified knowledge gaps, the objective of this paper was to experimentally compare three existing EMG-based control modes for upper limb prostheses (DC, PR control and CC) based on multiple test tasks (CRT and Southampton Hand Assessment Procedure (SHAP) - Door Handle Task (DHT)) and multifaceted human-centered evaluation, including: task performance, cognitive load, and usability assessment.

Based on the efficiency of control mode algorithms and previous studies [18,19], PR control and CC were expected to produce superior task performance (Hypothesis (H)1), lower cognitive workload (H2), and greater perceived usability (H3) than DC. In addition, since the SHAP-DHT requires fewer gestures and simpler operations, as compared to the CRT, it was expected that participants would experience higher cognitive workload in CRT vs. SHAP-DHT performance (H4). In general, the results of this study were expected to highlight the advantages and challenges of each EMG control mode and inform future design of EMG-based interfaces for upper limb prostheses. Furthermore, this study was expected to define an effective evaluation protocol and metrics to assess EMG-based control interfaces.

II. Method

A. Participants

The experiment protocol was approved by the Institutional Review Board of the University of North Carolina at Chapel Hill. Thirty-six participants without disabilities were recruited. Persons with prior experience using a prosthetic device and researchers involved in the study were excluded from participation. All participants were assigned to either DC, PR control or CC control at random on a rolling basis while ensuring an equal number assigned to each control mode and balancing of the sample for gender. However, data from six participants were excluded from the analysis. Five of these participants were unable to complete the control mode training. Specifically, two participants were unable to complete the PR mode training, two were unable to complete the CC mode training, and one was unable to complete the DC mode training. Moreover, we had to exclude data from another participant whose notably long eyelashes interfered with the eye-tracking system data collection, causing near-zero confidence levels for observations. Consequently, 10 participants were included in the data analysis for each of the three control modes. Out of the 30 participants, 12 were females. The mean and standard deviation of age across participants was 22.9 years and 2.8 years, respectively. All participants consented to the study and were compensated for participation.

B. EMG Control Modes and Setup

One commercial prosthetic hand [ETD, Motion Control Inc., USA], with 2-DOF of actuation in hand open/close and wrist pronation/supination, was used to test all three control algorithms in the CRT and SHAP-DHT. The EMG system was equipped with an anti-aliasing low-pass filter set at 500Hz for all data collection. The total weight of the device, which encompasses the hand, able-bodied adapter, and cable components, was approximately 4.54 pounds. A custom prosthetic hand adapter was designed and fabricated such that people without amputation could control the prosthetic hand with their right arm. The adapter was locked at 90 degrees at the elbow but allowed for a full range of motion at the hand and for pronation/supination. Table I shows the setup for each of the three control modes. The utilization of muscles for control of prosthetic limbs varies depending on the control mode employed. DC is a simple, non-machine learning method where a single agonist-antagonist muscle pair is used to control a single degree of freedom, such as the wrist or hand. The degree of freedom is switched through co-contraction. EMG signals from the flexor carpi radialis (FCR) and extensor carpi radialis longus (ECRL) were used for DC. On the other hand, PR and CC modes require multiple muscle signals to control multiple degrees of freedom for a more intuitive control experience. These methods use all four channels to predict the movement and velocity of joints, eliminating the need for manual switching between degrees of freedom. Therefore, two additional EMG signals, measured from the flexor digitorum superficialis (FDS) and the extensor digitorum communis (EDC) muscles, were used for PR control and CC. Electrode placements were based on muscle palpation and confirmation of EMG signal quality. An EMG system (MA400, Motion Lab Systems Inc., USA) captured signals at 1000 Hz. The control algorithms for the DC and PR control modes have been reported previously [15,19]. Here we briefly discuss their implementation.

Direct Control: DC utilizes a pair of agonist-antagonist muscles to control a single DOF (either hand open/close or wrist pronation/supination) at one time. The movement speed of the prosthetic hand is set proportional to the magnitude of measured EMG signals. Switching from one DOF to another is accomplished with a co-contraction of the muscle pair (closing the hand and clinching a fist). To enhance the integrity of the EMG signals, high-pass (20Hz) and low-pass (450Hz) filters were applied before further processing with a sliding window analysis technique. The analysis window had a size of 150ms with a 50ms overlap, and the increment between consecutive windows was 100ms. The latency for the DC control mode was half of the window size, ensuring a responsive and efficient control experience.

Pattern Recognition control: PR control classifies movement intent based on patterns of EMG features. In this study, five movement classes were included: hand close, hand open, wrist pronation, wrist supination, and no movement. To extract meaningful information from the EMG signals, four time-domain features were calculated, including mean absolute values (MAVs), number of zero crossings, waveform length, and number of slope sign changes. Similar to the DC mode, the collected EMG signals were filtered with 20Hz and 450Hz bandpass filters and processed using a sliding window technique with a 150ms window size, a 50ms overlap, a 100ms increment between consecutive windows, and half of the window size latency. The collected data were used to train a linear discriminant analysis (LDA) classifier. Intended movement classes were predicted in real-time by the classifier while the movement speed was set proportional to the magnitudes of EMG signals.

Continuous Control: In this mode, EMG data are recorded simultaneously with kinematic data from a Leap Motion Controller (Leap Motion, Inc., USA). The controller uses a camera to accurately estimate positions of hand and forearm segments [31-34]. Position estimates for phalangeal, palm and forearm segments are recorded at 120 Hz and are used to estimate wrist pronation/ supination and metacarpophalangeal (MCP) flexion/ extension joint angles. Additionally, our preprocessing for CC incorporated a MAV sliding window, which served as a straightforward form of low-pass filtering. This approach provided necessary refinement without compromising the integrity of the data. Fig. 1 graphically presents the EMG CC mode used in this study.

TABLE I
PARAMETER COMPARISON AMONG THREE CONTROL MODES

Control mode	Direct control	Pattern recognition control	Continuous control
Control method	N/A	LDA	ANN
Number of EMG channels	2	4	4
Number of control gestures	3	4	4
DOF per control	1	1	Multiple
Co-contraction is required to switch control DOF	Yes	No	No
Sequential control	Yes	Yes	No

Note: DOF - "degree of freedom"; N/A - "not applicable"; LDA - "linear discriminant analysis"; ANN - "artificial neural networks".

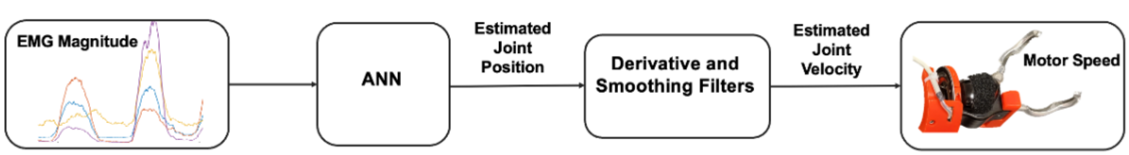

Figure 1. The control process for the CC mode.

Data were first collected from participants while they performed MCP flexion and extension and wrist pronation and supination movements. All motions were performed in a pattern in which participants moved their wrist/MCP between a fully flexed/pronated, relaxed, and fully extended/supinated position to a metronome set at a 1Hz frequency. Three 10-second trials were recorded for each motion type to be used for training.

An artificial neural network (ANN) was created for each participant for both the wrist and MCP using the Deep Learning Toolbox in MATLAB 2018b (Mathworks Inc., USA). The ANNs were trained to map processed EMG signals to joint positions. Velocity was estimated by differentiating the estimated positions:

$$\dot{\theta}(t) = \frac{\theta(t) - \theta(t-\Delta t)}{\Delta t}$$

where θ and $\dot{\theta}$ are the estimated position and velocity respectively and t is the current timestep. The value of Δt was tuned to 100 ms to provide relatively smooth velocity estimations. Additionally, estimated kinematics were smoothed in real time using a moving average filter. Finally, to prevent constant small motions in the prosthetic device, all estimated velocities that fell below a predetermined threshold were set to 0.

To train the ANN, we collected three sets of 10-second duration data for each type of movement, amounting to a total of 90 seconds of data per participant. More specifically, we acquired three data files for isolated hand movements, three for isolated wrist movements, and three for simultaneous movements. The collected data were then divided using 70% for ANN training, 15% for validation, and 15% for testing. For each participant, multiple parameters were manually adjusted by experimenters based on offline performance achieved by the ANN and on the feedback from participants testing basic motions in real time using the prosthetic device. The number of hidden layers and then the number of neurons in each hidden layer were incremented until maximum performance was found. The optimal size for the sliding windows was then determined by calculating the MAV for windows in the range of 100ms to 300ms with a 100ms increment. The size of the sliding windows used for the moving average filter was determined by gradually increasing values starting from 0ms (no smoothing) until participants indicated they could comfortably control the device. If needed, the thresholds were incremented starting from zero while subjects relaxed their upper limb until any small movements of the prosthetic motors were eliminated. The latency for the CC control mode was half of the window size. Lastly, based on participant feedback, an output gain was adjusted to allow the motors to move at an appropriate speed proportional to the estimated velocity.

C. Tasks Selection and Experimental Setup

In our prior research [35], we investigated the relationship between human upper body movements in real-world activities, such as driving a vehicle and activities of daily living (ADL), as documented in the literature. Our study determined that the CRT and SHAP-DHT encompass all upper body movements necessary for such activities (including shoulder, elbow, forearm, wrist, and hand movements) and were identified as the most sensitive ADL tests in previous research assessing the usability of prosthetic devices. It is important to note that while we did not collect every combination of input gestures and classified outcomes, we implemented a consistent calibration process for all participants to minimize potential confounding factors arising from differences in calibration quality. Furthermore, all participants met the training criteria for each device configuration, ensuring that any unintentional correct hook movements for task completion, such as closing, occurred consistently across the entire participant pool.

Clothespin Relocation Task: The CRT [36] consists of a bin with a horizontal bar, vertical bar, and three plastic clothespins (Fig. 2). The objective of the task is for participants to use two DOFs in motion to transport clothespins between the two bars, as quickly as possible.

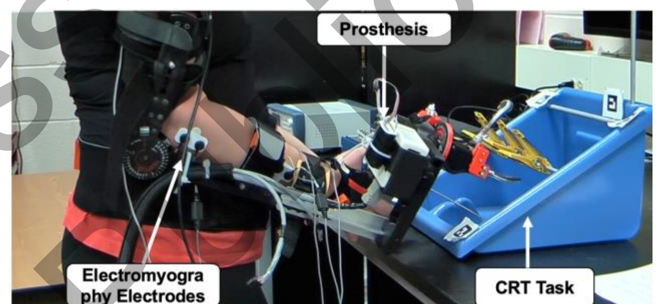

Figure 2. A participant using the prosthetic adapter for CRT task performance.

Southampton Hand Assessment Procedure: The door handle task (Fig. 3), as part of the SHAP-DHT [37], was also required in this study. The task involves grasping the handle, rotating it downwards and then back to its original position, and releasing it as quickly as possible.

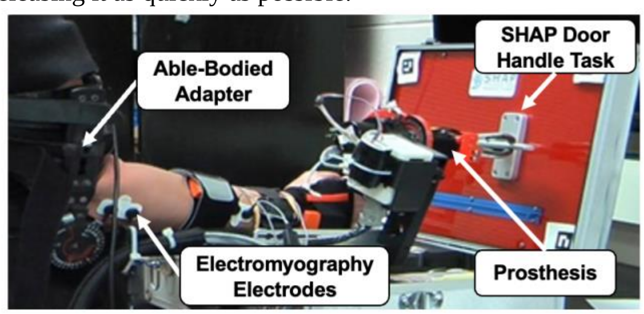

Figure 3. A participant using the prosthetic adapter for the SHAP-DHT.

D. Human-Centered Evaluation Methods and Metrics

To evaluate the performance of the three control modes in the two different tasks, we defined performance criteria for each of the tasks. CRT performance was measured in terms of the number of pins moved by a participant in a 2-minute period. The easier a control mode was to use, the more pins a participant could move and the better the task performance evaluation for the control mode. For the SHAP-DHT, we measured the time it took participants to rotate the door handle five times in sequence. The more difficult a control mode was to use, the longer the task time and the worse the control mode evaluation. In this study, we used eye-tracking measures (e.g., BR and PCPS) combined with NASA-TLX scores to compare the cognitive load implications of different device control modes for two degree-of-freedom myoelectric prosthetics.

With respect to usability assessment, we evaluated several commonly used methods, such as the Modified Client Satisfaction with Device module of the Orthotics and Prosthetic Users' Survey (CSD-OPUS), the System Usability Scale (SUS), the Trinity Amputation and Prosthesis Experience Scale (TAPES), the Quebec User Evaluation of Satisfaction with assistive Technology (QUEST 2.0) method, and the Usefulness, Satisfaction and Ease of Use (USE) scale. Since this research involved assessment of prosthetic device control modes rather than a comparison of existing commercial products, the QUEST 2.0 and USE scales were identified as stronger fits for the study than the CSD-OPUS, SUS, and TAPES. The QUEST 2.0 method was considered to be particularly relevant, as the survey includes an assistive device section, specifically designed for prosthetic devices. The overall structure of our human-centered evaluation of the EMG-based upper-limb prosthetic control modes is presented in Fig. 4.

E. Procedures

The experiment took place in a laboratory with constant illumination to limit the effect of fluctuations in lighting level on pupillometry. Based on multiple photometer readings during the experiment, the illuminance level was relatively consistent at 170–200 lux in the area where participants were tested. Participants wore a head-mounted eye-tracking system throughout the experiment to capture changes in their pupil size and blink rate. Participants also donned the upper-extremity prosthetic adapter. EMG electrodes were placed on the skin surface based on the assigned control mode. A verbal description of the prosthesis DOFs and control strategy was provided for participants. They practiced controlling the device until they reported comfort with the assigned control mode. Participants then advanced to a formal study training period.

The training session required participants to use the prosthesis to perform the CRT task while an experimenter recorded the time to move three clothespins. Training criteria were established based on pilot test data and a learning curve analysis revealing when participants achieve asymptotic performance (on average) with the specific device and at what level (i.e., minimum task time). If the average task completion time of three sequential trials was within 15–25s for the PR control, 20–35s for the DC and 16-23s for the CC mode, the participant passed the training sessions and proceeded to actual experimental trials.

Upon completion of the training trials, the eye-tracking system was calibrated for participants, and they were permitted to begin experiment trials after a 5-minute break.

During the experiment, all participants completed a total of three trials for both the CRT and SHAP-DHT. After each test trial, participants filled-out the NASA-TLX questionnaire and had a 2-min rest period. The QUEST 2.0 and USE forms were presented after a participant had completed all trials for a specific task and were used to collect usability assessments of the control mode during the experiment.

F. Experiment Design

This study followed a 3×2 mixed within-between-subjects experiment design, with three prosthetic control modes (DC, PR control and CC) and two types of test tasks (CRT and SHAP-DHT). Each participant was assigned a unique control mode to perform the CRT and SHAP-DHT, each of which was repeated three times. Therefore, our experimental design ensured an equal number of observations across device control modes and tasks.

G. Data Processing

For the CRT, the number of pins moved in a 2-minute period was recorded. For the SHAP-DHT, the time required to rotate the door handle 5 times was recorded.

Although we strictly controlled the lab illumination level and performed calibration of the eye-tracking system, such devices are highly sensitive to environmental conditions and physiology states. Consequently, we applied rigorous post-processing to the tracking data. Pupil size estimation depends on an accurate 3D eye model. Unfortunately, headset slippage during testing can seriously affect pupillometry. To address this issue, we used the approach of regularly updating the 3D model for each participant [38]. However, in some cases, this periodic update process can lead to increases in the likelihood of incorrect pupil size measurements, specifically if there is no slippage in the headset [38]. To further address this issue, we analyzed the system confidence level in pupil size (or data quality). When values were less than 0.9 (90% confidence), the data were excluded for experimental evaluation. This confidence level was determined based on discussions with the manufacturer (Pupil Labs) and has been used in prior studies with eye-tracking devices [19,39]. In addition, we watched recorded videos of test trials to assess the device calibration for pupil tracking and to identify inaccurate 3D eye models. This procedure was recommended by the manufacturer and has been applied to other studies [40,41]. We then calculated the PCPS by subtracting the baseline pupil size (collected prior to the experiment and when the participant was looking at a black monitor display in a relaxed state) from the measured pupil size in each trial and then divided by the baseline pupil size. BR is defined as the number of eye closures in a given period [19]. The eye-tracking system we used automatically captures blink information, including each blink time and blink duration. Based on this information, we used blinks per minute to represent BR.

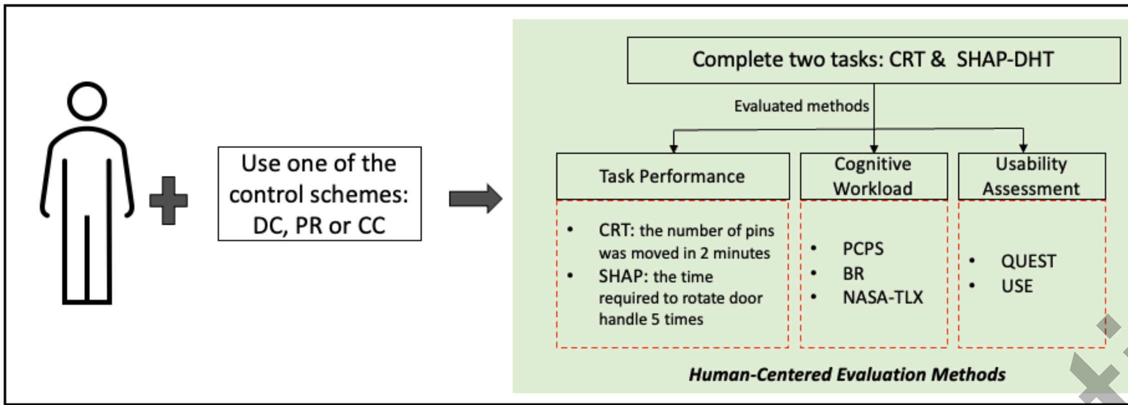

Figure 4. Structure of human-centered evaluation of EMG-based upper-limb prosthetic control modes.

Regarding the subjective workload measure, the NASA-TLX includes six dimensions of demand, including: mental, physical, temporal, performance, effort, and frustration [42]. Participants were required to rank each dimension at the start of the experiment, based on their task training session. Subsequently, after each trial, participants were asked to further rate the dimensions for the specific test conditions. The NASA-TLX total score was calculated as a rank-weighted sum of demand ratings from 0 to 100.

Usability was measured using the two questionnaires, including QUEST 2.0 and USE. Although the QUEST 2.0 and USE surveys are both usability quantification methods, they focus on different aspects of the target technology. The QUEST 2.0 questionnaire assesses user satisfaction with the assistive technology across eight different dimensions, including: device size, weight, durability, comfort, effectiveness, ease of use, ease in adjustment, safety, and security [43]. The USE survey is a composite assessment including four components of usefulness, satisfaction, ease of use, and ease of learning. The USE satisfaction component, for example, presents seven questions, such as whether the user is satisfied, whether they would recommend the device to a friend, whether the device is "fun to use", and other questions [44]. In general, the USE survey focuses more on user experience evaluation, while the QUEST 2.0 survey focuses more on characteristics of the technology, relative to design for usability. Participants were asked to rate the usability of the device after the last test trial. Due to variability in subjective evaluations, normalization was applied to QUEST 2.0 and USE scores.

H. Data Analysis

Descriptive statistical analyses were performed on all response measures, including calculation of means and standard deviations, along with graphical analysis. In the graphical analysis, the error bars reflect the standard error of the mean. These analyses were conducted to identify the overall distribution of datasets, relationships between the independent and dependent variables, and whether there was evidence of interactions among the independent factors. The descriptive statistics provided a basis for additional inferential analyses.

Based on the experiment design, the statistical model for inferential analysis was a 3-way mixed effects model with two fixed effects and one random effect. Since there was a full crossing of control mode and task factor settings, the model also included an interaction term. The random subject effect was a product of convenience sampling from the population. The subject term was involved in multiple interactions, which were pooled in the model error term.

For inferential statistical analysis, we adopted various methods depending on the type of response measure. For responses that were continuous in observation and satisfied parametric test assumptions, including PCPS, BR, task performance (TP) and the NASA-TLX, we applied an analysis of variance (ANOVA) to the statistical model along with Tukey's post-hoc tests (honestly significant difference; HSD) for multiple condition comparisons. For the usability measures, which did not uphold parametric model assumptions due to their discrete quantitative nature, we applied non-parametric methods. Tests included the Kruskal-Wallis test, as a one-way alternative to the ANOVA, and the Wilcoxon rank sum test for multiple comparisons of significant factor settings. In order to provide a clear and effective visual representation of our results, in all graphs we use error bars to represent the standard error of the mean and asterisks to denote significant differences in the control modes as well as interactions. Specifically, one asterisk is used when $p < 0.05$, two asterisks are used when $p < 0.01$, and three asterisks are used when $p < .001$.

III. Results

Results of the inferential analyses are summarized in Table II. For the training portion of the study, results indicated that, on average, participants using the PR control mode required fewer trials to meet the training criteria, as compared to those using the DC and CC modes. The means and standard errors were as follows: PR=7.4 ± 1.16 trials; DC=8.2 ± 1.51 trials; CC=8.2 ± 1.25 trials.

A. Task Performance

Fig. 5 (a) and (b) present the mean TP for the CRT and SHAP-DHT by control modes, accordingly. Fig. 5 (a) reveals that, on average, when performing the CRT task, participants exhibited highest performance when using the PR control mode and the lowest TP when using the DC mode (CC=9.23 pins; DC=7.33; PR=10.50). However, as shown in Fig. 5 (b), on average, participants achieved greater performance in the SHAP-DHT when using the DC device and worse performance when using the CC device (CC=94.53s; DC=60.63s; PR=65.27s).

Since the TP measures for both tasks were continuous with random effects, the 3-way ANOVA model was applied to both the CRT and SHAP-DHT responses. Normality tests (Shapiro–Wilks test: $p_{(CRT)} = 0.16$; $p_{(SHAP)} = 0.42$) indicated no parametric test assumption violations. ANOVA results revealed that for the CRT, participant performance differed significantly among control modes ($F_{(2,27)} = 4.08, p = 0.03, \eta^2 = 0.23$). Tukey's tests showed that the number of clothes pins relocated using DC mode was significantly less than for PR control ($p < 0.05$). The asterisks on the histograms indicate statistically significant differences between the means. However, no significant differences were detected among control modes when participants performed the SHAP-DHT ($F_{(2,27)} = 1.87, p = 0.17, \eta^2 = 0.12$).

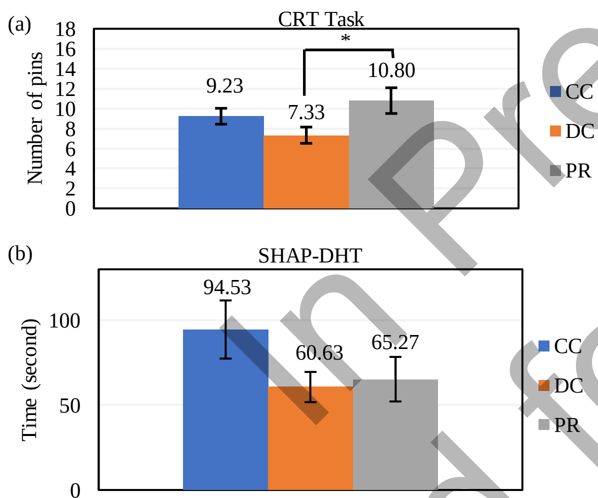

Figure 5. Comparison of TP across different prosthetic device control modes. Error bars represent standard errors on means. (a) Mean number of clothes pins moved using different device control modes. (Significant difference between DC and PR ($p < 0.05$) indicated by asterisk.) Moving more pins indicates superior performance. (b) Mean task completion time using different device control modes, where a shorter duration indicates better performance.

B. Cognitive Workload Measurements

Fig. 6 (a) and (b) present the mean PCPS and BR values by device control mode and task type. The descriptive statistics indicated that participants, whether performing the CRT or SHAP-DHT, exhibited the highest PCPS and lowest BR when using the DC mode. Fig. 6 (a) also reveals that when participants used the CC and PR control to perform the same or different test tasks, on average, there were no apparent differences in PCPS responses. In contrast, as shown in Fig. 6 (b), participant BR responses were higher when performing the SHAP-DHT vs. the CRT, regardless of the device control mode. Indicating that the CRT was likely more demanding. However, when considering the same task type, the influence of the CC and PR control modes on BR responses appeared to be similar.

As with the TP variable, the 3-way ANOVA was applied to the PCPS and BR response data. Normality assumption violations led to a transformation on each response (specifically $Y' = \sqrt{Y}$) and statistical outliers were replaced with within-group averages [45]. Shapiro–Wilks tests subsequently confirmed normality of the transformed responses ($p_{(PCPS)} = 0.24$; $p_{(blink\ rate)} = 0.09$). Furthermore, according to [46], the ANOVA is robust for balanced or near-balanced designs when the data are determined to be normal, even in the presence of heterogeneity of variance.

ANOVA results on PCPS revealed significant differences among control modes ($F_{(2,27)} = 3.53, p = 0.044, \eta^2 = 0.21$) and the interaction between control mode and task ($F_{(2,147)} = 3.12, p = 0.047, \eta^2 = 0.04$). Post-hoc analysis using Tukey's HSD tests revealed participants to exhibit significantly lower PCPS when using the CC mode than with the DC mode ($p < 0.05$). Furthermore, the CC condition produced the lowest PCPS while the DC mode led to the highest PCPS in the SHAP-DHT ($p < 0.05$).

ANOVA results on the BR response revealed significant differences among task types ($F_{(1,147)} = 14.32, p < 0.001, \eta^2 = 0.09$). The SHAP-DHT appeared to produce the highest BR across control modes. However, the control mode had no significant effect on the BR of participants. In addition, from the BR measurement, there was no significant interaction between control mode and task. The asterisks on the histograms indicate statistically significant differences between the means.

TABLE II
SUMMARY OF ANALYSIS RESULTS FOR ALL RESPONSES

Responses	Device control mode	Task	Interaction
TP	CRT: Sign. (p=0.028), DC<PR; SHAP: NS.	N/A	N/A
PCPS	Sign. (p=0.044); DC>CC	NS.	Sign. (p=0.047);
BR	NS.	Sign. (p<.001); CRT < SHAP	NS.
NASA-TLX	NS.	Sign. (p<.001); CRT > SHAP	Sign. (p<.001)
QUEST	NS.	N/A	N/A
USE	NS.	N/A	N/A

Note: Sign. Means statistically significant effect (p<0.05); NS means not significant; N/A means not applicable.

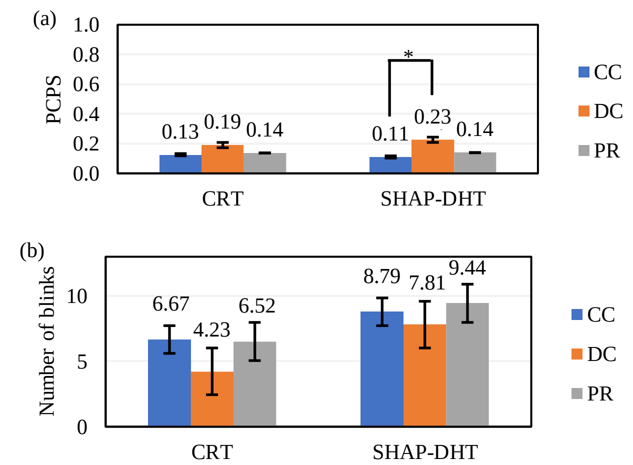

Figure 6. Comparison of prosthetic device control modes. Error bars represent standard errors on means. (a) Mean PCPS by control modes and task group. (Significant difference between CC and DC in SHAP-DHT task (p<0.05) indicated by asterisk.); (b) mean BR by control modes and task group.

Fig. 7 presents results of the cognitive workload assessment using the NASA-TLX rating method. When performing the CRT task, participants indicated the highest average workload levels (TLX scores) for the DC mode, while NASA-TLX scores appeared similar for the CC and PR control modes (DC=65.17; CC=56.89; PR=54.46). When performing the SHAP-DHT, participant NASA-TLX scores were highest when using the CC mode (DC-52.66; CC-54.56; PR-49.79). In addition, when using the DC mode, participant NASA-TLX scores appeared to be much lower for the SHAP-DHT than the CRT task.

Given that the NASA-TLX score is a continuous response, and the data did not violate the normality assumption, the 3-way ANOVA model was applied to the response. Test results revealed that participant TLX scores varied significantly between tasks ($F_{(1,27)} = 33.53, p < .001, \eta^2 = 0.19$). In addition, there was a significant interaction between the control mode and task type for the TLX response ($F_{(2,147)} = 7.52,\ p < .001, \eta^2 = 0.09$). The asterisks on the histograms indicate statistically significant differences between the means. By analyzing the ratings for the different demand components of the NASA-TLX, we found a trend of responses across components to support the general trend of the total cognitive workload score.

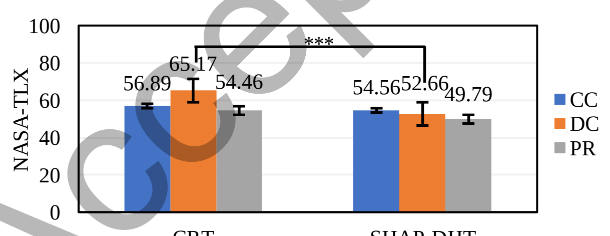

Figure 7. Mean NASA-TLX scores by control mode and task group. (Significant difference between CRT and SHAP-DHT tasks when using DC control mode (p<0.001) indicated by asterisks. Error bars represent standard errors on means.

C. Usability Assessments

Fig. 8 presents the mean usability scores from the QUEST 2.0 and USE surveys for each device control mode. The PR control mode produced the lowest scores for both the QUEST 2.0 and USE methods. Meanwhile, the CC mode produced the highest scores for QUEST 2.0 and the DC mode generated the highest score for the USE survey. Due to a small dataset and variation in survey response residuals, a non-parametric analysis method was applied to normalized QUEST 2.0 and USE scores. Results of the Kruskal-Wallis test revealed no significant differences in participant usability evaluations among the different control modes for either the QUEST 2.0 ($\chi^2 = 0.16, p = 0.92$) or USE ($\chi^2 = 0.08, p = 0.96$) methods.

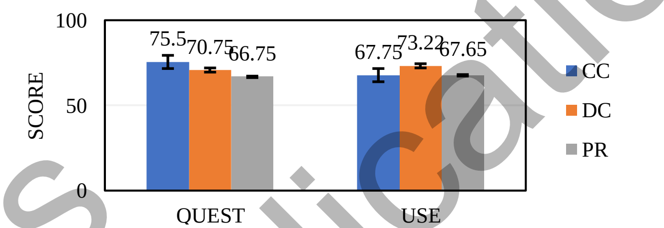

Figure 8. Mean usability assessments scores using different device control modes. Error bars represent standard errors on means.

IV. Discussion

The evidence obtained in this study provides partial support for H1. According to TP results, we inferred that when participants performed the CRT task, the DC mode was less favorable in comparison to the PR control mode operation, as found in previous studies [18-20]. However, when participants used CC, their TP was between the DC and PR control modes and suggested that the performance of participants who used the CC mode was similar to that of the DC and PR control modes in the CRT. In addition to the CRT, we also included the SHAP-DHT in our study. Results indicated that the prosthesis control modes were comparable in TP when performing the SHAP-DHT. From this result, we reviewed the experimental videos and conducted a detailed task analysis. During the CRT, participants had to accurately adjust the angle of the device to grasp and release pins successfully, as part of relocations. Achieving this control sequence with the DC mode, while ensuring accurate angle control, was challenging. In the course of the SHAP-DHT, however, participants only had to adjust the angle of the device one time before they grasped the door handle. The rest of the task did not require angle adjustments. Therefore, in the case of SHAP-DHT, the disadvantage of the DC mode was not pronounced. This analysis reveals that the DC mode, compared to PR, has limitations for activities that require frequent prosthetic joint angle adjustment. However, for activities that do not require frequent and accurate angle control, TP for the DC, PR control and CC modes was not significantly different.

The observations from this study provide support for H2 and H4, regarding the influence of control modes and task types on cognitive workload responses. Based on the PCPS results, we can infer that the DC mode imposed a higher cognitive load on participants, as compared to CC. In general, the CC mode was more intuitive and natural than the DC mode. When using CC, participants were required to perform fewer cognitive, perceptual, and motor operations to complete tasks, based on our detailed task analyses. Furthermore, we found that DC and CC differed more significantly in SHAP-DHT performance than in the CRT. This observation may be explained by the experimental design. Since all participants completed three trials of the CRT followed by the SHAP-DHT trials, this sequence could have resulted in the second task (SHAP-DHT) amplifying the effect of each control mode on participant cognitive load. However, there were no significant differences in PCPS for PR control versus DC or CC, when participants performed either the CRT or the SHAP-DHT. It was not surprising that there was no difference between CC and PR control because the two modes were controlled based on the user's intent. Although the CC mode allowed simultaneous joint operations and natural arm motion in control, since the number of controllable joints in our study was limited to two, from a user's perspective, the cognitive demand in operating the prosthetic arm with CC and PR control was similar. Unlike our previous study [19], no significant differences in PCPS were detected between PR control and DC. Looking at Fig. 6, we can see that, on average there was indeed a difference between the modes; however, variability in participant performance could have resulted in a lack of statistical sensitivity in our analysis.

Different from the PCPS results, both BR and the NASA-TLX response suggested that there were no significant differences in cognitive load among the three control modes. It is possible that the prior claim that BR measures visual workload, rather than cognitive load, is an accurate assessment. In addition, another recent study evaluated the capability of the NASA-TLX for perceptual and mental workload measurement. The study criticized the TLX for not measuring the mental construct it claims to measure [47]. The particular claim was that the TLX measures perceived task difficulty and not perceived mental load. These issues could have led to the absence of significant differences in the responses among our device control modes. However, both the BR and NASA-TLX measures indicated that it was plausible that the CRT imposed greater visual and perceptual workload on participants than the SHAP-DHT. On this basis, we inferred that participant experienced a higher subjective cognitive load when performing the CRT as compared to the SHAP-DHT. The interaction effect also indicated that participants perceived higher cognitive load when performing the CRT with the DC mode, as compared to all other control mode and task type combinations.

H3 and H4 were not supported in terms of the influence of control modes and tasks on device usability assessments. From a usability perspective, participants did not perceive any differences across the three control modes. Relevant to this finding, the USE questionnaire was not specifically designed for evaluating prosthetic devices. The USE method addresses technology usefulness, ease of use, and ease of learning, which are broad concepts. It is possible that the measure is too general in inquiries to differentiate among prosthetic device control modes, which differ in terms of the control algorithms and DOFs. Although QUEST 2.0 was specifically designed for prosthetic devices, it focuses on user satisfaction. It is possible that this limited definition of usability could have led to the lack of observation of significant differences among the three control modes. We therefore suggest that it is necessary to develop or improve existing usability analysis methods so that they can be used specifically for evaluation of prosthetic device control features.

Finally, by comparison with other prior studies [15-20], we adopted a multifaceted human-centered approach to evaluation of the impact of the three EMG-based control modes on users in terms of TP, cognitive workload experiences, and usability responses. Furthermore, we integrated three different subjective and objective measures of cognitive load to promote sensitivity or likelihood of detection of differences in workload among device control modes. Based on the above discussion, we found that these three methods indeed amplified different aspects of workload and, therefore, may be complementary to each other in this type of analysis. In summary, the human-centered evaluation method applied in this study may be helpful for a more comprehensive understanding and comparison of prosthetic device control modes and providing design guidance for future EMG-based interface technologies.

V. Conclusions

The objective of this study was to experimentally test the effects of different EMG-based prosthetic control modes on user task performance, cognitive load, and usability assessments. In addition, we aimed to generate multifaceted human-centered evaluations to inform the future design and application of EMG-based interfaces for prosthesis control. To achieve the objective, we compared three control modes (DC, PR control and CC) in terms of performance in two different tasks representing activities of daily living (CRT and SHAP-DHT). Results revealed performance with each control mode to vary from task to task. The DC mode produced limitations in tasks with angle adjustments (i.e., the CRT), as compared with PR control. However, task performance with CC was not significantly different from PR control. In general, the PR control demonstrated superior performance over DC in the CRT and exhibited comparable efficacy to CC. However, when considering the SHAP-DHT, no significant differences in performance were observed among control modes.

From our investigation, in performance of the SHAP-DHT, the CC and PR control modes appear to be more favorable than DC in terms of cognitive load, as indicated by PCPS. However, these differences were not always statistically significant, and CC was not significantly different from DC control. It appears that while CC and PR control might offer some advantages over DC in terms of potential better TP and lower cognitive load, these advantages are dependent on the task-at-hand and the response measures. On this basis, these findings should be applied with caution. For example, prosthetic designers may consider use of a DC mode by optimizing the degree of precision in control of device angle adjustments to reduce user cognitive load and potentially improve TP.

It was also clear from our analysis that using multiple types of test tasks and multiple types of human-centered evaluation methods allowed us to avoid task-induced bias in assessment of the various control modes and to promote sensitivity in detecting differences in performance and workload when they did occur. As for the usability assessment results, we found that widely used methods, such as QUEST and USE, have limitations for evaluating prosthetic control modes by able-bodied individuals. Therefore, we believe it is necessary to develop a new usability evaluation framework to allow for sensitive testing of such devices and control mode features.

Although we designed and conducted this study with the greatest possible care, there were still several limitations. Firstly, the participants were healthy young individuals, not traumatic amputees, which may make our control mode findings differ from real-world device applications. The decision to work with an able-bodied population was made due to the limited number of trans-radial amputees in the surrounding area. In addition, since most patients currently use devices with DC modes (commonly used in myoelectric control), recruiting such patients could have produced a bias in their performance. We plan to include traumatic amputees in future studies to address this limitation.

Given the difficulty of donning the experimental devices, and potential for fatigue of participants due to the length of testing, we elected to follow a between-subjects experiment design to compare the control modes. Beyond this, the control actions for DC differ from CC and PR. The presentation of all three modes to a single user could create mode confusion and lead to learning interference; thereby potentially compromising the reliability of experimental results. However, we recognize that our choice of a between-subjects design could have biased test results due to inter-individual differences in task performance, pupillary changes, blinking and subjective assessments of workload (the NASA-TLX scores) and usability (the QUEST 2.0 and USE responses). Another limitation of our study was the non-randomized order of tasks, which may have caused learning effects among participants. However, we conducted a statistical analysis on trial order, which did not reveal any significant effect on our findings. Nonetheless, it is still possible that the order of tasks influenced participant performance. Future studies may consider a randomized task order method. In addition, we only tested three prosthetic control modes (DC, PR control and CC). There are many commercial devices on the market that have yet to be tested for performance, workload, and usability outcomes.

Another limitation of the present study was manual selection of classification algorithm hyperparameters. The possibility that a more optimal set of hyperparameters exists for each participant is an aspect that future research could consider. Additional research should maximize the performance of the control algorithms relative to task demands.

For future work, we plan to use virtual reality (VR) to evaluate the performance of the three control modes, eliminating the constraints of physical prosthetic use and fatigue. We believe that this work will also lead to the development of novel methods for training on prosthetic device control methods in VR and their generalization to practical applications. In our forthcoming design, we plan to increase our sample size to ensure a richer data set for analyses. Furthermore, our aim is to quantify the prediction accuracy and robustness of the three control methods. For the PR method, we will use classification accuracy as the primary performance metric. This measure represents the percentage of correctly classified motion classes relative to the total number of classes and is highly relevant for evaluating the PR method. For the CC method, we plan to employ the root mean square error or the mean absolute error to assess performance. These metrics measure the difference between the predicted and actual joint angles or velocities, making them suitable for evaluating CC methods. However, we would like to note that the DC method does not have an associated "accuracy" metric. This is because DC is not based on machine learning, and as such, it does not make predictions or have ground truth values. We believe that this research will contribute to the advancement of prosthetic technology and can improve the lives of those who rely on these devices.

In conclusion, the number of participants in the present study was limited, which inherently restricts the statistical power and generalizability of our findings. Furthermore, the experiment followed a between-subjects design, which may introduce additional variability in results due to variations among individual participants. While our results provide some insights into the impact of different control modes and tasks on task performance and cognitive workload, these findings should be interpreted with caution. Future studies with larger samples and perhaps within-subject designs could help confirm our results and expand on understanding of prosthetic device features and workload relationships. The present work serves as an initial step in a broader assessment of such dynamics.

ACKNOWLEDGMENT

This project was funded by the National Science Foundation (No. IIS-1856676/1856441/1900044).

REFERENCES

[1] T. R. Dillingham, “Rehabilitation of the upper limb amputee,” Rehabilitation of the Injured Combatant, vol. 1, pp. 33-77, 1998.

[2] A. D. Roche et al., “Prosthetic Myoelectric Control Strategies: A Clinical Perspective,” Current Surgery Reports, vol. 2, Jan. 2014, Art. No. 44, doi: 10.1007/s40137-013-0044-8.

[3] P. A. Parker, R. N. Scott, “Myoelectric control of prostheses,” Crit Rev Biomed Eng., vol. 13, pp. 283-310, 1986.

[4] N. Parajuli et al., “Real-Time EMG Based Pattern Recognition Control for Hand Prostheses: A Review on Existing Methods, Challenges and Future Implementation,” Sensors, vol. 19, no. 20, p. 4596, Oct. 2019, doi: 10.3390/s19204596.

[5] M. Sartori et al., “Robust simultaneous myoelectric control of multiple degrees of freedom in wrist-hand prostheses by real-time neuromusculoskeletal modeling,” Journal of Neural Engineering, vol. 15, no. 6, Oct. 2018.

[6] D. L. Crouch, and H. Huang, “Lumped-parameter electromyogram-driven musculoskeletal hand model: A potential platform for real-time prosthesis control,” Journal of Biomechanics, vol. 49, no. 16, pp. 3901-3907, 2016, doi: 10.1016/j.jbiomech.2016.10.035.

[7] L. Pan, D. L. Crouch and H. Huang, “Comparing EMG-Based Human-Machine Interfaces for Estimating Continuous, Coordinated Movements,” in IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 27, no. 10, pp. 2145-2154, Oct. 2019, doi: 10.1109/TNSRE.2019.2937929.

[8] T. Kapelner, M. Sartori, F. Negro, and D. Farina, “Neuro-Musculoskeletal Mapping for Man-Machine Interfacing,” Scientific Reports (Nature Publisher Group), vol. 10, no. 1, Apr. 2020, Art. No. 5834.

[9] J. M. Hahne et al., "Linear and Nonlinear Regression Techniques for Simultaneous and Proportional Myoelectric Control," in IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 22, no. 2, pp. 269-279, March 2014, doi: 10.1109/TNSRE.2014.2305520.

[10] S. Muceli and D. Farina, "Simultaneous and Proportional Estimation of Hand Kinematics From EMG During Mirrored Movements at Multiple Degrees-of-Freedom," in IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 20, no. 3, pp. 371-378, May 2012, doi: 10.1109/TNSRE.2011.2178039.

[11] J. Berman, R. Hinson and H. Huang, "Comparing Reinforcement Learning Agents and Supervised Learning Neural Networks for EMG-Based Decoding of Continuous Movements," 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), 2021, pp. 6297-6300, doi: 10.1109/EMBC46164.2021.9630744.

[12] J. G. Ngeo, T. Tamei, and T. Shibata, "Continuous and simultaneous estimation of finger kinematics using inputs from an EMG-to-muscle activation model," Journal of Neuroengineering and Rehabilitation, vol. 11, no.1, pp. 122, Aug. 2014, doi: 10.1186/1743-0003-11-122.

[13] S. Abbaspour, M. Lindén, H. Gholamhosseini, A. Naber, and M. Ortiz-Catalan, "Evaluation of surface EMG-based recognition algorithms for decoding hand movements," Medical & Biological Engineering & Computing, vol. 58, no. 1, pp. 83-100, Jan. 2020, doi: 10.1007/s11517-019-02073-z.

[14] A. Krasoulis, and K. Nazarpour, "Myoelectric digit action decoding with multi-output, multi-class classification: an offline analysis," Scientific Reports, vol. 10, no. 1, 2020, Art. No. 16872, doi: 10.1038/s41598-020-72574-7.

[15] L. Resnik, H. Huang, A. Winslow, D. L. Crouch, F. Zhang, and N. Wolk, "Evaluation of EMG pattern recognition for upper limb prosthesis control: a case study in comparison with direct myoelectric control," Journal of NeuroEngineering and Rehabilitation, vol. 15, no. 1, 2018, Art. No. 23, doi: 10.1186/s12984-018-0361-3.

[16] L. Hargrove, L. Miller, K. Turner, and T. Kuiken, "Myoelectric Pattern Recognition Outperforms Direct Control for Transhumeral Amputees with Targeted Muscle Reinnervation: A Randomized Clinical Trial," Scientific Reports, vol. 7, no.1, 2017, Art. No. 13840, doi: 10.1038/s41598-017-14386-w.

[17] T. A. Kuiken, L. A. Miller, K. Turner and L. J. Hargrove, "A Comparison of Pattern Recognition Control and Direct Control of a Multiple Degree-of-Freedom Transradial Prosthesis," in IEEE Journal of Translational Engineering in Health and Medicine, vol. 4, pp. 1-8, 2016, Art no. 2100508, doi: 10.1109/JTEHM.2016.2616123.

[18] W. Zhang et al., "Cognitive workload in conventional direct control vs. pattern recognition control of an upper-limb prosthesis," 2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC), 2016, pp. 002335-002340, doi: 10.1109/SMC.2016.7844587.

[19] M. M. White et al., "Usability Comparison of Conventional Direct Control Versus Pattern Recognition Control of Transradial Prostheses," in IEEE Transactions on Human-Machine Systems, vol. 47, no. 6, pp. 1146-1157, Dec. 2017, doi: 10.1109/THMS.2017.2759762.

[20] M. Zahabi et al., "Application of Cognitive Task Performance Modeling for Assessing Usability of Transradial Prostheses," in IEEE Transactions on Human-Machine Systems, vol. 49, no. 4, pp. 381-387, Aug. 2019, doi: 10.1109/THMS.2019.2903188.

[21] M. Young, K. Brookhuis, C. Wickens, P. Hancock, "State of science: mental workload in ergonomics," Ergonomics, vol. 58, no. 1, pp. 1-17, 2014.

[22] Fréard, Dominique et al., "Subjective Measurement of Workload Related to a Multimodal Interaction Task: NASA-TLX vs. Workload Profile," in International Conference on Human-Computer Interaction, Berlin, Heidelberg, 2007, vol. 4552, doi: 0.1007/978-3-540-73110-8_7.

[23] B. Wilbanks, E. Aroke, and K. Dudding, "Using Eye Tracking for Measuring Cognitive Workload During Clinical Simulations," CIN: Computers, Informatics, Nursing, vol. 39, no. 9, pp. 499-507, Sep. 2021, doi: 10.1097/CIN.0000000000000704.

[24] O. Palinko, A. Kun, A. Shyrokov, and P. Heeman, "Estimating cognitive load using remote eye tracking in a driving simulator," in Proc. ACM 2010 Symposium on Eye-Tracking Research & Applications, New York, NY, USA, PP. 141-144, doi: 10.1145/1743666.1743701.

[25] G. Cardona, and N. Quevedo, "Blinking and Driving: the Influence of Saccades and Cognitive Workload," Current Eye Research, vol. 39, no. 3, pp. 239-244, 2014, doi: 10.3109/02713683.2013.841256.

[26] R. Martins, and J. Carvalho, "Eye blinking as an indicator of fatigue and mental load-a systematic review," Occupational Safety and Hygiene III, vol. 10, pp. 6, 2015.

[27] J. Brooklings, G. Wilson, and C. Swain, "Psychophysiological responses to changes in workload during simulated air traffic control," Biological Psychology, vol. 42, no. 3, pp. 361-377, 1996.

[28] M. Connan, E. Ramírez, and B. Vodermayer, "Assessment of a Wearable Force-and Electromyography Device and Comparison of the Related Signals for Myocontrol," Frontiers in Neurorobotics, vol. 10, no. 17, 2016, doi: 10.3389/fnbot.2016.00017.

[29] S. Deeny, C. Chicoine, L. Hargrove, T. Parrish, and A. Jayaraman, "A Simple ERP Method for Quantitative Analysis of Cognitive Workload in Myoelectric Prosthesis Control and Human-Machine Interaction," Plos One, vol. 9, no. 11, 2014, doi:10.1371/journal.pone.0112091.

[30] M. Markovic et al., "The clinical relevance of advanced artificial feedback in the control of a multi-functional myoelectric prosthesis," Journal of NeuroEngineering and Rehabilitation, vol. 15, no. 1, 2018, Art. No. 28, doi: 10.1186/s12984-018-0371-1.

[31] R. C. Oldfield, "The assessment and analysis of handedness: the Edinburgh inventory," Neuropsychologia, vol. 9, no. 1, pp. 97-113, 1971.

[32] J. Tiffin, and E. Asher, "The Purdue Pegboard: norms and studies of reliability and validity," Journal of Applied Psychology, vol. 32, no. 3, pp. 234 – 247, 1948, doi: 10.1037/h0061266.

[33] A. H. Butt, E. Rovini, C. Dolciotti, G. De Petris, P. Bongioanni, M. C. Carboncini, et al., "Objective and automatic classification of Parkinson disease with Leap motion controller", BioMedical Engineering OnLine, vol. 17, no. 1, 2018, Art. No. 168.

[34] Y. Lin, J. Breugelmans, M. Iversen, D. Schmidt, "An Adaptive Interface Design (AID) for enhanced computer accessibility and rehabilitation," Int J Hum Comput Stud, vol. 98, pp. 14-23, 2017, doi: 10.1016/j.ijhcs.2016.09.012.

[35] J. Park, M. Zahabi, D. Kaber, J. Ruiz, & H. Huang. (2020). Evaluation of activities of daily living Tesbeds for assessing prosthetic device usability. 2020 IEEE International Conference on Human-Machine Systems (ICHMS). https://doi.org/10.1109/ichms49158.2020.9209553

[36] L. Miller, "Occupational Therapy Outcomes with Targeted Hyper-reinnervation nerve transfer surgery: Two case studies," in Proc. Myoelectric Symposium, 2005.

[37] C. M. Light, P. H. Chappell, and P. J. Kyberd, "Establishing a standardized clinical assessment tool of pathologic and prosthetic hand function: Normative data, reliability, and validity," Archives of Physical Medicine and Rehabilitation, vol. 83, no. 6, pp. 776-783, 2002, doi: 10.1053/apmr.2002.32737.

[38] "Core – Best practices," Pupil Labs. [Online]. Available: https://docs.pupil-labs.com/core/best-practices/

[39] F. Shahini, M. Zahabi, B. Patranella, and A. M. A. Razak, "Police Officer Interactions with In-vehicle Technologies: An On-Road Investigation," in Proc. The Human Factors and Ergonomics Society Annual Meeting, SAGE Publications, Los Angeles, CA, USA, 2020, vol. 64, no. 1, pp. 1976-1980, doi: 10.1177/1071181320641476.

[40] W. Fuhl, M. Tonsen, A. Bulling, and E. Kasneci, "Pupil detection for head-mounted eye tracking in The wild: An evaluation of the state of the art," Machine Vision and Applications, vol. 27, no. 8, pp. 1275–1288, 2016.

[41] F. Shahini, J. Park, K. Welch, and M. Zahabi, "Effects of unreliable automation, non-driving related task, and takeover time budget on drivers' takeover performance and workload," Ergonomics, pp. 1–16, 2022.

[42] S. G. Hart, and L. E. Staveland, "Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research," Advances in Psychology, vol. 52, pp. 139-183, 1988.

[43] L. Demers, R. Weiss-Lambrou, and B. Ska, "The Quebec User Evaluation of Satisfaction with Assistive Technology (QUEST 2.0): an overview and recent progress," Technology and Disability, vol. 14, no. 3, pp. 101-105, 2022.

[44] A. M. Lund, "Measuring usability with the use questionnaire12," Usability Interface, vol.8, no.2, pp. 3-6, 2001.

[45] S. K. Kwak, and J. H. Kim, "Statistical data preparation: management of missing values and outliers," Korean Journal of Anesthesiology, vol. 70, no. 4, pp. 407-411, 2017.

[46] M. G. Larson, "Analysis of Variance," Circulation, vol. 117, pp. 115-121, Jan. 2008, doi: 10.1161/CIRCULATIONAHA.107.654335.

[47] R. D. McKendrick, and E. Cherry, "A Deeper Look at the NASA TLX and Where It Falls Short," in Proc. The Human Factors and Ergonomics Society Annual Meeting, SAGE Publications, Los Angeles, CA, USA, 2018, vol. 62, no. 1, pp. 44-48, doi: 10.1177/1541931218621010.